Robust Statistical Radio
Interferometric Methods for the
Detection of the Epoch of Reionization

Matyas David Molnar
Fitzwilliam College

UNIVERSITY OF
CAMBRIDGE

August 2022

This thesis is submitted for the degree of Doctor of Philosophy
In 21 cm cosmology, we are seeking to detect the highly redshifted spin-flip line of H\textsubscript{i} to constrain the properties of the Epoch of Reionization and to better understand the astrophysical and cosmological processes of the early Universe. The sensitivity requirements for such a detection are exacting, since astrophysical foregrounds overshadow the 21 cm signal by roughly five orders of magnitude. These foregrounds are, in principle, separable from the 21 cm signal, as they are naturally sequestered in Fourier space owing to their spectral smoothness. However, real-world factors such as calibration errors or array non-redundancy can taint this separability, which have the effect of proliferating foreground power into otherwise signal-dominated regions. A foreground avoidance strategy can be used; however, extreme calibration and instrumental requirements are needed to minimize this foreground leakage.

Radio astronomy is synonymous with big data: large data volumes need to be reduced, much of which is corrupted by radio-frequency interference, systematics and other non-thermal effects. Traditional outlier rejection algorithms are ad-hoc, often require manual input, and can be intricate and costly; these can miss anomalous effects and cause overflagging. Furthermore, commonly used calibration methods assume an underlying Gaussian noise distribution for visibilities, which can lead to sub-optimal results due to radio-frequency interference contamination and array non-redundancy. With the advent of the Square Kilometre Array and other next-generation radio telescopes, the flagging requirements for such traditional processes will be insurmountable. Automated methods that employ robust statistics will be able to adequately reduce these immense streams of interferometric data to produce uncorrupted estimates, with little to no manual input. This thesis focuses on robust and alternative statistical techniques that better deal with non-normal or contaminated radio data, with the aim of improving 21 cm power spectrum results.

The thesis outline is as follows: I start by reviewing the physics of the Epoch of Reionization and the 21 cm signal. I subsequently introduce the fundamentals of radio interferometry, and present the Hydrogen Epoch of Reionization Array (HERA) experiment and its data reduction pipeline. Following this theoretical and experimental underpinning, HERA data is scrutinized and its normality is probed, thus supporting the use of robust statistics for any interferometric
data reduction. Robust multivariate location estimators are employed to obtain new visibility results, which are compared to those from the standard pipeline in both the visibility and power spectrum domains. In addition, robust multivariate outlier detection, high-pass filtering before robust averaging and alternative inpainting methods are also discussed. Shifting the focus to calibration, I present a generalized approach to redundant calibration with non-Gaussian maximum likelihood estimation. As an extension of this work, redundant calibration solutions are compared by degenerate translation, and a unified redundant calibration solver across days is developed that provides optimal gain estimates for whole multi-day datasets. Lastly, I use multiresolution analysis with wavelets to better characterize the 21 cm signal, as it is non-stationary across redshifts due to the light-cone effect. I also show how the wavelet transform can be used for error detection by inspecting frequency-delay scaleograms.

Through this research, I demonstrate that robust statistics in radio interferometric data reduction and calibration are crucial for obtaining uncontaminated estimates, and eliminate any reliance on abstruse and ad hoc flagging procedures. While these robust methods are computationally more expensive, they can be accelerated with machine learning libraries and hardware. These robust techniques will become a cornerstone of radio astronomy on account of their ability to reduce large amounts of data with confidence and with little intervention.
## Contents

Summary iii  
Contents v  
List of Figures ix  
List of Tables xiii  
Declaration xv  
Acknowledgements xvii  

1 The Epoch of Reionization 1  
1.1 Introduction and cosmological background 1  
1.2 Structure formation 2  
1.2.1 The dark ages 2  
1.2.2 The cosmic web 3  
1.2.3 Virialization of dark matter halos 4  
1.2.4 Collapse of baryonic matter 6  
1.2.5 Gas cooling 7  
1.2.6 The first stars 7  
1.3 The Epoch of Reionization 8  
1.3.1 Reionization from the first luminous sources 9  
1.3.2 Probes for reionization 13  
1.3.3 The 21 cm line 18  

2 Radio interferometry and HERA 33  
2.1 Radio interferometry 33  
2.1.1 Theory 34  
2.1.2 Sensitivity 40  
2.1.3 Foreground mitigation 43  
2.1.4 EoR experiments 45  
2.2 HERA 46  
2.2.1 Array design 47  
2.2.2 Traditional calibration, deconvolution and imaging 49  
2.2.3 Analysis pipeline 57  
2.2.4 Latest upper limits on the 21 cm power spectrum 62
## 3 Normality and robust location estimates of HERA visibilities

3.1 Normality of HERA visibilities

3.1.1 Data visualization

3.1.2 Multivariate normality

3.2 Overview of principles of robust statistics

3.2.1 Breakdown value

3.2.2 Affine equivariance

3.2.3 Efficiency

3.3 Multivariate robust outlier detection

3.3.1 Outlier detection with robust Mahalanobis distances

3.3.2 LST-binning results and comparison to MAD-clipping

3.3.3 The role of outlier rejection

3.4 Robust location estimates of HERA visibilities

3.4.1 Multivariate median estimators

3.4.2 Preliminary results

3.4.3 H1C_IDR2.2 re-analysis

3.4.4 High-pass filtering

3.5 Gaussian process interpolation

3.6 Discussion

## 4 Robust redundant calibration

4.1 Relative redundant calibration

4.1.1 MLE under Gaussian noise

4.1.2 MLE under Cauchy noise

4.1.3 Constraining degeneracies

4.1.4 Student’s $t$-distribution

4.2 JAX

4.2.1 Automatic differentiation

4.2.2 JIT and XLA

4.3 Results for generalized redundant calibration

4.3.1 Redundant visibility solutions

4.3.2 Performance

4.3.3 Noise in MLE calculations for redundant calibration

4.4 Comparing visibility solutions from redundant calibration

4.4.1 MLE to translate between degeneracies

4.4.2 Phase gradients

4.4.3 Pairwise comparison results

4.4.4 Statistics across degenerately transformed redundant visibilities

4.5 Unified generalized redundant calibration

4.5.1 Unified MLE

4.5.2 Results

4.5.3 Performance

4.6 Tolerance for termination in redundant calibration

4.7 Discussion

## 5 Wavelets and multiresolution analysis

5.1 The short-time Fourier transform
## List of Figures

1.1 Reionization history of the IGM ........................................ 12
1.2 Quasar spectra at $5.74 < z < 6.42$ ........................................ 16
1.3 Thomson scattering of CMB photons .................................. 17
1.4 Global history of the 21 cm signal .................................... 24
1.5 Phases of the 21 cm signal .................................................. 25
1.6 21cmFAST light-cone simulations ....................................... 29
1.7 Dimensionless PS of a simulated EoR signal ......................... 30

2.1 Interferometer coordinate system ......................................... 36
2.2 The EoR window and foreground wedge ................................. 45
2.3 Illustration of the evolution of the Universe ............................ 47
2.4 HERA antenna configuration ............................................... 48
2.5 Global Sky Map at 150 MHz ............................................... 50
2.6 HERA-52 array layout and $u$-$v$ sampling ............................ 54
2.7 HERA-19 sample raw and CASA calibrated visibilities .............. 55
2.8 Bandpass gain calibration .................................................. 55
2.9 HERA-52 PSF and synthesized Fornax A image ...................... 56
2.10 Per-baseline PS analysis of 14 m EW baselines ........................ 57
2.11 Diagram of the full HERA pipeline ...................................... 58
2.12 Visibility phases after selected reduction steps ....................... 59
2.13 Final H1C_IDR2.2 visibilities for baseline (12, 13, pl) .......... 63
2.14 H1C_IDR2.2 PS upper limits by band and field ..................... 64
2.15 Published EoR limits as of August 2022 ............................... 65

3.1 Example scatter plot of LST-binned visibilities ....................... 70
3.2 Example KDE plot of LST-binned visibilities .......................... 70
3.3 Additional example KDE plots of LST-binned visibilities .......... 72
3.4 Per-baseline KDE plots of LST-binned visibilities .................... 73
3.5 Per-JD KDE plots of redundantly grouped visibilities .............. 74
3.6 $H.Z$ statistic for LST-binned + redundantly grouped visibilities before and after MAD-clipping .............................. 75
3.7 $H.Z$ statistic outcomes for either LST-binned or redundantly grouped visibilities ......................... 76
3.8 MAD- and RMD-clipping for an example visibility slice .......... 83
3.9 MAD- and RMD-clipping comparison for an example visibility slice .......................... 83
3.10 MAD- and RMD-clipped mean visibility amplitudes ............... 86
3.11 MAD- and RMD-clipped mean visibility phases .................... 86
3.12 Flagging percentages from MAD- and RMD-clipping ............... 87
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.13</td>
<td>MAD- and RMD-clipping PS results</td>
<td>87</td>
</tr>
<tr>
<td>3.14</td>
<td>2D Tukey regions for an example visibility slice</td>
<td>94</td>
</tr>
<tr>
<td>3.15</td>
<td>3D Tukey regions for a simulated Gaussian distribution</td>
<td>95</td>
</tr>
<tr>
<td>3.16</td>
<td>Visibility locations for various estimators</td>
<td>97</td>
</tr>
<tr>
<td>3.17</td>
<td>PS of location estimates of LST-binned + redundantly grouped visibilities</td>
<td>98</td>
</tr>
<tr>
<td>3.18</td>
<td>PS of location estimates of LST-binned visibilities</td>
<td>98</td>
</tr>
<tr>
<td>3.19</td>
<td>H1C_IDR2.2 PS limits by band and field with comparison to geometric median results</td>
<td>100</td>
</tr>
<tr>
<td>3.20</td>
<td>H1C_IDR2.2 PS limits by band and field with comparison to the results that omit the 570–573 channel flags</td>
<td>101</td>
</tr>
<tr>
<td>3.21</td>
<td>Example CLEAN visibility model</td>
<td>103</td>
</tr>
<tr>
<td>3.22</td>
<td>Delay filtering process</td>
<td>103</td>
</tr>
<tr>
<td>3.23</td>
<td>PS with and without HPFing</td>
<td>103</td>
</tr>
<tr>
<td>3.24</td>
<td>Visibility normality after HPFing</td>
<td>104</td>
</tr>
<tr>
<td>3.25</td>
<td>PS comparison between the geometric median and the HERA mean with and without HPFing</td>
<td>105</td>
</tr>
<tr>
<td>3.26</td>
<td>Complex Gaussian process regression of sample visibilities</td>
<td>107</td>
</tr>
<tr>
<td>3.27</td>
<td>Inpainting with Gaussian process regression</td>
<td>108</td>
</tr>
<tr>
<td>3.28</td>
<td>PS comparison between Gaussian process and CLEAN inpainted visibilities</td>
<td>108</td>
</tr>
<tr>
<td>3.29</td>
<td>Log-marginal-likelihood of 1D complex Gaussian process regression</td>
<td>109</td>
</tr>
<tr>
<td>4.1</td>
<td>Visibility amplitude fit with Gaussian and Cauchy distributions (typical)</td>
<td>115</td>
</tr>
<tr>
<td>4.2</td>
<td>Visibility amplitude fit with Gaussian and Cauchy distributions (outlier)</td>
<td>115</td>
</tr>
<tr>
<td>4.3</td>
<td>Redundant calibration negative log-likelihoods from MLE</td>
<td>116</td>
</tr>
<tr>
<td>4.4</td>
<td>Student’s t-distribution</td>
<td>119</td>
</tr>
<tr>
<td>4.5</td>
<td>JAX IR computational graph</td>
<td>124</td>
</tr>
<tr>
<td>4.6</td>
<td>$-\ln(L_{\text{rel}})$ and $R_{\text{man}}$ for redundant calibration</td>
<td>126</td>
</tr>
<tr>
<td>4.7</td>
<td>Redundant calibration visibility solution comparison between Gaussian and Cauchy assumed noise</td>
<td>127</td>
</tr>
<tr>
<td>4.8</td>
<td>Redundant calibration gain solution comparison between Gaussian and Cauchy assumed noise</td>
<td>128</td>
</tr>
<tr>
<td>4.9</td>
<td>Raw visibilities by redundant baseline group for an example visibility slice</td>
<td>129</td>
</tr>
<tr>
<td>4.10</td>
<td>Redundantly calibrated visibilities by redundant baseline group for an example visibility slice</td>
<td>130</td>
</tr>
<tr>
<td>4.11</td>
<td>Calibrated visibility comparison for redundant calibration with Gaussian and Cauchy assumed noise</td>
<td>131</td>
</tr>
<tr>
<td>4.12</td>
<td>$-\ln(L_{\text{rel}})$ with and without noise input</td>
<td>133</td>
</tr>
<tr>
<td>4.13</td>
<td>Visibility amplitude solutions for redundant calibration with and without noise input</td>
<td>133</td>
</tr>
<tr>
<td>4.14</td>
<td>Pairwise degenerate comparison between days</td>
<td>137</td>
</tr>
<tr>
<td>4.15</td>
<td>HERA analysis pipeline flags</td>
<td>137</td>
</tr>
<tr>
<td>4.16</td>
<td>$-\ln(L_{\text{deg}})/\sigma_{14\text{m}}^2$ and $R_{\text{man}}$ histograms from degenerate comparison</td>
<td>138</td>
</tr>
<tr>
<td>4.17</td>
<td>Potential flags from degenerate comparison on redundantly calibrated visibilities</td>
<td>140</td>
</tr>
<tr>
<td>4.18</td>
<td>Statistics across JDs of redundantly calibrated and degenerately transformed visibilities for each baseline group</td>
<td>142</td>
</tr>
<tr>
<td>4.19</td>
<td>Statistics across time of redundantly calibrated and degenerately transformed visibilities for the 14 m EW baselines</td>
<td>142</td>
</tr>
<tr>
<td>4.20</td>
<td>Statistics across JDs of $-\ln(L_{\text{rel}})$</td>
<td>142</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>HERA design specifications</td>
<td>49</td>
</tr>
<tr>
<td>2.2</td>
<td>Conversion between delay and $k_{\parallel}$</td>
<td>56</td>
</tr>
<tr>
<td>2.3</td>
<td>H1C_IDR2.2 bands and fields</td>
<td>60</td>
</tr>
<tr>
<td>3.1</td>
<td>Flagging metrics for MAD- and RMD-clipping</td>
<td>85</td>
</tr>
<tr>
<td>3.2</td>
<td>Properties of multivariate median estimators</td>
<td>96</td>
</tr>
<tr>
<td>3.3</td>
<td>Smoothness of location estimates</td>
<td>99</td>
</tr>
<tr>
<td>4.1</td>
<td>Redundant and optimal calibration speed comparison</td>
<td>132</td>
</tr>
<tr>
<td>4.2</td>
<td>Outlier detection from degenerate comparison</td>
<td>139</td>
</tr>
<tr>
<td>4.3</td>
<td>Identified outliers from redundantly calibrated and degenerately transformed visibilities</td>
<td>143</td>
</tr>
</tbody>
</table>
This thesis is the result of my own work, except where stated otherwise, carried out in the Astrophysics Group of the Cavendish Laboratory, University of Cambridge, between October 2017 and August 2022 (minus the period from May to October 2019, during which I did my industrial placement at Cantab Capital Partners, as required by my Centre for Doctoral Training in Data Intensive Science programme). This thesis has not been submitted for a degree or other qualification at any other university and does not exceed the 60,000 word limit set out by the Degree Committee for the Faculty of Physics and Chemistry.

Matyas David Molnar
August 2022
First and foremost, I would like to express my deepest thanks to my supervisor Bojan Nikolic for his mentorship, continued guidance and patience throughout the course of my PhD. Bojan’s deep knowledge of radio interferometry and his ability to explain concepts from orthogonal perspectives have been invaluable to my learning. His pragmatic approach to academic research is also refreshing, and I feel very grateful for having him as a mentor. I have also greatly enjoyed the countless anecdotes shared during our meetings.

Thank you to James Kent, my fellow PhD student and desk neighbour in G07 of the Battcock Centre for Experimental Astrophysics, who was a fountain of knowledge and a massive help throughout the years over which we overlapped. I thank the members of the HERA collaboration for their dedication and tireless work in getting the array up and running, and for pushing the boundaries of science. In particular, my thanks go to Nick Kern, Josh Dillon, Aaron Ewall-Wice, Nithya Thyagarajan and Ruby Byrne for all the helpful and insightful dialogue, and for taking the time to help with some of the nitty-gritty aspects of my research. I am also grateful to both Keith Grainge and Richard McMahon for the enjoyable and stimulating discussions during my viva voce and for their suggested revisions, which have assuredly improved this thesis. I would also like to thank Dave Green for his \LaTeX\ template.

My sincere gratitude also goes to Fitzwilliam College, which has been truly stellar on all fronts over the course of my doctorate, for all the support offered to me; hands down, the finest college of them all. I would also like to thank the Science and Technology Facilities Council for awarding me a studentship as part of the Centre for Doctoral Training in Data Intensive Science and the Cambridge Philosophical Society for their research studentship award.

I would like to humbly thank all the friends who have supported me throughout my PhD. I list a few noteworthy groups who were integral to this odyssey. To the Support Bubble, $SB = \{Charlotte, Daniel, Hannah\}$: \textit{dear lord}, what a sad little coterie. Sorry, but, not going to lie, you are some of the strangest and most alienating people I know (no offence); these are also the reasons why I am so fond of you. Thank you for all the silliness, memories and emotional support, and for munificently hosting me over the past few years. I can’t wait for our future adventures together. To the Blackett Boys, $BB = \{Alexander, Andrew, Joshua, Samuel\}$: to
this day, I am still utterly clueless as to why we all chose to study physics. Thank you for the unconditional support (in your own abnormal ways) and for being part of this journey with me, through thick and Jing. To my former housemates, the Noobs, $N = \{\text{Kieran, Matthew}\}$: I genuinely could not have finished this doctorate without you two; thank you for putting up with me and for making my time at Cambridge all the more enjoyable – we completely crushed it.

Lastly, my eternal thanks go to my family, Nadia, Thomas and Zoltan, for going above and beyond to support me all the way through my education. I will forever be indebted to you for your unfailing care and encouragement. Cette thèse est pour vous.
This chapter describes and lays out the mathematical foundations for the important astrophysical concepts that underpin the cosmic reionization of the Universe. I introduce the foundational principles of structure formation and the formation of the first stars and galaxies in Section 1.2. Reionization is outlined in Section 1.3, and the main methods to study this era are then introduced, with Section 1.3.3 wholly dedicated to the 21 cm hyperfine line of $\text{H}^1$ since its understanding is central to 21 cm cosmology.

1.1 Introduction and cosmological background

Our understanding of cosmology has progressed significantly over the past couple of decades. Advances in observational techniques have set increasingly tight constraints on the $\Lambda$-CDM parameterization of the Big Bang cosmological model, which describes an expanding Universe dominated by non-baryonic cold dark matter (CDM) and dark energy (which account for 26% and 69% of the total energy density; Planck Collaboration et al. 2020), whose large-scale evolution is dictated by the equations of general relativity and thermodynamics. While the Big Bang theory is well-established and provides a basic picture of the evolution of the Universe, all the way from its genesis to today, some 13.8 billion years later, there remain poorly understood and unobserved epochs in this history. An important episode is the emergence of complex cosmic structures and objects from the growth of primordial density fluctuations in the smooth and simple medium that emerged from the Big Bang. Many unanswered questions regarding this formation of structure persist. In particular, there are two large gaps in this chronology: the
period spanning from the last scattering surface of the Cosmic Microwave Background (CMB), 400,000 years after the Big Bang, to the formation of the first luminous structures, termed the dark ages, is yet unobserved; and the Epoch of Reionization (EoR), the era during which the Universe underwent its last cosmic phase transition, as the first luminous structures reionized the intergalactic medium (IGM), is still poorly understood. These deeply intertwined epochs, between redshifts $1100 \gtrsim z \gtrsim 6$, are amongst the last frontiers of cosmology.

This chapter reviews some of the theory and key concepts associated with structure formation and the EoR. In particular, the physics of the 21 cm line of H\textsc{i} is discussed in Section 1.3.3, which is key to make a statistical detection of the EoR.

For detailed reviews of the physics of reionization and the potential of the 21 cm line as an observational probe, see Furlanetto et al. (2006); Morales & Wyithe (2010); Pritchard & Loeb (2012); Mesinger (2016); Mesinger (2019); Liu & Shaw (2020).

1.2 Structure formation

The basic framework for structure formation consists of the growth of inhomogeneities in the CMB (with temperature fluctuations of the order $\sim 10^{-5}$), through gravitational instability, into the cosmic web seen today, with the first collapsed and gravitationally bound structures serving as a breeding ground for the first stars, galaxies and black holes (BHs). The present, although incomplete, model of structure formation is described in this section. Further observational constraints, simulations and theoretical work are needed to provide a more detailed picture of the evolution of the Universe and complete the narrative of large-scale structure formation.

1.2.1 The dark ages

As we look back in cosmic time, we observe increasingly high-redshifted sources that map out the history of the Universe. This cosmic archaeology, however, has a limit, as the Universe can only be imaged as long as it is transparent (at least, not with light).\(^b\) This is only the case after photon decoupling, which shortly followed recombination.

Recombination occurred when electrons and protons became bound to form H\textsc{i} around 290,000 years after the Big Bang (at $z \approx 1320$). Earlier than this epoch, the Universe was opaque as it consisted of a dense plasma, with photons tightly coupled to free electrons through

\(^a\)Redshift $z \equiv (\lambda_{\text{obs}} - \lambda_{\text{emit}})/\lambda_{\text{emit}}$, where $\lambda_{\text{obs}}$ and $\lambda_{\text{emit}}$ are the observed and emitted wavelengths of the radiation. This shift in wavelength due to cosmic expansion can be used to specify times or distances in the Universe, although the relationships between these quantities are highly non-linear.

\(^b\)We note that the early Universe could be probed (even before recombination) with gravitational waves. It is expected that there is a diffuse gravitational wave background that permeates the cosmos due to hypothetical violent processes in the early Universe. See e.g. Caprini & Figueroa (2018) for a review on primordial sources that can lead to cosmological backgrounds of gravitational waves.
Compton scattering, which in turn interacted with protons via Coulomb scattering. Following the start of recombination, photons were still coupled to the primordial plasma through their interactions with electrons at a rate $\Gamma_\gamma \propto n_e$, with $n_e$ the number density of free electrons. As $n_e$ decreased due to recombination and cosmic expansion, $\Gamma_\gamma$ similarly decreased and photons decoupled from the plasma when $\Gamma_\gamma(t_{\text{dec}}) \sim H(t_{\text{dec}})$, the Hubble parameter (expansion rate of the Universe), with $t_{\text{dec}} \approx 380,000$ year ($z \approx 1100$). As $n_e$ continued to drop, the photon mean free path increased rapidly and surpassed the horizon distance, and photons then propagated freely across the Universe. The signature of recombination is the CMB, which allows us to probe the conditions of last-scattering and gives insight into the fundamental physics of the Big Bang.

Following decoupling, an era of low photon emission called the dark ages ensued, where the only emission throughout this period was due to the hyperfine structure splitting of H i (the 21 cm line, see Section 1.3.3). This era ended when the first stars and quasars formed and ‘lit up’ the Universe, at $z \approx 20$.

The physics of the dark ages is uneventful. Following hydrogen recombination, only a few physical phenomena contribute to the evolution of the Universe: expansion, gravity, the interaction between the CMB and residual electrons, and recombination of these remaining electrons. Measurements of this era could further shed light on the evolution of the Universe and early structure formation, and has the potential to constrain cosmological parameters.

### 1.2.2 The cosmic web

The cosmic web, which emerged during the dark ages, is the largest scale manifestation of the inhomogeneous gravitational collapse of matter. Redshift survey campaigns have revealed that the spatial matter distribution of the early Universe follows a pervasive foam-like pattern on scales 1–100 Mpc. It is understood that most of the volume in the Universe resides in underdense void regions, bounded by matter concentrated in sheet-like walls embedded in a network of filaments, with dense clumps forming at the intersection of such filaments that are the birthplaces of rich clusters of galaxies (Bond et al. 1996). On scales larger than ~ 100 Mpc, the Universe is found to be statistically homogeneous and isotropic, in accordance with the cosmological principle (Scrimgeour et al. 2012).

This large-scale morphology of the mass distribution in the Universe evolved from gravitational instabilities of density perturbations in the early Universe. The Zel’dovich ‘pancake’ theory (Zel’dovich 1970) describes structure formation on supergalactic scales, where an ellipsoidal of dark matter (DM) collapses (most rapidly) along its shortest axis, resulting in sheets of planar pancake form, which is the first non-linear structure to form. This pancake can then collapse along the second axis, making a filament, which then, in turn, can drain into a dense
Chapter 1. The Epoch of Reionization

A halo. Note that high-density nodes can form at the intersection of filaments or sheets. A snapshot in cosmic time of the distribution of DM should show a mixture of such geometries, as separate regions reach different evolutionary stages of structure formation at a given time.

A competing theory to structure formation is that of hierarchical clustering, which proposes that DM structure proceeds hierarchically: low-mass halos are thought to continuously accrete matter and merge under their mutual gravitational interaction to form ever more massive halos, while filamentary networks build up into massive elongated channels that transport matter to dense clusters at the nodes of the network (Press & Schechter 1974; White & Rees 1978; Davis et al. 1985). Baryonic matter eventually falls into the potential wells of the collapsed DM structures (as the latter decouples much earlier than the former), and galaxies form and evolve in DM halos. This method of structure formation is backed by the results of $N$-body computer simulations; see e.g. the Millennium Simulation (Springel et al. 2005; Boylan-Kolchin et al. 2009) for recent results.\footnote{In the Millennium Simulation, hierarchical growth of DM structure through gravitational instability is modelled using 2160$^3$ particles, from $z = 127$ to today, in a cube-shaped region with side of length 2230 Gly.}

The Zel’dovich pancake theory and that of hierarchical clustering are ‘top-down’ and ‘bottom-up’ pictures of structure formation, respectively, and are, hence, in tension, although the latter approach is preferred in recent times. In reality, it is likely that a hybrid of these two methods contributed to the structure seen today.

1.2.3 Virialization of dark matter halos

DM halos grew from initial density perturbations of the smooth cosmological background. As soon as these perturbations reached order unity ($\rho/\bar{\rho} \sim 1$), the full regime of gravitational collapse followed.

A simplified toy model of the dynamical collapse of a spherically symmetric ‘top-hat’ can be considered to describe halo formation. Initially, the overdensity expands with Hubble flow, although excess gravity slows down the expansion until turnaround, with recollapse ensuing. A shell collapsing at a redshift $z_c$ has a linearized overdensity extrapolated to today (also termed the critical density of collapse) of (Barkana & Loeb 2001)

$$\delta_{\text{crit}}(z) = \frac{1.686}{D(z_c)} \approx 1.686 (1 + z_c)$$

(1.1)

where $D(z)$ is the linear growth factor in a flat matter-dominated Universe, given by

$$D(z) \propto \left( \frac{\Omega_\Lambda a^3 + \Omega_m}{a^2} \right)^{\frac{1}{2}} \int_0^a \frac{da'}{a^2} \left( \frac{\Omega_\Lambda a'^3 + \Omega_m}{(\Omega_\Lambda a'^3 + \Omega_m)^{\frac{1}{2}}} \right)^{\frac{1}{2}}$$

(1.2)
1.2. Structure formation

where \( a = 1/(1+z) \) is the scale factor, and \( \Omega_{\Lambda} \) and \( \Omega_m \) are the vacuum density (or cosmological constant) and matter (CDM + baryons) density parameters.

In reality, any slight violation from exact symmetry, as well as net angular momentum contributions, prevent halos from recollapsing to zero size. Instead, matter will virialize through dynamical relaxation (Henriksen & Widrow 1999), resulting in a DM halo with a centrally concentrated mass distribution.

A halo of mass \( M \) collapsing at \( z \gg 1 \) has virial radius (Barkana & Loeb 2001)

\[
R_{\text{vir}} = 0.784 \left( \frac{\Omega_m}{\Omega_m(z)} \frac{\Delta_c}{18\pi^2} \right)^{-\frac{1}{3}} \left( \frac{M}{10^8 M_\odot} \right)^{\frac{1}{3}} \left( \frac{1+z}{10} \right)^{-1} h^{-\frac{2}{3}} \text{kpc}
\]

and corresponding circular velocity

\[
V_c = \left( \frac{GM}{R_{\text{vir}}} \right)^{\frac{1}{2}} = 23.4 \left( \frac{\Omega_m}{\Omega_m(z)} \frac{\Delta_c}{18\pi^2} \right)^{\frac{1}{2}} \left( \frac{M}{10^8 M_\odot} \right)^{\frac{1}{2}} \left( \frac{1+z}{10} \right)^{\frac{1}{2}} h^{\frac{1}{3}} \text{km s}^{-1}
\]

and virial temperature

\[
T_{\text{vir}} = \frac{GM \mu m_p}{R_{\text{vir}} 2k_B} = 1.98 \times 10^4 \frac{\mu}{0.6} \left( \frac{\Omega_m}{\Omega_m(z)} \frac{\Delta_c}{18\pi^2} \right)^{\frac{1}{3}} \left( \frac{M}{10^8 M_\odot} \right)^{\frac{1}{3}} \left( \frac{1+z}{10} \right) h^{\frac{2}{3}} \text{K}
\]

where \( \Omega_m(z) = \Omega_m (1+z)^3/(\Omega_m(1+z)^3 + \Omega_\Lambda) \), \( \Delta_c = \rho_{\text{vir}}/\bar{\rho}_{\text{crit}} \) is the ratio of the virial density at the collapse redshift to the critical density, \( \mu \) is the mean atomic weight of the neutral primordial gas (in units of the proton mass \( m_p \)), \( h \) is the dimensionless Hubble parameter (such that Hubble’s constant \( H_0 = h \times 100 \text{km s}^{-1} \text{Mpc}^{-1} \)), and all other symbols have their usual meaning.

Baryonic matter accumulates into pre-existing or growing DM regions, with gas falling in at speeds comparable to \( V_c \) (see Equation 1.4) as the DM halo collapses, establishing a protogalaxy. As multiple gas streams collide, the gas eventually settles to a static configuration, and the gas shocks to \( T_{\text{vir}} \) (see Equation 1.5). At this point, two main forces are in competition: gravity and the pressure gradient of the gas. As the former exceeds the latter, the resulting instability causes runaway collapse.

In the hierarchical model of galaxy formation, small galaxies are formed first in low-mass DM halos. They then merge or accrete gas to form larger galaxies. It is, therefore, also important to characterize the abundance of halos in the early Universe, as this is indicative of the abundance of galaxies. The Press-Schechter mass function (Press & Schechter 1974),

\[
N(M) \, dM, \text{ describes the comoving number density of halos with masses in the range } [M, M +
\]

\textit{dThe Press-Schechter formalism assumes that the primordial density fluctuations were Gaussian, and that the perturbations grew according to linear theory until they reached a critical density contrast, where they subsequently evolved into bound objects of mass } M.\]
Chapter 1. The Epoch of Reionization

\[ dM \], and is given by

\[ N(M) \, dM = \frac{1}{\sqrt{\pi}} \left( 1 + \frac{n}{3} \right) \frac{\bar{\rho}}{M^2} \left( \frac{M}{M_*} \right)^{\frac{n+4}{2}} \exp \left[ -\left( \frac{M}{M_*} \right)^{\frac{n+4}{2}} \right] dM \] (1.6)

where \( n \) is the exponent of the power law of the power spectrum (PS) of primordial fluctuations (such that \( P(k) \propto k^n \)), \( \bar{\rho} \) is the mean density of the Universe, and \( M_* \) is the critical mass above which structures will form.

1.2.4 Collapse of baryonic matter

The minimum threshold mass for gas to gravitationally collapse into dense clumps in a cosmological setting is quantified by the Jeans mass, given by (Loeb & Furlanetto 2013)

\[
M_J = \begin{cases} 
1.35 \times 10^5 \left( \frac{\Omega_b h^2}{0.15} \right)^{-\frac{1}{2}} M_\odot, & \text{for } 1000 \geq z \geq 200 \\
4.54 \times 10^3 \left( \frac{\Omega_b h^2}{0.15} \right)^{-\frac{1}{2}} \left( \frac{\Omega_b h^2}{0.022} \right)^{-\frac{1}{2}} \left( \frac{1 + z}{10} \right)^{\frac{3}{2}} M_\odot, & \text{for } z \leq 100 
\end{cases} \] (1.7)

where \( \Omega_b \) is the baryonic density parameter. In CDM models with weakly-interacting massive particles (WIMPs), the Jeans mass of the DM is negligible. The cases in Equation 1.7 differ for the given redshift ranges as the Jeans mass depends on the IGM temperature \( T_{\text{IGM}} \). For high redshift \( z > 200 \), it is assumed that the temperature of the baryons in the IGM traces the CMB, such that \( T_{\text{IGM}} \propto T_\gamma \propto (1 + z) \), and that for \( z \leq 100 \) baryons cool adiabatically according to \( T_{\text{IGM}} \propto \rho_b^{\gamma_i - 1} \propto (1 + z)^2 \), where \( \rho_b \) is the baryon density and \( \gamma_i \) is the adiabatic index. Radiative processes, such as photoionization, can significantly increase the Jeans mass, and must be considered once the first luminous sources form.

In the context of baryonic matter collapsing within DM halos, the Jeans mass equals the total mass of the galaxy. The DM forms a halo inside which gas may cool and condense at the centre, with the collapse mass \( M > M_J \) determining the mass of the galaxy that will form. The centrifugal force of the protogalaxy prevents all the gas from falling into the centre and directly forming a BH. Fragmentation of the central gas core will give birth to stars as long as the local Jeans mass of the fragmented region drops to the mass scale of individual stars.

\(^{\text{e}}\)WIMPs are collisionless so do not feel a pressure force (although they do have a nonzero velocity dispersion), so the Jeans mass of CDM is of the order \( M_J^{\text{CDM}} \sim M_b \). All halos between \( M_J^{\text{CDM}} \lesssim M \lesssim 10^5 M_\odot \) are expected to contain mostly DM and little baryonic matter (Loeb & Zaldarriaga 2005; Loeb & Furlanetto 2013).
1.2. Structure formation

1.2.5 Gas cooling

The first stars are predicted to form at $z \sim 20$ in DM minihalos of mass $\sim 10^6 M_\odot$ (Haiman et al. 1996; Tegmark et al. 1997). As the cores of these minihalos undergo further collapse, temperatures $\lesssim 10^4$ K are required for the (metal-free) gas to form stars. However, at these temperatures, atomic transitions are not effective cooling mechanisms, as collisions are not sufficiently energetic to excite atoms so that they can, subsequently, de-excite (and cool) through line emission. For metal-free gas clouds at temperatures $\lesssim 10^4$ K, it is understood that dihydrogen $\text{H}_2$ acted as the coolant (Saslaw & Zipoy 1967), as it has adequately low energy levels and a correspondingly low excitation temperature. Hydrogen molecules cooled the gas through rotational-vibrational transitions to temperatures as low as $\sim 200$ K, limited by the spacing of the first two rotational energy levels of the molecule. Furthermore, cooling via $\text{H}_2$ becomes inefficient at the critical density $n_{cr} \approx 10^4$ cm$^{-3}$, when collisions become important enough and the populations shift to local thermal equilibrium.

1.2.6 The first stars

The details behind the process of primordial star formation are presently unknown, largely due to our ignorance of the fragmentation mechanism and its relationship with the thermodynamics of the gas. Numerical simulations have, however, helped develop a basic picture of the physics of this matter, from which a standard model of formation of the first stars emerged.

For typical halos of mass $\sim 10^6 M_\odot$, the central massive clump undergoes runaway collapse, with the collapse threshold again determined by the local Jeans mass, given by (Bromm 2013)

$$M_J \approx 500 \left( \frac{T}{200 \text{ K}} \right)^{\frac{3}{2}} \left( \frac{n}{10^4 \text{ cm}^{-3}} \right)^{-\frac{1}{2}} M_\odot$$

Typically, no sub-fragmentation of the core is expected. This last stage of collapse initiates the formation of the first stars in the Universe, categorized as Population III (Pop III) stars. This protostar will also accrete the surrounding gas, which will contribute to the final stellar mass. From dimensional arguments, the growth rate is $\dot{M}_{\text{acc}} \sim c_s^3/G \propto T^{3/2}$ (Stahler et al. 1980), although for typical stars this rate is expected to taper off with time, as radiative feedback from the star influences the surrounding gas to a greater extent than gravity (Dijkstra & Loeb 2008), and can eventually curtail accretion.

$^1$Primordial $\text{H}_2$ is thought to have formed through the $\text{H}_2^+$ channel with free protons as catalysts, with a fractional abundance of $\approx 10^{-7}$ for $z \gtrsim 400$. For $z \lesssim 110$, when CMB intensity becomes weak enough for considerable $\text{H}^-$ formation, $\text{H}_2$ abundance grew through the following two-stage chemical reaction in which a free electron acts as a catalyst: $\text{H} + e^- \rightarrow \text{H}^- + h\nu$; $\text{H}^- + \text{H} \rightarrow \text{H}_2 + e^-$. Detailed calculations show that ‘successful’ halos, where the conditions for star formation are satisfied, require a $\text{H}_2$ fraction in excess of $f_{\text{H}_2} \sim 10^{-4}$ (Ciardi & Ferrara 2005).

$^2$Stellar populations are classified by their metallicity: Population I – metal-rich stars (young); Population II – metal-poor stars (old); Population III – metal-free stars (oldest; the first stars in the Universe).
Current estimates predict the first stars to be predominantly very massive ($M_\star > 30 M_\odot$), with typical masses $M_\star \gtrsim 100 M_\odot$ (Omukai & Nishi 1998; Yoshida et al. 2008), and upper mass limits set to $M_\star \lesssim 500 M_\odot$ (Abel et al. 2002; Bromm & Loeb 2004). Some models even suggest final stellar masses of $M_\star \gg 10^3 M_\odot$; however, initial conditions to reach such masses are somewhat exotic. It is unclear if Pop III stars ever actually reach these gargantuan masses.

Constraining the masses of the first stars is crucial to understanding the evolution of the early Universe, including cosmic reionization, as their mass determines the rate of emission of ionizing photons. Furthermore, their mass determines their contribution to the metallicity of the IGM, as stars of different masses follow different collapse pathways.\(^h\) Abundances of Pop III stars must also be constrained.

Super-massive black holes (SMBHs) also formed from the merging and accretion of BHs (possible product of the collapse of Pop III star) and led to quasar activity (see e.g. Bromm & Loeb 2003; Valiante et al. 2016), thus, becoming powerful reionization sources.

1.3 The Epoch of Reionization

Reionization followed from the formation of the first luminous sources, when the light emerging from stars and quasars spread through the dark and cold Universe and began ionizing the hydrogen in the IGM, heralding the cosmic dawn. This is the final cosmic phase transition, which sees the Universe change from a fully neutral state to one in which 99.9% of atoms are ionized. This era is termed the Epoch of Reionization (EoR), and is one of the last frontiers of modern cosmology.

Studying the physical processes driving reionization, especially through observational probes, will answer many fundamental astrophysical questions. How did the first generations of stars and galaxies form, and what were their properties? How massive were Pop III stars, what are their remnants and how did they interact with each other? Through what processes did the metallicity of the IGM and galaxies increase? What is the structure of the IGM? Was it clumpy or smooth at these early times? How did feedback (radiative, mechanical, and chemical) impact galaxy formation? How did SMBHs form and what were their roles in the formation of galaxies? What is the thermal and ionization history of baryons?

The timings, duration and character of reionization contain crucial information regarding structure formation that together determine the fate of later generations of baryonic objects, and thus also the fate of the entire Universe. In addition, cosmological parameters can be further constrained from measurements of this epoch.

\(^h\)e.g. the expected collapse pathways of Pop III stars with main sequence masses: between $40 M_\odot < M_\star < 140 M_\odot$ or $M_\star > 260 M_\odot$ – directly into BHs; $140 M_\odot < M_\star < 260 M_\odot$ – pair-instability supernovae; $15 M_\odot < M_\star < 40 M_\odot$ – core-collapse supernovae (Woosley et al. 2002; Stacy & Bromm 2014).
In this section, the physics of reionization and the growth of ionization fronts are discussed, and current observational constraints from different physical probes are presented.

### 1.3.1 Reionization from the first luminous sources

The current view of reionization is that the first protogalaxies and quasars ionized the surrounding gas creating H II ‘bubbles’ (individual Strömgren spheres of a galaxy’s stars quickly join to form a single ionization front). The first galaxies formed in the most massive halos, which were biased as they were preferentially located in high-density regions with high recombination rates. As photons emerged from such dense regions, the ionization fronts propagated into the low-density IGM. The overlap stage followed, where H II bubbles merged and grew to continue ionizing the diffuse low-density IGM. When two H II bubbles joined, each point inside their common boundary became exposed to ionizing radiation from both sources. This caused a rapid rise in ionizing intensity inside the bubble, allowing the common ionization front to expand into previously inaccessible (due to a high rate of recombinations) high-density regions. Since each bubble coalescence accelerated the reionization process, the overlap phase is thought to have occurred rapidly. As the overlap stage ended, most regions in the IGM were able to see several unobscured sources, the ionizing intensity had greatly increased, and the IGM was much more homogenous. Furthermore, hierarchical structure formation models predict a rapid rise in galaxy formation rate for the relevant redshift range. This led to a highly ionized state for the low-density IGM, where H II regions were ubiquitous, bar gas inside self-shielded high-density clouds. This marked the end of reionization, which can be defined to be complete at $z_{\text{ion}}$, when the volume-weighted H I fraction fell below $10^{-3}$. Some dense neutral clumps remained in high-density structures that correspond to Lyman limit systems and damped Ly $\alpha$ systems (DLAs), although these gradually ionized with the rise in ionizing intensity (which also becomes more uniform as an increasing number of ionizing sources are visible to every point in the IGM). This post-overlap phase continued indefinitely, as collapsed objects retain neutral gas, even in the present Universe.

The course of reionization can be determined by counting photons from all ionizing sources as a function of redshift. While both stars and quasars contributed to reionization, the early Universe was dominated by small galaxies with disproportionately small central BHs, with the former thought to have had a higher impact on reionization. In Section 1.3.1.1 stellar models are considered as fiducial cases.

---

$^1$Lyman limit systems and DLAs have H I column densities $10^{17} < n_{\text{H I}} < 2 \times 10^{20} \text{cm}^{-2}$ and $n_{\text{H I}} > 2 \times 10^{20} \text{cm}^{-2}$, respectively. Systems with $n_{\text{H I}} < 10^{17} \text{cm}^{-2}$ are referred to as Ly $\alpha$ forest absorbers; these are optically thin at the Lyman limit.
1.3.1.1 \textbf{H} \textit{ii} bubbles

The radiation output from a single stellar source is initially considered, which ionizes the IGM in a growing volume to create a \textit{H} \textit{ii} bubble. Most ionizing radiation for stellar spectra is just above the \textit{H} \textit{i} ionizing threshold of $h \nu = 13.6 \text{ eV}$, implying a large absorption cross-section. Even a thin layer of \textit{H} \textit{i} can absorb all ionizing radiation emitted by stars; thus, a sharp ionization front is seen.

In the absence of recombinations in the IGM, each hydrogen atom would only have to be ionized once. However, at such high redshifts, the number density of the IGM is considerable (which varies as $n \propto a^{-3}$ (t), where we recall the scale factor $a$), and recombinations must be taken into account, as the recombination rate $\Gamma_{\text{rec}}$ is proportional to $n^2_{\text{H}}$. Hence, in a non-uniform IGM, $\Gamma_{\text{rec}}$ will be significant for gas being ionized in high-density clumps. This is accounted for by introducing a volume-averaged clumping factor $C$:

$$C = \frac{\langle n^2_{\text{H}} \rangle}{\bar{n}^2_{\text{H}}}$$  \hspace{1cm} (1.9)

where $\langle n_{\text{H}} \rangle$ and $\bar{n}_{\text{H}}$ are the local space-averaged, and mean intergalactic, hydrogen atom number densities.

For a large ionized volume compared to the typical scale of clumping (such that many clumps are averaged over and $C$ can be taken to be uniform), then the solution for a spherical ionized comoving volume $V$ around a source turned on at $t = t_i$ is given by (Barkana & Loeb 2001)

$$V(t) = \int_{t_i}^t \frac{1}{\bar{n}^0_{\text{H}}} \frac{dN_\gamma(t')}{dt'} e^{F(t', t)}$$  \hspace{1cm} (1.10)

where

$$F(t', t) = -\alpha_B \bar{n}^0_{\text{H}} \int_{t'}^t \frac{dC(t'')}{a^3(t'')}$$  \hspace{1cm} (1.11)

and $dN_\gamma/dt$ is the source production rate of ionizing photons, $\bar{n}^0_{\text{H}}$ is the current mean hydrogen number density ($= 1.88 \times 10^{-7} (\Omega_b h^2/0.022) \text{ cm}^{-3}$), and $\alpha_B$ is the recombination coefficient for hydrogen at $T = 10^4 \text{ K}$. The total number of ionized atoms within this volume is given by $N_i(t) = \bar{n}^0_{\text{H}} V(t)$.

The size of the \textit{H} \textit{ii} bubble depends on the properties of the halo that produces it. We consider a halo of mass $M$ and baryon fraction $\Omega_b/\Omega_m$, where baryons are incorporated into stars with efficiency $f_\star = 10\%$, and escape fraction $f_{\text{esc}} = 10\%$.\footnote{The escape fraction is defined as the fraction of ionizing photons that escape outside the source galaxy’s virial radius to the IGM, after accounting for recombinations within the galaxy and its halo.} Furthermore, $N_\gamma \approx 4000$
ionizing photons produced per baryon in stars is assumed. To take helium ionization into account, the correction factor $A_{\text{He}} = 4/(4 - 3Y_p) = 1.22$, where $Y_p$ is the mass fraction of helium, is introduced, which converts the number of ionizing photons to the number of $\text{H~ii}$ species (assuming that He is only singly ionized).\footnote{Helium ionization must also be considered because $\text{He~i}$ has its first ionization potential at $24.4 \text{ eV}$, which is sufficiently close to the $13.6 \text{ eV}$ required for $\text{H~i}$ so that typical stellar radiation can ionize both species together.}

Combined, these parameters determine the overall ionizing efficiency $\zeta$, defined as

$$\zeta \equiv A_{\text{He}} f_* f_{\text{esc}} N_\gamma$$

Neglecting recombinations, the maximum comoving radius of the region that a halo of mass $M$ can ionize is (Barkana & Loeb 2001)

$$r_{\text{max}} = \left(\frac{3}{4\pi} \frac{N_{\text{ion}}}{n_0^0 H_0} \right)^{1/3} = \left(\frac{3}{4\pi} \frac{\zeta}{\Omega_b \Omega_m m_p} \right)^{1/3} = 675 \left(\frac{\zeta}{40} \frac{M}{10^9 M_\odot}\right)^{1/3} \text{kpc}$$

where $N_{\text{ion}}$ is the total number of ionizing photons produced by the source. We note that $r_{\text{max}}$ is $\sim 20 \times R_{\text{vir}}$, the virial radius of a collapsing DM halo inside which galaxies form (see Equation 1.3), with this factor independent of halo mass and redshift. In reality, the radius of the $\text{H~ii}$ bubble never reaches $r_{\text{max}}$ if the recombination time is shorter than the lifetime of the ionizing source.

### 1.3.1.2 Reionization of the IGM

The fraction of the volume of the Universe that is (hydrogen) ionized is quantified by the filling factor $Q_{\text{H~ii}}$, whose evolution can be described from simple considerations. The model of individual $\text{H~ii}$ bubbles (described in Section 1.3.1.1) can be used to gain a basic understanding of the development of $Q_{\text{H~ii}}$. The latter depends on the overall production of ionizing photons, which itself depends on the collapse fraction $F_{\text{col}}$, the fraction of all the baryons in the Universe that are in galaxies (i.e. the fraction of gas that settles into DM halos and cools efficiently inside them). Assuming instantaneous photon production and neglecting complications due to density inhomogeneities, Equation 1.10 can be converted to statistically describe the reionization of the Universe (Barkana & Loeb 2001):

$$Q_{\text{H~ii}}(t) = \int_0^t \frac{dF_{\text{col}}}{dt} e^{F_{\text{col}}(t')} dt'$$

The elements of $\zeta$ (see Equation 1.12) each have large uncertainties and may evolve in time. The clumping factor $C$ also heavily depends on the pattern of ionization in the IGM, and is difficult to estimate robustly.

\footnote{This figure is valid for Pop II stars (as a fiducial model). However, Pop III stars produced $\sim 10^5$ ionizing photons per baryon. Quasars typically produce $\sim 10^4$ per baryon incorporated into the BH, but with low $f_*$ ($< 0.1\%$ in local Universe) and $f_{\text{esc}} \sim 1$. Quasars also produce a significant number of photons with $h\nu > 13.6 \text{ eV}$, resulting in a thick transition region as photons penetrate deeper into the surrounding neutral gas (Barkana & Loeb 2001).}
Figure 1.1: Evolution of the global H\textsc{i} fraction from observational constraints; these include: (i) the dark fraction in the Ly\textsc{a} and Ly\textsc{b} forests from quasi-stellar objects (QSOs) [Dark pixels]; (ii) the redshift evolution of galactic Ly\textsc{a} emission [Ly\textsc{a} fraction]; (iii) the clustering of Ly\textsc{a} emitters [LAE Clust.]; (iv) the IGM damping wing imprint in the spectra of QSO ULAS J1120+0641 [QSO DW]; (v) the CMB optical depth [Planck]; (vi) the patchy kinetic Sunyaev–Zel’dovich signal [kSZ]. The combined datasets constrain the reionization history of the Universe to $z = (8.52, 7.57, 6.82)$ (with about a 10% error at the 1\textsigma level) for an average neutral fraction $\bar{x}_\text{H} = (0.75, 0.50, 0.25)$. The red and yellow regions show the 1\textsigma and 2\textsigma constraints, respectively. Reprinted from Greig & Mesinger (2017). See e.g. Greig et al. (2022) for the latest constraints on the IGM neutral fraction.

1.3.1.3 Reionization history

There are still large uncertainties in the physical parameters that govern reionization, and this whole epoch is still poorly understood and as yet unobserved. There is much debate on the reionization history and its effect on structure formation. Nonetheless, rapid progress in observational evidence and theoretical tools have provided valuable hints on the form of the EoR. Recent results have constrained the timings and shape of reionization, as seen in Figure 1.1.

Current EoR instrumentation cannot yet observe nor fully determine the details of the EoR, and recent results still cannot place tight constraints on astrophysical parameters. Next-generation experiments will confront these arduous tasks and help answer many of the outstanding questions by imaging the EoR morphology and its temporal evolution (i.e. the 3D distribution of cosmic ionized and neutral patches as a function of space and redshift). This would be possible with 21 cm tomography, and the resulting observations would encode a wealth of information about the EoR and structure formation.
1.3. Probes for reionization

1.3.2.1 The Lyman-α line

Young high-redshift metal-poor star-forming galaxies are predicted to be bright in Lyman-α (Lyα) emission ($\lambda_{\alpha} = 1216$ Å), and searching for such emission can place constraints on the formation and luminosity function of high-redshift galaxies, the stellar initial mass function (IMF) at these high redshifts, and the reionization process of the IGM (Johnson et al. 2009).

The radiative transfer of Lyα photons through their host galaxies is generally quite complex. We can consider the simplified case whereby a fraction of ionizing photons are absorbed within their source galaxy, forming H ii regions. Within these bubbles, the newly created protons and electrons undergo recombination, producing Lyα photons. In ionization equilibrium, the rates of recombination and ionization must be equal, although around a third of recombinations are to the ground state and, thus, do not result in Lyα transitions.

The stars producing ionizing photons at these high redshifts were hot, massive and short-lived; hence, it can be assumed that the rate of ionizing photons is proportional to the galaxy’s instantaneous star formation rate $\dot{M}_*$, with the proportionality constant depending on the IMF. In Schaefer (2003), stellar evolution models predict that a galaxy with a constant star formation rate, a Salpeter IMF,\(^m\) a metallicity $Z = 0.05 Z_\odot$, and no binary stars produces $Q_i = 4.3 \times 10^{53}$ ionizing photons $s^{-1} \text{year}^{-1} M_\odot^{-1}$, where the last unit indicates per $M_\odot$ of stars created.

The intrinsic line luminosity of a galaxy is given by (Loeb & Furlanetto 2013)

$$L_{\text{Ly}\alpha}^{\text{int}} = \frac{2}{3} Q_i h_P \nu_\alpha (1 - f_{\text{esc}}) \dot{M}_*$$  \hspace{1cm} (1.15)

where $f_{\text{esc}}$ is the escape fraction and $h_P$ is Planck’s constant.

There is much uncertainty in the radiative transfer of these photons through their own galaxy and the more far-flung IGM. In such media, the Lyα line is optically thick with photons likely to scatter multiple times, sometimes away from the line-of-sight (LoS). Besides the scattering diminishing the overall brightness of the Lyα line, it also affects its frequency structure and relation to the galaxy’s continuum photons. The observed line luminosity of a galaxy is, therefore, given by (Loeb & Furlanetto 2013)

$$L_{\text{Ly}\alpha}^{\text{int}} = \frac{2}{3} T_{\text{Ly}\alpha}^{\text{IGM}} T_{\text{Ly}\alpha}^{\text{ISM}} Q_i h_P \nu_\alpha (1 - f_{\text{esc}}) \dot{M}_*$$  \hspace{1cm} (1.16)

where $T_{\text{Ly}\alpha}^{\text{IGM}}$ and $T_{\text{Ly}\alpha}^{\text{ISM}}$ are the fraction of Lyα photons transmitted through the IGM and the galaxy’s interstellar medium (ISM).

---

\(^m\)Salpeter IMF (Salpeter 1955) describes Pop I stars with a broken power law of the form $\frac{dN}{d\log M_*} \propto M_*^{\beta}$, where $\beta \approx -1.35$ for $M_* > 0.5 M_\odot$; 0.0 for $0.008 M_\odot < M_* < 0.5 M_\odot$. 
The diffuse prevalent neutral IGM during the EoR was opaque to Lyα emission by background galaxies. The spatial distribution of this diffuse, neutral IGM can, thus, be constrained by the distribution of Lyα emitting galaxies during the EoR (Haiman & Spaans 1999). A drop in the Lyα flux for galaxies at $z > 6$ has been observed (Dijkstra 2016), confirming the presence of neutral IGM at these early epochs.

**Gunn-Peterson trough**

The Lyα line can be used to probe the ionization state of the IGM in the spectra of luminous background sources, such as quasars and galaxies. It was found, by Gunn & Peterson (1965) and independently by Scheuer (1965), that the Lyα cross-section is so large that the IGM should be opaque to it for a hydrogen neutral fraction $x_{\text{HI}} \gtrsim 10^{-5}$.

Continuum photons emitted blueward of the Lyα frequency during the reionization era will redshift as they traverse the IGM due to cosmic expansion. Eventually, the wavelength of such photons will stretch near the Lyα resonance, where they can be absorbed (if the gas is still sufficiently neutral) and reemitted in a different direction. The effect of absorption from H i is quantified by the optical depth of a photon as it redshifts through the Lyα resonance.

The cross-section of a single atom is

$$\sigma_\alpha (\nu) = \frac{3 \lambda_\alpha^2 \Lambda_\alpha^2}{8 \pi} \frac{(v/v_\alpha)^4}{4 \pi^2 (v - v_\alpha)^2 + \left( \Lambda_\alpha^2/4 \right)} \left( \frac{v_\alpha}{v} \right)^6$$

(1.17)

where $\Lambda_\alpha = 8 \pi^2 e^2 f^{\text{osc}} / 3 m_e c \lambda_\alpha^2$ is the Lyα decay rate, $f^{\text{osc}}$ is the oscillator strength, and $v_\alpha = c/\lambda_\alpha$ is the frequency of the Lyα line.

If the photon is assumed to have been emitted with a wavelength $\lambda$ much farther from $\lambda_\alpha$ than the line width, the line can be approximated as narrow:

$$\sigma_\alpha (\nu) \approx \frac{3 \lambda_\alpha^2 \Lambda_\alpha^2}{8 \pi} g(v - v_\alpha)$$

(1.18)

since, in this case, the velocity shifts due to the thermal and peculiar velocities (and turbulence) in the IGM can be ignored, as these effects are small compared to cosmological redshift.

The optical depth is then given by (Fan et al. 2006)

$$\tau_\alpha = \int ds \sigma_\alpha (s)n_{\text{HI}}(s)$$

(1.19)

$$\approx 4.9 \times 10^5 \left( \frac{\Omega_m h^2}{0.15} \right)^{-1/2} \left( \frac{\Omega_b h^2}{0.022} \right) \left( \frac{1 + z}{7} \right)^{3/2} \left( \frac{n_{\text{HI}}}{n_{\text{H}}} \right)$$

(1.20)

where $s$ is the photon’s proper distance from the observer, $n_{\text{HI}} = x_{\text{HI}} n_{\text{H}}$, $n_{\text{H}}$ is the hydrogen number density, $\Omega_b$ is the baryon density parameter, and the integral is evaluated over $ds = \ldots$
1.3. The Epoch of Reionization

\[ c \, dt = c \, da/ \dot{a} = c (da/ aH), \] with the Hubble parameter \( H = \dot{a}/a \) evaluated in the matter-dominated era.

This enormous optical depth means that even \( x_{HI} \sim 10^{-4} \) gives rise to complete Gunn-Peterson (GP) absorption, as photons that redshift across the Ly\( \alpha \) line will be completely extinguished.

Each continuum photon redshifts into resonance at a particular distance from the source that depends on its initial wavelength. The photon is absorbed if there is sufficient H\( \text{I} \) at the point in the IGM where its wavelength is resonant with the Ly\( \alpha \) line. After it passes that point and redshifts further, it no longer interacts with H\( \text{I} \). Hence, the continuum photons of different wavelengths emitted by the source sample the IGM at different points along the LoS, thus, enabling the mapping of the H\( \text{I} \) distribution along the line from the source.

The resulting absorption profile is called the Ly\( \alpha \) forest, due to the strong variability of these absorption features depending on the detailed structure of the IGM along the LoS.

GP absorption can be used to constrain the last stages of reionization, with complete absorption seen in the spectra of quasars at \( z \gtrsim 6 \), as seen in Figure 1.2.

1.3.2.2 CMB constraints

As the IGM became ionized during the cosmic dawn, CMB photons began to scatter off free electrons causing small and large-scale secondary anisotropies, which can be used to probe the ionization and structure of the Universe during the EoR.

Mean optical depth

Thomson scattering by free electrons released by cosmic reionization produces additional linear polarization to the local CMB quadrupoles at the horizon scale during the EoR. This scattering suppresses CMB temperature anisotropies by a factor of \( e^{-2 \tau_T} \), where the Thomson scattering optical depth is given by

\[ \tau_T = \int dz \, n_e(z) \sigma_T \left( \frac{c \, dt}{dz} \right) \]

(1.21)

where \( n_e(z) \) is the redshift dependent free electron number density, \( \sigma_T = 6.65 \times 10^{-25} \text{cm}^{-2} \) is the Thomson scattering cross-section, and the integral is performed over the path taken by the photon.

This scattering partially washes out the CMB temperature anisotropies, as a fraction \( e^{-\tau_T} \) of the photons appearing to originate from a given direction actually come from elsewhere. Each LoS samples only photons from a finite (approximately) cylindrical region whose radius equals the causal horizon at the time of scattering (see Figure 1.3). The angular PS of CMB fluctuations \( C_l \) (\( l \) being the multipole moment) is, therefore, damped by a factor of \( e^{-2 \tau_T} \) (the
factor of 2 in the exponential is due to two factors of temperature appearing in the expression for $C_I$.

In effect, this scattering can occur at all radii about the observer. Therefore, electrons at each radius damp temperature anisotropies across their local horizon (damping proportional to $\tau_f$ by these electrons). On large scales, the PS is unaffected.

The suppression in anisotropies is highly degenerate with the amplitude of the primordial PS of scalar perturbations $A_s$, rendering temperature data alone insufficient to meaningfully separate these quantities. Reionization, however, leads to polarization, as photons scattered from an electron in a quadrupole radiation field (in this case, the large-scale CMB quadrupole) will be linearly polarized. The polarized signal produces a distinct peak on large scales (signal amplitude and therefore power scale as $\tau_f$ and $\tau_f^2$ respectively), referred to as the ‘reionization
1.3. The Epoch of Reionization

Thomson Scattering of CMB Photons

![Diagram of Thomson scattering](image)

**Figure 1.3:** Diagram of the damping of primordial CMB anisotropies from Thomson scattering with free electrons during reionization. The observer is the star at the centre of the diagram; the thick outer circle represents the surface of last scattering; the shaded clouds are parcels of free electrons at $z \sim 6$ (this distance is demarcated by a dashed circle around the observer); thin circles around each electron parcel delimits their causal horizon. CMB photons (arrows) propagate from the surface of last scattering, most of them travelling straight to the observer without interaction due to $\tau_T$ being low. However, some photons scatter into the LoS from other directions (due to interactions with electron clouds), which suppresses primordial anisotropies on small scales (relative to the horizon at the time of scattering).

bump’, which cannot be mimicked by any other parameter in the cosmological model.

Recent results from Planck Collaboration et al. (2020) found a value of $\tau_T = 0.054 \pm 0.007$, corresponding to a mid-point reionization redshift of $z_{re} = 7.64 \pm 0.74$.

**kSZ effect**

The peculiar velocities of free electrons during the EoR introduce a Doppler shift to the scattered photons, thus, further imprinting temperature anisotropies on the CMB. This effect, also known as the kinetic Sunyaev–Zel’dovich (kSZ) effect (Sunyaev & Zel’dovich 1980) has several observable consequences. In the non-relativistic limit, the kSZ effect shifts the observed CMB temperature as $(u/c)n_e$, where $u$ is the LoS bulk electron velocity, $c$ is the speed of light and $n_e$ is the density of free electrons. This implies a hot spot in the CMB if the ionized gas is moving towards the observer and a cold spot if moving away.

---

\[\text{The reionization mid-point in this model assumes simple tanh parameterization for the ionization fraction with width } \Delta z_{re} = 0.5.\]
More precisely, the total kSZ signal along the LoS $\hat{n}$ is given by
\[
\frac{\Delta T_{\text{kSZ}}}{T_\gamma}(\hat{n}) = \int d\eta a(\eta)^{-2}e^{-\tau_T(\eta)}\sigma_T n_e(\eta, \hat{n}) \frac{\hat{n} \cdot u(\eta)}{c}
\]
and $u$ is the peculiar velocity of the ionized gas, $\bar{x_i}(\eta)$ is the mean ionization fraction, $n_{e,0}$ is the mean electron density of the Universe today, $\delta_x$ and $\delta_b$ are the baryonic mass and ionization fraction overdensities, and the integral is over conformal time $\eta = \int_0^t dt'/a(t')$. A net kSZ signal requires perturbations in $n_e(\eta, \hat{n})$ that are correlated with the large-scale velocity field.

The kSZ signal can be considered to be composed of two distinct contributions: a homogeneous signal due to $\delta_b$ perturbations after the Universe is fully ionized and a patchy signal from $\delta_x$ perturbations during reionization. The kSZ signal from patchy reionization is sourced by $\text{H}^\text{ii}$ regions (see Section 1.3.1.1). The proper motion of such regions generates angular anisotropy through the kSZ effect. The $\text{H}^\text{ii}$ bubbles around the first stars, galaxies and quasars are correlated with the velocity field because their sources of ionization are biased tracers of the matter distribution. It is found that the amplitude of the patchy kSZ signal is predominantly dependent on the duration of reionization. The longer the contrast between neutral and ionized regions persists, the longer the patchiness can induce fluctuations in the CMB temperature. Alternatively, the kSZ power is linearly proportional to the number of $\text{H}^\text{ii}$ bubbles along $\hat{n}$, which scales with duration. The shape of the signal is also dependent on the distribution of bubble sizes, as the kSZ power will peak at larger scales and vice versa. Furthermore, the signal can offer insight into when reionization started, as an early reionization (of a set duration) would take place in a denser Universe, leading to more kSZ power (Gruzinov & Hu 1998; Loeb & Furlanetto 2013; Reichardt 2016).

Using the 2500 deg$^2$ South Pole Telescope-Sunyaev-Zel’dovich (SPT-SZ) survey, the kSZ power was found to be $D_{\text{kSZ}} = 2.9 \pm 1.3 \mu\text{K}^2$ at $l = 3000$, with the patchy kSZ PS constraining the duration of reionization$^6$ to $\Delta z < 5.4$ (George et al. 2015).

### 1.3.3 The 21 cm line

Of the probes that can investigate the ionization state of the IGM and constrain reionization, which include the Ly$\alpha$ line and the kSZ effect (see Sections 1.3.2.1 and 1.3.2.2), the strongest candidate to study the early phases of structure formation is the ‘spin-flip’ line of H$\text{i}$, which was theoretically predicted by van de Hulst (1945) and first detected by Ewen & Purcell (1951).

---

$^6$Here, the duration of reionization is defined as the time the Universe takes to go from a 20% to 99% volume-averaged ionization fraction (at the 95% confidence level) (George et al. 2015).
1.3. The Epoch of Reionization

This corresponds to the transition between the $F = 0$ and $F = 1$ hyperfine levels of ground state $\text{H}_1^1\text{S}_{1/2}$ (where $n^{28+1}\text{L}_J$ spectroscopic notation is used). This emission has frequency $\nu_{21} \approx 1420\text{ MHz} \ (\lambda_{21} \approx 21\text{ cm}, \ E_{21} \approx 6\ \mu\text{eV}$), and is driven by the interaction of the magnetic dipole moments of the nucleus and electron, resulting in an increase in energy when the spins are parallel and a decrease when antiparallel.

The transition is highly forbidden (since $\Delta L = 0$), with the probability between these hyperfine structure levels being $\sim 10^{23}$ smaller than that of allowed optical transitions, predominantly due to the $v_{21}^3$ term appearing in the equation for the Einstein $A_{ij}$ coefficient. This transition is extremely weak, with a half-life for spontaneous emission of $t_{1/2} \approx 1.1 \times 10^7\text{ year}$, and is optically thin through the IGM. Hydrogen, being the most abundant element in the Universe, can, thus, serve as a tracer for local properties of the gas. While the signal is very faint, the transition energy is sufficiently low that it provides a sensitive calorimeter of the thermal state of the IGM. This low-frequency radio transmission, which can be detected through interferometric measurements, can be used to better understand the physics of the early Universe, and to even observe it. Furthermore, the effect of redshift allows the Universe to be ‘sliced’ at different redshifts in the radial direction. This 3D mapping tomography would image the distribution of $\text{H}_1$ as a function of space and cosmic time, and highlight the ‘Swiss-cheese’ structure of the IGM during reionization.

1.3.3.1 Radiative transfer and brightness temperature of the 21 cm line

The radiative transfer equation for the specific intensity $I_\nu$ of a line is

$$\frac{dI_\nu}{ds} = \frac{\phi(\nu) h\nu}{4\pi} \left[ n_1 A_{10} - (n_0 B_{01} - n_1 B_{10}) I_\nu \right]$$

(1.24)

where $ds$ is a proper path length element, $\phi(\nu)$ is the normalized line profile function ($\int d\nu \phi(\nu) = 1$), subscripts 0 and 1 denote the lower (singlet) and upper (triplet) atomic levels, $n_i$ is the number density of level $i$, and $A_{ij}$ and $B_{ij}$ are the Einstein coefficients for the transitions between levels $i$ and $j$, the initial and final states. For the 21 cm transition, $\nu_{10} = 1420\text{ MHz}$ and $A_{10} = 2.85 \times 10^{-15}\text{ s}^{-1}$, with the Einstein coefficients interrelated through the standard Einstein relations

$$A_{10} = \frac{2h\nu^3}{c^2} B_{10}$$

(1.25)

$$g_0 B_{01} = g_1 B_{10}$$

(1.26)

where $g_i$ is the spin degeneracy of the $i^{\text{th}}$ energy level (here $g_0 = 1$ and $g_1 = 3$).

The relative populations of the two spin states in $\text{H}_1$ are characterized by the spin temperature $T_S$, such that

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{-T_S/T_S}$$

(1.27)
where \( T_\star = h\nu_{10}/k_B = 0.068 \) K is the temperature equivalent to the energy difference from the hyperfine splitting, with \( E_{10} = h\nu_{10} = 5.9 \times 10^{-6} \) eV. It is always found that \( T_\star \ll T_S, T_\gamma \), so all related exponentials can be expanded to leading order. This implies that \( \sim 3/4 \) atoms are in the excited state, making stimulated emission an important process that must be considered.

We now quantify \( I_\nu \) by the brightness temperature \( T_b(\nu) \), defined according to the spectrum of a blackbody radiator using the Rayleigh–Jeans formula:

\[
I_\nu = B_\nu(T_b) \approx \frac{2\nu^2 k_B T_b}{c^2}
\]  

which is an excellent approximation for the Planck curve for the range of frequencies and temperatures relevant to the 21 cm line. In this limit, we may solve Equation 1.24, by means of the integrating factor \( e^{\int d\tau_\nu} \), along the LoS path through a uniform cloud of hydrogen:

\[
T'_b(\nu) = T_S \left( 1 - e^{-\tau_\nu} \right) + T'_R(\nu) e^{-\tau_\nu}
\]  

where \( T'_R \) is the brightness of the background radiation field incident on the cloud along the ray (i.e. the CMB such that \( T'_R = T_\gamma(z) \)), and the optical depth \( \tau_\nu \) is defined as

\[
\tau_\nu \equiv \int ds \alpha_\nu(s)
\]  

with the integral performed along the ray through the cloud, where \( s \) is the proper distance and \( \alpha_\nu \) is the absorption coefficient, given by

\[
\alpha_\nu = \phi(\nu) \frac{h\nu}{4\pi} (n_0 B_{01} - n_1 B_{10})
\]  

and we have used

\[
\frac{\phi(\nu) h\nu}{4\pi \alpha_\nu} n_1 A_{10} = \frac{j_\nu}{\alpha} = B_\nu(T_S) = \frac{2\nu^2 k_B T_S}{c^2}
\]  

where we have related the emissivity \( j_\nu \) to the source function \( B_\nu \) to the spin temperature \( T_S \) through the Rayleigh–Jeans formula (Equation 1.28).

Applying the quantities relevant to the 21 cm transition, the optical depth can be rewritten as

\[
\tau_{10} = \int ds \frac{3A_{10}}{8\pi \nu^2} \left( 1 - e^{-T_\star/T_S} \right) \phi(\nu) n_0
\]  

where we have used Equations 1.25 to 1.27.

Due to cosmological redshift, the emergent brightness temperature \( T'_b(\nu_0) \) measured in the cloud’s comoving frame creates an apparent brightness at the observer on Earth of \( T_b(\nu) = T'_b(\nu_0)/(1 + z) \), where the observed frequency is \( \nu = \nu_0/(1 + z) \), and \( \nu_0 \) is the rest frequency of the emission. We proceed to work with these observed quantities.
1.3. The Epoch of Reionization

The most important component of line broadening comes from cosmic expansion, leading to a Doppler shift of photons of the form $\phi(\nu) \sim c/[sH(z)\nu]$. Evaluating the optical depth yields (Madau et al. 1997)

$$
\tau_{10} = \frac{3}{32\pi} \frac{hc^{3}A_{10}}{k_{B}T_{S}v_{10}^{2}} \frac{x_{H_{I}}n_{H}}{(1+z) \left(\frac{dv_{//}}{dr_{//}}\right)}
$$

$$
\approx 0.0092 \left(1 + \delta_{b}\right) (1+z)^{2} \frac{x_{H_{I}}}{T_{S}} \left[\frac{H(z)/(1+z)}{\frac{dv_{//}}{dr_{//}}}\right]
$$

(1.35)

where the Hubble parameter in a matter-dominated Universe (valid for $3400 > z > 0.4$) is

$$
H(z) = H_{0}\sqrt{\Omega_{m} (1+z)^{3} + \Omega_{k} (1+z)^{2} + \Omega_{\Lambda}}
$$

(1.36)

and $H_{0}$ is the Hubble constant, $x_{H_{I}}$ is the H I fraction, $n_{H}$ is the local hydrogen number density, $\frac{dv_{//}}{dr_{//}}$ is the gradient of proper velocity along the LoS, including both the Hubble expansion and the peculiar velocity, and $\delta_{b}$ is the baryon density contrast. The last term of Equation 1.35, in the square brackets, amounts to the redshift-space distortion due to the peculiar velocity of the gas relative to the Hubble flow along the LoS. For fiducial values of these parameters, $\tau_{10} \sim O(1\%)$ at $z \sim 7$.

The cosmological signature of interest is the contrast between the 21 cm emission and the CMB observed on Earth. The observed flux is expressed by the differential antenna temperature between the patch in the sky and the CMB, $\delta T_{b} = T_{b} - T_{\gamma}(z)$, and, for $\tau_{10} \ll 1$, is given by (Furlanetto et al. 2006)

$$
\delta T_{b}(\nu) = \frac{T_{S} - T_{\gamma}(z)}{1+z} \left(1 - e^{-\tau_{10}}\right)
$$

$$
\approx \frac{T_{S} - T_{\gamma}(z)}{1+z} \tau_{10}
$$

$$
\approx 9x_{H_{I}}(1+\delta_{b}) (1+z)^{2} \left[\frac{T_{S} - T_{\gamma}(z)}{T_{S}}\right] \left[\frac{H(z)/(1+z)}{\frac{dv_{//}}{dr_{//}}}\right] \text{mK}
$$

(1.39)

Further substitution of components into Equation 1.39 yields the alternative formulation commonly seen in the literature that explicitly depends on cosmological density parameters:

$$
\delta T_{b} \approx 27x_{H_{I}}(1+\delta_{b}) \left[\frac{T_{S} - T_{\gamma}(z)}{T_{S}}\right] \left[\frac{\Omega_{b}h^{2}}{0.022}\right] \left[\frac{1+z}{10} \frac{0.15}{\Omega_{m}h^{2}}\right] ^{1/2} \left[\frac{H(z)/(1+z)}{\frac{dv_{//}}{dr_{//}}}\right] \text{mK}
$$

(1.40)

The brightness temperatures can be replaced by specific intensities to obtain

$$
\delta I_{b} = 9x_{H_{I}}(1+\delta_{b}) (1+z)^{2} \left[\frac{I_{H_{I}} - I_{\gamma}(z)}{T_{S}}\right] \left[\frac{H(z)/(1+z)}{\frac{dv_{//}}{dr_{//}}}\right] \text{W} \text{m}^{-2} \text{sr}^{-1} \text{Hz}^{-1}
$$

(1.41)
or alternatively

$$\delta I_b(\nu) \approx 27 \chi_{\text{HI}} (1 + \delta_0) (1 + z)^2 \left[ \frac{I_{\text{HI}} - I_\gamma(z)}{T_0} \right] \left[ \frac{\Omega_0, h^2}{0.022} \right] \left( \frac{1 + z}{1 + 0.15 \Omega_m, h^2} \right)^{\frac{1}{2}} \left[ \frac{H(z)/(1 + z)}{d v_\parallel / d r_\parallel} \right] \text{ W m}^{-2} \text{ sr}^{-1} \text{ Hz}^{-1}$$ (1.42)

where $I_{\text{HI}}$ and $I_\gamma(z)$ are the emission brightnesses of $\text{HI}$ and the CMB.

The redshift-space distortion term in Equations 1.39 to 1.42 is often omitted in the literature, as the effect of LoS peculiar velocity can be assumed to be negligible (to first order) compared to the Hubble flow. See e.g. Bharadwaj & Ali (2004); Barkana & Loeb (2005b) for studies of the effects of peculiar velocities on the 21 cm brightness temperature.

For Equations 1.39 to 1.42, $\delta T_b < 0$ yields an absorption signal, while $\delta T_b > 0$ an emission signal. Moreover, $\delta T_b$ saturates for $T_S \gg T_\gamma$, but can become arbitrarily large (and negative) if $T_S \ll T_\gamma$. The detectability of the fluctuations is highly dependent on $T_S$, which must deviate from the background temperature for a signal to be discerned.

### 1.3.3.2 The spin temperature

Three competing processes determine $T_S$ (Wouthuysen 1952; Field 1958, 1959): (i) particle collisions; (ii) resonant absorption and stimulated emission of the CMB; (iii) scattering of Ly$\alpha$ ultraviolet (UV) photons. The rate of these processes is fast compared to the de-excitation time of the line; hence, $T_S$ is approximately given by the equilibrium balance of these effects:

$$\frac{d n_0}{dt} = n_0 (C_{01} + R_{01} + W_{01}) = n_1 (C_{10} + R_{10} + W_{10}) = \frac{d n_1}{dt}$$ (1.43)

where $C_{ij}, R_{ij}$ and $W_{ij}$ are the collisional, radiative, and Ly$\alpha$ transitional rates from state $i \rightarrow j$, with 0 being the ground state and 1 the excited state. Using the principle of detailed balance, the relationship between the upward and downward rates are given by

$$C_{01} = \frac{g_1}{g_0} C_{10} e^{-T_\gamma/T_K} \approx 3 C_{10} \left( 1 - \frac{T_\gamma}{T_K} \right)$$ (1.44)

$$R_{01} = \frac{g_1}{g_0} R_{10} e^{-T_\gamma/T_\gamma} \approx 3 R_{10} \left( 1 - \frac{T_\gamma}{T_\gamma} \right)$$ (1.45)

$$W_{01} = \frac{g_1}{g_0} W_{10} e^{-T_\gamma/T_C} \approx 3 W_{10} \left( 1 - \frac{T_\gamma}{T_C} \right)$$ (1.46)

where $T_K$ is the gas kinetic temperature, $T_\gamma$ is the CMB temperature, $T_C$ is the colour temperature of the Ly$\alpha$ radiation field at the Ly$\alpha$ frequency (closely coupled to $T_K$ by recoil during repeated scattering), and the first order approximations on the RHS for these Boltzmann distributions are
valid since \( T_* \) is much smaller than all temperatures. The radiative rates relate to the Einstein coefficients in the radiative transfer equation (Equation 1.24) through

\[
R_{01} = B_{01} I_\gamma \\
R_{10} = A_{10} + B_{10} I_\gamma
\]

such that \( R_{01} \) is responsible for stimulated absorption, and \( R_{01} \) responsible for stimulated and spontaneous emission.

Further working with these relations and making use of Equations 1.25, 1.26 and 1.28, the spin temperature can be reduced to (Field 1958; Zaldarriaga et al. 2004; Furlanetto 2016)

\[
T_S^{-1} = \frac{T_\gamma^{-1} + x_C T_K^{-1} + x_\alpha T_C^{-1}}{1 + x_C + x_\alpha}
\]

where and \( x_C \) and \( x_\alpha \) are the collisional and UV scattering coupling coefficients, which are given by

\[
x_C = \frac{C_{10} T_*}{A_{10} T_\gamma}
\]

\[
x_\alpha = \frac{W_{10} T_*}{A_{10} T_\gamma}
\]

We note that \( C_{10} \), the collisional de-excitation rate of the \text{H\,i} triplet hyperfine level, is proportional to \( n_H \) (Allison & Dalgarno 1969), and \( W_{10} \), the indirect de-excitation rate of the triplet level due to absorption of an Ly\(\alpha \) photon followed by decay to the singlet level, is approximately given by \( 1.3 \times 10^{-12} J_{-21} \) s\(^{-1} \), with \( J_{-21} \) the intensity of the background radiation field at the Ly\(\alpha \) frequency in units of \( 10^{-21} \) erg cm\(^{-2} \) s\(^{-1} \) Hz\(^{-1} \) sr\(^{-1} \).

In most situations of interest, \( T_C \to T_K \), which simplifies Equation 1.49 to

\[
1 - \frac{T_\gamma}{T_S} = \frac{x_C + x_\alpha}{1 + x_C + x_\alpha} \left( 1 - \frac{T_\gamma}{T_K} \right)
\]

The relative strengths of the processes quantified in Equations 1.49 and 1.52 vary with redshift, and couple \( T_S \) to \( T_\gamma \) when resonant absorption of the CMB dominates, and to \( T_K \) when collisional coupling or UV scattering become important. See Furlanetto et al. (2006); Pritchard & Loeb (2012) for detailed reviews on the physics that determine \( T_S \).

1.3.3.3 The global signal

While there are many physical phenomena that govern the 21 cm signal, the 21 cm brightness temperature can ultimately be expressed as a function of four variables: \( T_b = T_b(T_K, x_{\text{H\,i}}, J_\alpha, n_H) \) (Pritchard & Loeb 2012), where \( J_\alpha \) is the specific flux of the background radiation field evaluated at the Ly\(\alpha \) frequency. The dependence of \( T_b \) on each of these variables saturates at certain
Figure 1.4: The predicted 21 cm cosmic signal, from decoupling through to the end of the EoR. Upper panel: time evolution of fluctuations in the 21 cm brightness temperature, pieced together from redshift slices from a simulated cosmic volume. Colouration indicates the strength of the 21 cm brightness: purple and blue – absorption (of which there are two phases); red – emission; black – no signal. Lower panel: expected evolution of the sky-averaged (global) 21 cm brightness temperature signal against the CMB. During the dark ages, gas cooled adiabatically with the expansion of the Universe, seen as an absorption trough in the global 21 cm signal, with fluctuations arising from variation in density. As the first stars and galaxies formed and lit up the Universe, their radiation greatly altered the properties of the IGM. Scattering of Lyα photons leads to a strong coupling between the excitation of the 21 cm line spin states and the gas temperature, which initially caused a strong varying absorption signal due to the strong clustering of the first generation of galaxies. Following this, X-ray emission from these galaxies heated the gas leading to an emission signal. As reionization ran its course, UV photons further ionized the IGM, creating Hii bubbles that surrounded groups of galaxies, thus marking dark holes in the 21 cm brightness temperature (upper panel). Eventually, all the hydrogen gas, except for a few dense pockets, became ionized. There is much debate and uncertainty revolving around the shape of this signal due to the unknown properties of the first galaxies. Adapted from Loeb & Furlanetto (2013).

points, leading to different regimes where variations in only one of these parameters dominate fluctuations of the 21 cm signal. A summary of the latest predictions of this global signal, separated into these different physical regimes, is shown in Figure 1.4, with the key features indicated as a function of cosmic time. The different stages affecting the 21 cm signal are also shown in the simplified illustration of Figure 1.5.

To summarise, the signal progresses from an early phase of collisional coupling to a late phase of Lyα coupling, separated by a brief period devoid of any 21 cm signal. Following this, fluctuations in the signal originate from X-rays, then ionizing UV radiation backgrounds. After reionization is complete, there is a faint residual signal from H I in galaxies.
1.3. The Epoch of Reionization

The properties of the first luminous sources. Reprinted from Pritchard & Loeb (2012).

The chronology in this cartoon is still debated, although the presented version appears to be the most likely based on the relative energetics of the relevant physical processes and the properties of the first luminous sources. Reprinted from Pritchard & Loeb (2012).

**Figure 1.5:** Cartoon of the different phases and critical points between the dark ages and the end of reionization that contribute to fluctuations in the 21 cm signal. $1100 \geq z \geq 200$: Compton scattering from residual free $e^-$ from recombination maintains thermal equilibrium between the gas and CMB; thus, $T_K = T_\gamma$. Collisional coupling is also still effective as the gas density remains high enough at this early stage, so $T_S = T_\gamma$. This period is expected to be devoid of any signal with $T_b = 0$. $200 \geq z \geq 40$: the gas decouples from the CMB and cools adiabatically with $T_K \propto (1+z)^2$, while $T_\gamma$ falls steadily as $\propto (1+z)$. We, thus, get $T_K < T_\gamma$. Collisional coupling still causes $T_S$ to track $T_K$, so $T_S < T_\gamma$, which gives rise to an initial absorption signal (around $z \sim 80$) as $T_b < 0$. During this regime, fluctuations in $\delta T_b$ arise from density fluctuations. $40 \geq z \geq z_\alpha$: further expansion ensues that leads to decreased gas density; collisions become less frequent and this coupling becomes inefficient. CMB radiative coupling, however, persists, and $T_S$ reflects $T_\gamma$ again, meaning that no 21 cm signal is observed. $z_\alpha \geq z \geq z_\gamma$: the first stars alight at $z_\alpha$ and flood the Universe with both Ly$\alpha$ and X-ray photons that begin to ionize the Universe. Ly$\alpha$ photons resonantly scatter off $H_1$ via the Wouthuysen-Field effect, coupling the spin and gas temperatures ($T_\gamma \sim T_K$). The gas is still cold due to the adiabatic expansion, so $T_S \sim T_K < T_\gamma$ and an absorption feature is seen. Variations in both density and Ly$\alpha$ flux affect fluctuations in $\delta T_b$. Further star formation eventually saturates Ly$\alpha$ coupling ($x_\alpha \gg 1$). We denote $z_\alpha$ as the redshift where the gas will be strongly Ly$\alpha$ coupled everywhere. $z_\alpha \geq z \geq z_\delta$: Following Ly$\alpha$ coupling saturation, the 21 cm signal is no longer depends on the Ly$\alpha$ flux. At this stage, heating is considerable, and fluctuations in $\delta T_\gamma$ are primarily sourced by $T_K$ fluctuations. For $T_K < T_\gamma$, absorption is seen, but as $T_K$ approaches $T_\gamma$, some emission may start to be seen from hotter regions. At $z_\delta$, the gas is fully heated such that $T_K = T_\gamma$. $z_\delta \geq z \geq z_T$: Heating pervades the Universe and drives $T_K > T_\gamma$, with $T_S$ tracing $T_K$; therefore, we expect an emission signal. Saturation of $\delta T_b$, however, occurs at $z_T$, when $T_S \sim T_K \gg T_\gamma$. At this point, $x_i \gtrsim 10^{-2}$. Fluctuations in $\delta T_b$ are now due to contributions from $x_i$, $T_K$, $n_H$. $z_T \geq z \geq z_R$: at this later stage when $z = z_T$, heating drives $T_K \gg T_\gamma$ and temperature fluctuations are negligible. We now have $T_5 \sim T_K \gg T_\gamma$, so we can simplify Equation 1.39 by removing the dependence on $T_5$, which also leads to a more straightforward PS. By this point, $Q_{H\alpha}$ becomes important, and the 21 cm signal is primarily sensitive to fluctuations in the density and ionization fields. $z \lesssim z_R$. Following the end of the EoR, any remaining 21 cm signal is emitted by collapsed overdense and self-shielding pockets of $H_1$ (i.e. DLA). The boundaries of these separate regimes are not clear-cut, and there is likely to be overlap between them. The exact chronology in this cartoon is still debated, although the presented version appears to be the most likely based on the relative energetics of the relevant physical processes and the properties of the first luminous sources. Reprinted from Pritchard & Loeb (2012).
In March 2018, Bowman et al. (2018) claimed to have detected a flattened absorption profile in the sky-averaged radio spectrum centred at 78 MHz with a full-width at half-maximum (FWHM) of 19 MHz and an amplitude of 0.5 K, using observations from the EDGES experiment (Bowman & Rogers 2010). It is claimed that the profile appears consistent with predictions from the mean 21 cm global signal induced by the first stars. The low-frequency edge of the profile suggests that stars had existed and already produced a background of Lyα photons by 180 million years after the Big Bang, and the high-frequency edge indicates that the gas was heated to above the radiation temperature \(\lesssim 100\) million years later. This result for the absorption profile, however, has a large depth, and a flat-bottomed shape, both of which are unexpected; the best-fitting amplitude is more than a factor of two greater than the largest predictions, which gives cause for concern. Such a result would suggest that either the primordial gas was much colder than expected or the background radiation temperature was hotter than expected. It is theorized that solely the cooling of the gas as a result of interactions between DM and baryons seems to explain the observed amplitude, which has sparked many new DM models. There is much debate surrounding the validity of this ‘detection’ (e.g. Hills et al. 2018). The SARAS 3 experiment (Nambissan T. et al. 2021) refutes this claimed detection in Singh et al. (2022). The EoR community awaits further confirmation of the (in)validity of Bowman et al. (2018) from other global experiments.

1.3.3.4 Power spectrum analysis

Imaging sensitivity to \(\delta T_b\) is difficult to achieve due to its faintness, with the noise of first-generation EoR experiments being comparable to or even exceeding the signal itself, rendering tomographic measurements out of reach with current instrumentation. Consequently, efforts have been focused on constraining the statistical properties of the IGM with low signal-to-noise (S/N) data (Furlanetto et al. 2006).

We first define the fractional perturbation to \(\delta T_b\): \(\delta_{H_1}(x) \equiv [\delta T_b(x) - \bar{\delta T}_b] / \bar{\delta T}_b\). The PS of \(\delta_{H_1}(x)\) is defined as the Fourier transform (FT) of the two-point correlation function (Barkana & Loeb 2005a,b):

\[
\langle \tilde{\delta}_{H_1}(k)\tilde{\delta}^*_{H_1}(k') \rangle \equiv (2\pi)^3 \delta^D(k-k')P_{H_1}(k)
\] (1.53)

where \(\tilde{\delta}_{H_1}(k)\) is the FT of the hydrogen density field, \(k\) is the comoving wavevector, \(\delta^D\) is the Dirac delta function, and \(\langle \ldots \rangle\) denotes ensemble average.

Expanding Equations 1.39 and 1.49 to linear order in each of the perturbations (Furlanetto et al. 2006):

\[
\delta_{H_1} = \beta_b \delta_b + \beta_x \delta_x + \beta_\alpha \delta_\alpha + \beta_T \delta_T - \delta_\nu
\] (1.54)
showing the dependence of $\delta T_b$ on its input parameters, where $\delta_i$ is the fractional variation in $i$, with $i$ covering the baryon density ($b$), the neutral fraction ($x$), the Ly$\alpha$ coupling coefficient ($\alpha$), the gas temperature ($T$), and the LoS peculiar velocity gradient ($\partial v$). The expansion coefficients $\beta_i$ are given by

$$\beta_b = 1 + \frac{x_C}{x_{\text{tot}} (1 + x_{\text{tot}})}$$  (1.55)

$$\beta_x = 1 + \frac{x_{\text{HH}}}{x_{\text{tot}} (1 + x_{\text{tot}})}$$  (1.56)

$$\beta_\alpha = \frac{x_\alpha}{x_{\text{tot}} (1 + x_{\text{tot}})}$$  (1.57)

$$\beta_T = \frac{T_\gamma}{T_K} + \frac{1}{x_{\text{tot}} (1 + x_{\text{tot}})} \left( x_{\text{C}} \frac{d \ln x_H}{dT_K} + x_{\text{HH}} \frac{d \ln x_{\text{HH}}}{dT_K} \right)$$  (1.58)

where $x_{\text{tot}} \equiv x_C + x_\alpha$, $x_i^{ij}$ is the collisional coupling coefficient between species $i$ and $j$, $x_\alpha$ is the Ly$\alpha$ coupling coefficient, and $k_{10}^{ij}$ is the specific scattering rate coefficient for spin de-excitation by collisions between species $i$ and $j$ (units of cm$^3$ s$^{-1}$). By linearity, $\tilde{\delta}_{\text{H}i}$ can be written in a similar fashion.

Each of the expressions in Equations 1.55 to 1.58 has a straightforward physical interpretation: $\beta_b$ – the first term describes the increased matter content and the second describes the increased collisional coupling efficiency in dense gas; $\beta_x$ – the two terms describe direct fluctuations in the ionized fraction and the effects of the increased electron density on $x_C$ (the latter is only consequential in partially ionized regions); $\beta_\alpha$ – measures the fractional contribution of the Wouthuysen–Field effect to the coupling; $\beta_T$ – the first term parameterizes the speed at which $T_S$ responds to fluctuations in $T_K$, while the others include the explicit temperature dependence of the collision rates (Furlanetto et al. 2006). We note that fluctuations in $\delta_x$ can be $\sim 1$ if the ionization field is built from H ii regions, meaning that terms such as $\delta_b \delta_x$ are, in fact, first order, and must still be retained as they can still significantly contribute to the PS.

By Equation 1.53, the PS contains all possible terms of the form $P_{\delta_i \delta_j}$. In most instances, the terms in $\delta_j$ will be correlated in some way; statistical 21 cm observations hope to measure these quantities separately.

The homogeneity and isotropy of the Universe suggest that the contributions to the PS should be spherically symmetric in Fourier space, i.e. $\delta(k) = \delta(k)$. However, peculiar velocity gradients introduce redshift-space distortions since the direction of the observer becomes important, and hence only cylindrical symmetry is preserved. It can be shown that $\tilde{\delta}_{\partial \nu} = -\mu^2 \tilde{\delta}$, where $\tilde{\delta}$ is the FT of the underlying density field (for most purposes, the baryon density field

---

7The Wouthuysen–Field effect (Wouthuysen 1952; Field 1959) is a coupling mechanism whereby the $1/2 S_{1/2}$ and $1/2 S_{1/2}$ hyperfine sublevels of H i can mix through the absorption (jumping to either the $2/2 P_{1/2}$ or $2/2 P_{3/2}$ sublevels) and re-emission of Ly$\alpha$ photons. Here, $n_F L_n$ notation is used to describe hyperfine sublevels. See e.g. Furlanetto (2016) for further details on the Wouthuysen–Field effect.
is equivalent to the total matter density \( \delta_b = \delta \), and \( \mu \equiv \hat{k} \cdot \hat{r} \) (cosine between wavevector and LoS) (Kaiser 1987). The PS of the fractional perturbation to the differential brightness temperature is, hence, given by (Pritchard & Loeb 2012)

\[
P_{\text{HI}}(k, \mu) = P_{\mu^0}(k) + \mu^2 P_{\mu^2}(k) + \mu^4 P_{\mu^4}(k) + P_{f(k, \mu)}(k, \mu)
\] (1.59)

where

\[
P_{\mu^0}(k) = P_{bb} + P_{sx} + P_{a\alpha} + P_{TT} + 2P_{bx} + 2P_{b\alpha} + 2P_{bT} + 2P_{x\alpha} + 2P_{xT} + 2P_{aT}
\] (1.60)

\[
P_{\mu^2}(k) = 2P_{b\delta} + 2P_{x\delta} + 2P_{a\delta} + 2P_{T\delta}
\] (1.61)

\[
P_{\mu^4}(k) = P_{\delta \delta}
\] (1.62)

\[
P_{f(k, \mu)}(k, \mu) = 2P_{x\delta \delta_{\beta\gamma} \chi} + P_{x\delta_{\beta\gamma} \delta_{\beta\gamma} \chi} + \text{other quartic terms with } \delta_{\beta\gamma} \] (1.63)

High precision EoR experiments of the 3D PS should allow the separation of \( P_{\text{HI}}(k, \mu) \) into the 4 terms in Equation 1.59 by their angular dependence on the powers of \( \mu^2 \). The \( P_{f(k, \mu)}(k, \mu) \) term with its more complicated angular dependence, however, threatens this decomposition, as its contribution is important during the final stages of reionization. We, therefore, note that the angular decomposition of Equation 1.59 may not be possible when ionization fluctuations are considerable.

When evaluating the 21 cm PS from interferometric measurements (see Section 2.1), it is favourable to work in 3D \( k \)-space, since in the absence of cosmic evolution and velocity distortions, the 21 cm PS is spherically symmetric (it is imperative to consider such inherent symmetries to construct a high S/N statistic; this can also be exploited to distinguish foregrounds from the \( \text{H} \) signal). This symmetry is due to the isotropy of space on large spatial scales, implied by the cosmological principle. Furthermore, the signal can be considered to be homogeneous if it undergoes minimal evolution over a sufficiently short time interval (observationally corresponding to a narrow frequency band, as we recall that the frequency of the redshifted 21 cm line maps to LoS distance).

### 1.3.3.5 Simulation results

The cosmological 21 cm signal can be simulated using 21cmFAST\(^{\dagger}\) (Mesinger et al. 2011; Murray et al. 2020), a seminumerical modelling tool that can generate 3D realizations of the density, ionization, peculiar velocity and spin temperature fields, which together can be used to compute \( \delta T_b \), the brightness temperature and its PS.

\(^{\dagger}\)https://github.com/21cmfast/21cmFAST
1.3. *The Epoch of Reionization*

Figure 1.6: 21cmFAST light-cones showing the evolution of the neutral fraction $x_{\text{HI}}$ and the brightness temperature contrast $\delta T_b$ over redshifts $13.2 \geq z \geq 6.1$. Reionization is patchy for $z \gtrsim 9$, after which it gains momentum and larger $\text{H} \text{II}$ bubbles start forming, with the reionization midpoint occurring at $z = 7.5$. The (hydrogen in the) Universe is $\approx 85\%$ ionized at $z = 6.1$.

I use standard cosmological parameters and a fiducial model of reionization (taken to be the default 21cmFAST values) for our simulation. I evolve a box of length 600 Mpc with 1200 cells along each side over a redshift range $13.2 – 6.1$ (corresponding to 100 – 200 MHz) to get the brightness temperature and ionization fractions in Figure 1.6. This model indicates a midpoint of reionization at $z = 7.5$, and has a global reionization history consistent with Figure 1.1.

The corresponding 21 cm dimensionless PS $\Delta_{\text{HI}}^2(k) \equiv \frac{k^3}{2\pi^2} P_{\text{HI}}^I(k)$ are shown in Figure 1.7 at a selected few redshifts. The computed PS from 21cmFAST is fairly robust to slight changes in
Figure 1.7: 21 cm dimensionless PS of simulated EoR data generated by the 21cmFAST code at several selected redshifts. Neutral fractions $x_{\text{HI}}$ are also given for each of the indicated redshifts. At $x_{\text{HI}}$ close to 1, H II bubbles are unimportant and the 21 cm PS traces the matter PS. When the H II bubbles first appear, they suppress the power by ionizing the highest density regions first. Soon afterwards, $P_x$ becomes large, dominating fluctuations and creating a ‘shoulder’ at $k_{pk}$, which can clearly be seen in the lower redshift PS. The scale of the peak associated with the shoulder of the 21 cm PS is directly related to H II bubble size: $k_{pk}$ decreases as reionization progresses, and the sharper the peak, the sharper the bubble distribution at the evaluated redshift. For $k < k_{pk}$, $P_H \propto P_{\delta\delta}$, when $x_{\text{HI}}$ is still relatively big. However, once bubbles become large, their Poisson fluctuations take over, and $P_H$, starts to resemble white noise. We also notice that $\Delta^2_{\text{HI}}$ is relatively flat, meaning that it is accessible for detection at a wide range of $k$ modes.

reionization parameters and agrees with current observations. This result can, therefore, be used as a reference in EoR experiments or alternative simulations with different parameter values. See e.g. Pober et al. (2014) for another simulation with a vanilla reionization model. The shape and peak of the (dimensionless) PS depend on the ionization fraction and the parameters of the reionization model. The peak, in particular, is crucial to constraining reionization, as it is directly related to the characteristic size of H II bubbles.

1.3.3.6 From theory to observation

The main objective for current 21 cm interferometric experiments is to detect the 21 cm PS and to constrain the astrophysics and cosmology during the EoR. This chapter studied the physics of structure formation, reionization and the 21 cm line, and reviewed the mechanism by which
the PS of the contrast between the 21 cm signal and the CMB can be used to infer statistical properties on the IGM in the early Universe.

Next, in Chapter 2, I show how interferometric observation with the Hydrogen Epoch of Reionization Array can be used to measure the 21 cm PS. Thereafter, the research presented in this thesis focuses on the use of robust and alternative statistical data reduction and calibration methods; these aim to improve 21 cm PS results and make them more resilient to errors, since real-world interferometric data is plagued with outliers from e.g. radio-frequency interference and can exhibit non-normal behaviour.
Radio interferometry and HERA

Observational 21 cm cosmology is a hugely promising field that can set constraints on the EoR, including its timing and duration, as well as shedding light on the physical processes that underpin this cosmic phase transition. Low-frequency radio experiments aim to directly probe this era by measuring the redshifted 21 cm signal.

The research in this thesis analyses the interferometric measurements from the Hydrogen Epoch of Reionization Array (HERA), with the end target of the collaboration to set upper limits and even detect the 21 cm PS. The fundamentals of radio interferometry are laid out in the first part of this chapter (Section 2.1), and I show how interferometric visibility measurements are related to the 21 cm PS. Sensitivity requirements and foreground mitigation strategies are also discussed. Section 2.2 introduces the HERA experiment, and its data reduction and PS estimation pipelines are outlined. Sample visibility data and synthesized images are also shown.

2.1 Radio interferometry

Observations of 21 cm fluctuations, during and prior to reionization, require low-frequency telescopes since frequency decreases with redshift as \( v_{\text{obs}} = 1420/(1+z) \) MHz. The diffraction limit dictates the finest achievable angular resolution:

\[
\theta_D \sim \frac{\lambda}{D_{\text{max}}}
\]

(2.1)

where \( D_{\text{max}} \) is the maximum dimension of the telescope (i.e. the dish diameter for a single dish telescope or the baseline for an interferometer, which is a set of antennas that work together as a single telescope) and \( \lambda = \lambda_{\text{obs}} = \lambda_0 (1+z) \), with \( \lambda_0 = 21 \) cm for EoR observations. The largest
angular scales of an interferometer are also bound by the minimum spacing between antennas: 
\( \theta_M \sim \lambda/D_{\text{min}} \). Moreover, the field of view (FoV) is set by \( \sim \lambda/d \), where \( d \) is the size of an individual antenna in the array. With their higher angular resolution, interferometers can better resolve radio sources to create sharper images. This does, however, come at a cost in surface brightness sensitivity, which is discussed in Section 2.1.2.

As we will see in Section 2.1.1.1, an interferometer probes Fourier modes on the sky, which can then be transformed to the image domain to provide a sky map. The ultimate aim of 21 cm cosmology is to make full 3D tomographic maps of the neutral hydrogen in the early Universe. The sensitivity requirements for such a feat are strenuous, and there are also instrumental systematic concerns. Therefore, current experiments have focused on the statistical characterization of the 21 cm signal through global signal and PS experiments, which are both complementary to 21 cm tomography. The former seek to measure the mean brightness temperature \( \delta T_b \) over redshift, where the signal is averaged over all angles on the sky (also called the monopole of \( \delta T_b \); see Section 1.3.3.3). The latter aims to measure the PS of the 21 cm signal, which quantifies the power of the spatial fluctuations of the \( \delta T_b \) field as a function of length scales (see Section 1.3.3.4). As later shown in Section 2.1.1.2, the 21 cm PS relates directly to the Fourier modes measured by an interferometer. I note that second/next-generation interferometric experiments, such as the Square Kilometre Array (SKA) (Dewdney et al. 2009), which is currently under construction and uses existing experiments as precursor telescopes (e.g. MWA and HERA), will eventually pursue imaging tomography of H I.

This section outlines the salient principles of interferometry, with regard to 21 cm cosmology and the HERA array.

2.1.1 Theory

2.1.1.1 Fundamentals of interferometry

An interferometer works by measuring the electric field from the sky at each antenna in its array. Each antenna pair forms a baseline, and the signals from each antenna in a baseline are then cross-correlated to form a quantity that we call a visibility \( V \). More formally, each baseline measures the mutual coherence function

\[
V(r_1, r_2) = \langle E_1(r_1, t) E_2^*(r_2, t) \rangle
\]

(2.2)

where the signals \( E_i \) are measured at points \( r_i \), and the expected value of the product is taken (i.e. we take the integral over \( t \)). These visibilities are the output from the interferometer and have dimensions of spectral power flux density. In radio astronomy, we usually work with visibilities in units of Jansky, where 1 Jy = \( 10^{-26} \) W m\(^{-2}\) Hz\(^{-1}\).
2.1. Radio interferometry

Following Thompson (1999), it can be shown that this visibility response can be expressed in relation to the source intensity distribution \( I(l, m, \nu) \) (also called sky brightness distribution or specific intensity) for spatially incoherent radiation from the far field:

\[
V(u, v, w, \nu) = \iint \frac{dl \, dm}{\sqrt{1 - l^2 - m^2}} A(l, m, \nu) I(l, m, \nu) e^{-2 \pi i [u + v \sin \theta + w \sqrt{1 - l^2 - m^2} - 1]} 
\]

(2.3)

where \( \nu \) is the spectral frequency, \((u, v, w) = b/\lambda\) are the east-west (EW), north-south (NS) and LoS projections of the baseline vector \( b \) between the antennas of the interferometer in units of wavelength \( \lambda \), \((l, m, n)\) is the position vector on the sky plane measured with reference to a defined location, referred to as the phase centre, with \( \{l = \sin \theta_x, m = \sin \theta_y, n = \sqrt{1 - l^2 - m^2}\} \) denoting the direction cosines toward east, north and zenith, and \( A(l, m, \nu) \) is the primary beam of the antenna, which is a windowing function describing the FoV and bandpass response of an interferometer pair. The integrand is taken to be zero for \( l^2 + m^2 \geq 1 \). See Figure 2.1 for a diagram of the coordinate system used for interferometric observation.

In the flat-sky approximation, we take \( A(l, m, \nu) \) to be sufficiently small such that \( l, m \ll 1 \), making \( \sqrt{1 - l^2 - m^2} \approx 1 \). The dependence of the visibility on \( w \), which arises from non-coplanarity of the interferometric array, becomes very small and can, hence, be omitted.\(^a\) HERA operates in this regime. The visibility equation then becomes

\[
V(u_\nu, \nu) = \iint du \, dv \, A(\hat{s}, \nu) I(\hat{s}, \nu) e^{-2 \pi i u \cdot \hat{s}}
\]

(2.4)

where \( u_\nu \equiv (u, v) \), and \( \hat{s} \equiv (l, m) \).

Under certain assumptions, the van Cittert-Zernike theorem (Zernike 1938) states that there exists a Fourier relationship between the mutual spatial coherence function (i.e. the visibility) and the sky brightness \( I(\hat{s}, \nu) \), such that the inverse FT of the visibility of a distant, incoherent source is equal to its intensity distribution. See e.g. Born & Wolf (1999); Thompson (1999); Wilson et al. (2013) for derivations, assumptions and applications of the van Cittert-Zernike theorem. Hence, we can write the modified brightness distribution \( T = AI \) as

\[
T(\hat{s}, \nu) = \iint du \, dv \, V(u_\nu, \nu) e^{2 \pi i u \cdot \hat{s}}
\]

(2.5)

Each visibility recorded at a point in the \( u-v \) plane, therefore, samples a component of the FT of the brightness distribution, with visibilities at small \( u, v \) (small baselines) probing large-scale structure and vice-versa. Naturally, an interferometer has incomplete \( u-v \) coverage, which can be quantified with a weighting function \( W(u_\nu) \) (= 0 where \( V \) is not sampled). By performing

\(^a\)See Paul et al. (2016) for research on the effect of the \( w \)-term in HI PS estimation. They find that the \( w \)-term causes an effective shrinking of the primary beam \( A(l, m, \nu) \), which reduces the contribution of the HI signal.
Chapter 2. Radio interferometry and HERA

Interferometer Coordinate System

Figure 2.1: Diagram of the $u, v, w$ and $l, m, n$ right-handed coordinate system used to express the interferometer baselines and the source brightness distribution. Here, $d\Omega = dl \, dm \, dn$ is the differential solid angle for the source element, $\hat{s}_0$ is the phase centre, such that all other positions can be given by $\hat{s} = (\hat{s}_0 + \sigma)/\sqrt{|\hat{s}_0|^2 + |\sigma|^2}$. The geometric delay $\tau_g = \vec{b} \cdot \vec{s}/c$ is due to the wavefront from the source arriving at one antenna $\tau_g$ later than the other. The instrumental delay $\tau_i$ can be introduced to compensate for $\tau_g$ (setting $\tau_i = \tau_g$ ensures that signals arrive simultaneously at the correlator) or can be continuously adjusted for source tracking. The main lobe and sidelobes of the antennas are also shown in orange.

The inverse FT of the sampled visibility function, we recover the ‘dirty’ image:

$$T'(\hat{s}, \nu) = \iint du \, dv \, W(u, \nu) V(u, \nu) e^{2\pi i u \cdot \hat{s}}$$

$$= B(\hat{s}, \nu) * T(\hat{s}, \nu)$$

where

$$B(\hat{s}, \nu) = \iint du \, dv \, W(u, \nu) e^{2\pi i u \cdot \hat{s}}$$

is the point spread function (PSF) (also called the synthesized or dirty beam), and Equation 2.7 arises because $T'$ is the FT of the product of $W$ and $V$, which is equal to the convolution (represented by *) of the FTs of $W$ and $V$.

There exist standard imaging methods, such as the CLEAN algorithm (see Section 2.2.2.2), that use non-linear techniques to interpolate/extrapolate visibility samples into unsampled regions of the $u$-$v$ plane.
Returning to Equation 2.4, the sky intensity can be decomposed as

\[ I(\mathbf{s}) = \bar{I}(\nu) + \Delta I(\mathbf{s}, \nu) \]  

(2.9)

where \( \bar{I}(\nu) \) and \( \Delta I(\mathbf{s}, \nu) \) are the isotropic and fluctuating parts of the intensity distribution. Since the isotropic component does not contribute to interferometric measurement, the visibility can be expressed as

\[ V(u, \nu) = \int dl \, dm \, A(\mathbf{s}, \nu) \Delta I(\mathbf{s}, \nu) e^{-2\pi i u \cdot \mathbf{s}} \]  

(2.10)

In EoR experiments, the measured visibility receives contributions from the redshifted H i line, foregrounds and noise.

### 2.1.1.2 Relating complex visibilities to the 21 cm power spectrum

The H i fluctuations in Fourier space (FT of \( \delta_{\text{H}i} \) in Equation 1.54) determine the 21 cm signal contribution to \( \Delta I(\mathbf{s}, \nu) \):

\[ \Delta I_{\text{H}i}(\mathbf{s}, \nu) = \bar{I}(\nu) \int \frac{d^3k}{(2\pi)^3} \tilde{\delta}_{\text{H}i}(k) e^{ik \cdot r} \]  

(2.11)

where \( \mathbf{r} \equiv (\mathbf{s}, r_\nu) \) specifies the 3D position of the H i emission, with \( r_\nu \) the transverse comoving coordinate distance to the point of observation.\(^b\)

The H i PS can be constructed from the correlation of the observed visibilities. Following Paul et al. (2016), we substitute the form of intensity fluctuations from Equation 2.11 into Equation 2.10:

\[ V(u, \nu) = \bar{I}(\nu) \int \frac{d^3k}{(2\pi)^3} \tilde{\delta}_{\text{H}i}(k) e^{ik \cdot r} \times \int dl \, dm \, A(\mathbf{s}, \nu) \times \exp \left[ -2\pi i \left( \frac{\mathbf{u} \cdot \mathbf{s} - k_\perp r_\nu}{2\pi} \right) \cdot \mathbf{s} \right] \]  

(2.12)

where \( \mathbf{k} \) has been split into components on the plane of the sky \( k_\perp \), and along the LoS \( k_\parallel \). The integral over angles is the FT of the primary beam \( A(\mathbf{s}, \nu) \), so Equation 2.12 can be rewritten:

\[ V(u, \nu) = \bar{I}(\nu) \int \frac{d^3k}{(2\pi)^3} \tilde{\delta}_{\text{H}i}(k) e^{ik \cdot r} a \left( \frac{\mathbf{u} \cdot \mathbf{s} - k_\perp r_\nu}{2\pi} \right) \]  

(2.13)

where

\[ a \left( \mathbf{u} \cdot \mathbf{s} - \frac{k_\perp r_\nu}{2\pi} \right) \equiv \int dl \, dm \, A(\mathbf{s}, \nu) \times \exp \left[ -2\pi i \left( \frac{\mathbf{u} \cdot \mathbf{s} - k_\perp r_\nu}{2\pi} \right) \cdot \mathbf{s} \right] \]  

(2.14)

Now, using Equation 1.53, we can compute the visibility correlation function:

\[ \langle V(u, \nu) V^*(u', \nu') \rangle = \bar{I}_s \bar{I}_s \int \frac{d^3k}{(2\pi)^3} P_{\text{H}i}(k) e^{ik \cdot (r - r')} \times a \left( \frac{\mathbf{u} \cdot \mathbf{s} - k_\perp r_\nu}{2\pi} \right) a \left( \frac{\mathbf{u'} \cdot \mathbf{s} - k_\perp r'_\nu}{2\pi} \right) \]  

(2.15)

\(^b\)Recall, \( r_\nu = \int dz / H(z) \), with integral limits extending from zero to the emission redshift \( z = \nu_{\text{emit}} / \nu_{\text{obs}} - 1 \) (Hogg 1999). We often see \( r_\nu \) written as \( D_M \) in the literature.
This gives the correlation of the \( \text{H} \text{i} \) signal in 3D, in which the \( \bm{u}_\nu \) coordinates correspond to Fourier components of the \( \text{H} \text{i} \) signal, and \( \nu \) refers to the coordinate of fluctuation in position \((r, \nu)\) space.

We now take a FT over frequency, in a step called the delay approximation (Parsons et al. 2012; Liu et al. 2014a), to estimate the 3D 21 cm PS. Delay modes do not map directly to the LoS \( k_\parallel \), but for short baselines, we can approximate them to be the same.

\[
\tilde{V}(\bm{u}, \eta) = \int d\nu V(\bm{u}_\nu, \nu) e^{2\pi i \eta \nu}
\]

where the frequency dependence of the baseline has been suppressed through integration. Here, \( \eta \), the conjugate variable of \( \nu \) (sometimes written as \( \tau \)), is the total delay between antenna pairs, such that

\[
\eta = \tau_g - \tau_i = \frac{\bm{b} \cdot \hat{s}}{c} - \tau_i
\]

with \( \tau_i = 0 \) in most cases. This provides a mapping to celestial position.

Furthermore, the baseline vector can be expressed as \( \bm{u}_\nu = \bm{u}_0 \nu / \nu_0 \), where we take \( \nu_0 \) as a fixed frequency within the observed bandwidth (usually taken to be the central frequency). We take \( \bm{u} = \bm{u}_0 \) on the LHS of Equation 2.16. As seen in Section 2.1.3.1, the delay transformation step also helps to isolate the impact of foregrounds and to obtain a Fourier region dominated by the \( \text{H} \text{i} \) signal: the ‘EoR window’.

Following further manipulation and simplifications (see Paul et al. 2016), the autocorrelation of \( \tilde{V}(\bm{u}, \eta) \) can be shown to reduce to

\[
\tilde{V}^2(\bm{u}, \eta) = \langle \tilde{V}(\bm{u}, \eta) \tilde{V}^*(\bm{u}', \eta) \rangle = \iiint d\nu d\nu' \, I^2 \int \frac{d^3 k}{(2\pi)^3} P_{\text{H} \text{i}}(\bm{k})
\]

\[
\times \int dl dm \, A(\hat{s}, \nu) \exp \left[ 2\pi i \left( \bm{u}_\nu - \frac{\bm{k} \cdot \bar{r}_\nu}{2\pi} \right) \cdot \hat{s} \right]
\]

\[
\times \int dl' dm' \, A(\hat{s}', \nu) \exp \left[ -2\pi i \left( \bm{u}_\nu' - \frac{\bm{k} \cdot \bar{r}_\nu}{2\pi} \right) \cdot \hat{s}' \right]
\]

\[
\times \int dx \exp \left[ 2i x \left( 2\pi \eta + k_\parallel d\nu / d\nu' \right) \right]
\]

where \( x = (\nu' - \nu) / 2 \). The dominant contribution to the integral over \( x \) comes from \( \eta = k_\parallel d\nu / d\nu' / 2\pi \), which establishes the correlation scale in LoS direction. For a sufficiently narrow bandwidth, the variation of frequency dependence of the integral over \( dl dm \) and \( dl' dm' \) is expected to be small, so these integrals can be computed at some central frequency \( \nu_0 \) that lies within the bandwidth. If the frequency dependence of the primary beam \( A(\hat{s}, \nu) \) and the background intensity \( \bar{I}_s \) are neglected, the integral over \( \nu \) is trivial. This is a good approximation for interferometers such as HERA since the variation of frequency dependence of integrals over
Radio interferometry

2.1. Radio interferometry

$dl dm$ is expected to be slight for the bandwidths usually chosen for PS estimation, which are relatively small ($\sim 10$ MHz, since we are also limited by cosmic variance).

From Equations 2.15 and 2.18, several important properties of the $\text{H}_\text{i}$ signal can be inferred. Firstly, it can be shown that the mappings between the variables encountered are (Morales & Hewitt 2004)

$$\theta_x = \frac{r_x}{r_y}, \quad u = \frac{k_x r_y}{2\pi}$$
$$\theta_x = \frac{r_x}{r_y}, \quad v = \frac{k_y r_y}{2\pi}$$

$$\Delta \nu \approx \frac{\nu_{21} H(z)}{c (1 + z)^2} \Delta r_z, \quad \eta \approx \frac{c (1 + z)^2}{2\pi \nu_{21} H(z)} k_{||}$$

where $\{r_x, r_y\}$ are the comoving distances in the $\{x, y\}$ directions, $\{k_x, k_y\} = \{k_{\perp 1}, k_{\perp 2}\}$, $\Delta \nu = \nu' - \nu$ is the bandwidth, $\Delta r_z$ is the comoving depth along the LoS, and $H(z)$ is the Hubble parameter as defined in Equation 1.36. The frequency-dependent quantities in Equation 2.19 are to be evaluated at $\nu_0$. It is also found that correlations in $k_{\perp}$ and $k_{||}$ are nearly separable, which can be exploited to avoid the majority of the foreground contamination located at low $k_{||}/|k|$ (see Section 2.1.3.1).

Furthermore, Equations 2.18 and 2.19 allow us to simplify the relation between $\tilde{V}^2(u, \eta)$ and $P_{\text{H}_\text{i}}(k)$. Solving Equation 2.18, in the limit defined by Equation 2.19, gives (Paul et al. 2016)

$$\tilde{V}^2(u, \eta) = \frac{\tilde{I}(\nu) \Omega \Delta \nu}{r_y^2 \frac{dr_y}{d\nu}} P_{\text{H}_\text{i}}(k)$$

where $\Omega = \lambda^2 / A_{\text{eff}}$ is the primary beam solid angle, with $A_{\text{eff}}$ the effective area of the antenna. Here, $r_y$ and $dr_y/d\nu$ play the roles of conversion factors from angle and frequency to comoving distance. Since $\tilde{I}(\nu) = 2k_B T_b / \lambda^2$, we can express the $\text{H}_\text{i}$ signal as $T_b^2 P_{\text{H}_\text{i}}(k) = P_{\text{H}_\text{i}}^T(k)$ in units of $[\text{mK}^2 (h^{-1} \text{Mpc})^3]$. Hence, we obtain

$$\tilde{V}^2(u, \eta) = \left(\frac{2k_B}{\lambda^2}\right)^2 \frac{\Omega \Delta \nu}{r_y^2 \frac{dr_y}{d\nu}} P_{\text{H}_\text{i}}^T(k)$$

or alternatively

$$\tilde{V}^2(u, \eta) = \left(\frac{2k_B}{\lambda^2}\right)^2 \frac{\Omega \Delta \nu}{r_y^2 \frac{dr_y}{d\nu}} \frac{2\pi^2}{k^3} \Delta_{\text{H}_\text{i}}^2(k)$$

where we work with the dimensionless PS $\Delta_{\text{H}_\text{i}}^2(k) \equiv \frac{k^3}{2\pi} P_{\text{H}_\text{i}}^T(k)$ (assuming $P_{\text{H}_\text{i}}^T(k)$ to be isotropic), which, despite the name, has dimensions $[\text{mK}^2]$. In the literature we often see the cosmological conversion factors $r_y$ and $dr_y/d\nu$ written as $X$ and $Y$, respectively.

More generally, this can be recast into the quadratic estimator formalism that is used in the HERA PS estimation:

$$\tilde{\beta}_{\alpha} = M_{\alpha} x_1^\dagger R^\dagger Q_{\alpha} R x_2$$

(2.23)
where $\hat{q}_\alpha$ is the estimate of the $\alpha$ bandpower, $\mathbf{x}_i$ are the visibility vectors that are to be cross-multiplied to form PS (as to eliminate noise bias), $\mathbf{R}$ is the weighing matrix (in HERA, we set this as a diagonal matrix that applies a Blackman-Harris taper [see Appendix A.3; I use the Blackman-Harris window for all PS computations in this thesis] and also propagates data flags), $\mathbf{Q}_\alpha = \mathbf{c}_\alpha^\dagger \mathbf{c}_\alpha$ is the mapping from visibility to PS space, with $\mathbf{c}_\alpha$ the discrete Fourier transform (DFT) operator, and $\mathbf{M}_\alpha$ is the normalization matrix (this is set to be a diagonal matrix given by the conversion coefficients of Equation 2.21). See e.g. Liu & Tegmark (2011) for application of the quadratic estimator in 21 cm cosmology and discussion of the forms of the matrices in Equation 2.23.

2.1.2 Sensitivity

2.1.2.1 Single-dish telescope

The sensitivity of a radio telescope is quantified by the comparison between the signal output of the antenna and the system noise; both of these are expressed as temperatures $T_a$ and $T_{sys}$, respectively, where the temperatures are defined as those of matched resistive loads that would produce equal power levels ($P = k_B T \Delta \nu$ for the resistor). The system temperature includes contributions from the telescope, the receiver system, and the sky, with the latter dominating in most cases at low frequencies. For a single dish, the root mean square (RMS) noise fluctuations for an unresolved source are given by the radiometer equation (Wilson et al. 2013):

$$\Delta T_N = \frac{T_{sys}}{\sqrt{\Delta \nu t_{int}}}$$

which decreases with increased bandwidth $\Delta \nu$ and integration time $t_{int}$. We have assumed $T_{sys} \gg T_a$, which is the usual case in radio astronomy. Following Furlanetto et al. (2006), we can approximate $T_{sys} \approx T_{sky}$ by the sky brightness temperature, which, as a general ‘rule of thumb’, is given by

$$T_{sky} \approx 180 \text{ K} \left( \frac{\nu}{180 \text{ MHz}} \right)^{-2.6}$$

for quiet portions of the sky; more elaborate models exist that are used for rigorous foreground subtraction. The single dish telescope noise can, thus, be estimated to be

$$\Delta T_{N|sd} \approx 0.6 \text{ mK} \left( \frac{1 + z}{10} \right)^{2.6} \left( \frac{100 \text{ hr}}{\Delta \nu t_{int}} \right)^{\frac{1}{2}}$$

The mean 21 cm signal (relative to the CMB, see Equation 1.39) has $\delta T_b \sim 20 \text{ mK}$, meaning that single-dish telescopes should easily have the required sensitivity for a detection of the global 21 cm signal (see Section 1.3.3.3). The challenge, however, is to differentiate the foregrounds from this cosmological signal.
2.1.2.2 Interferometric array

In the case of interferometers, the total noise across the array includes an additional array filling factor $\eta_f = A_{\text{tot}}/D_{\text{max}}^2$, where $A_{\text{tot}}$ is the total effective collecting area of the array and $D_{\text{max}}$ is the longest baseline of the array (thus setting its resolution), such that (Furlanetto et al. 2006)

$$\Delta T^N = \frac{T_{\text{sys}}}{\eta_f \sqrt{\Delta \nu t_{\text{int}}}}$$

(2.27)

Further assuming $T_{\text{sys}} \approx T_{\text{sky}}$ from Equation 2.25 and inserting aspired values for a radio interferometer, we obtain

$$\Delta T^N_{\text{ia}} \sim 2 \text{ mK} \left( \frac{A_{\text{tot}}}{10^5 \text{ m}^2} \right) \left( \frac{10'}{\Delta \theta} \right)^2 \left( \frac{1 + z}{10} \right)^{4.6} \left( \frac{\text{MHz} \ 100 \text{ hr}}{\Delta \nu t_{\text{int}}} \right)^{\frac{1}{2}}$$

(2.28)

where $\Delta \theta = \lambda/D_{\text{max}}$ is the angular resolution scale ($\sim 10'$ for EoR experiments), and the bandwidth $\Delta \nu \sim 1$ MHz corresponds to a comoving distance $\sim 20$ Mpc. The current generation of telescopes, HERA included, have $A_{\text{tot}} < 10^5 \text{ m}^2$, so imaging (i.e. mapping pixels with S/N $\gg 1$) is only possible on coarse scales that exceed those of typical H ii bubbles that are seen during most of reionization. Hence, the focus of forthcoming imaging experiments is on mapping the largest of H ii bubbles created during the EoR, such as those generated by the most luminous quasars of that epoch; these giant H ii regions will also have sharp contrast on either side of the ionization front (Wyithe & Loeb 2004). We note that Equation 2.28 does not include antenna configuration and should, therefore, only be used as a rough guide.

As we are interested in the PS of the 21 cm signal, it is imperative to find the noise associated with it. We start by considering the noise contribution to the RMS amplitude of the delay transformed visibility (Parsons et al. 2012):

$$\bar{V}_N = \frac{2k_B}{\lambda^2} T_{N,\text{RMS}} \Omega \Delta \nu$$

(2.29)

where $T_{N,\text{RMS}}$ is the detector RMS noise for a single visibility measurement, which is closely related to Equation 2.24, and is given by (Furlanetto et al. 2006)

$$T_{N,\text{RMS}} = \frac{\lambda^2 T_{\text{sys}}}{A_e \sqrt{\Delta \nu t_{\text{int}}}}$$

(2.30)

and $A_e$ is the collecting area of each antenna element.

We then substitute the visibility from Equation 2.22 into Equation 2.29 to get the noise contribution to the dimensionless PS, and use Equation 2.30 to get the sensitivity to any one mode of the dimensionless PS in terms of $T_{\text{sys}}$ (Parsons et al. 2012; Pober et al. 2013a):

$$\Delta^2 N(k) \approx \left( \frac{r^2}{2\pi^2 2t_{\text{int}}} \right) \frac{\Omega}{\Delta \nu} k^3 T_{\text{sys}}^2$$

(2.31)

More precisely, an angular scale $\theta_D$ corresponds to a comoving distance $2.7(\theta_D/1')[(1 + z)/10]^{0.2}$ Mpc, and a bandwidth corresponds to $1.8(\Delta \nu/0.1 \text{ MHz})[(1 + z)/10]^{1/2}$ Mpc (Furlanetto et al. 2006).
where $\Omega$ is the solid angle of the primary beam of one element in steradians,$^d$ $t_{\text{int}}$ is the integration time for sampling a particular $k$-mode, $T_{\text{sys}}$ is the system temperature and we have used $T_{\text{sys}}^2 \equiv T_{N,\text{RMS}}^2 \Delta v t_{\text{int}}$, and the factor of two in the denominator arises from the inclusion of two orthogonal polarizations to measure the total unpolarized signal. The cosmological scalar that converts observed bandwidths and solid angles to cosmological distances is also given by (Furlanetto et al. 2006)

$$r^2 \frac{dr_\nu}{d\nu} \approx 540 \left( \frac{1 + z}{10} \right)^{0.9} (h^{-1} \text{Mpc})^3 \text{sr}^{-1} \text{Hz}^{-1}$$

Moreover, from substituting $r^2 \frac{dr_\nu}{d\nu}$ at 150 MHz ($z = 8.5$) into Equation 2.31 and by combining independent $k$-mode measurements in a redundant array (i.e. multiple LoS, time and $u$-$v$ samples) it can be shown that the total sensitivity of the dimensionless PS, with fiducial values in line with the PAPER experiment (Parsons et al. 2010), is given by (Parsons et al. 2012)

$$\Delta^2_{N,H_1}(k) \approx 60 \text{mK}^2 \left[ \frac{k}{0.1 \text{ Mpc}^{-1}} \right]^{2} \left[ \frac{6 \text{ MHz}}{\Delta v} \right]^{2} \left[ \frac{1}{\Delta \ln(k)} \right]^{2} \left[ \frac{\Omega}{0.76 \text{ str}} \right] \left[ \frac{T_{\text{sys}}}{500 \text{ K}} \right]^{2} \left[ \frac{6 \text{ h}}{t_{\text{per,day}}} \right]^{2} \left[ \frac{120 \text{ days}}{t_{\text{days}}} \right] \left[ \frac{32}{N} \right] \left[ \frac{10^4 f_0}{f} \right]^{2}$$

where 120 days of observation with a baseline length $|\mathbf{b}| = 20$ (wavelengths) that allows 13 min of integration per day for a total integration time of $9 \times 10^4$ s per $(u, v, \eta)$-mode is assumed. Here, $\Delta \ln(k)$ is the logarithmic bin size for LoS modes, $t_{\text{per,day}}$ is the amount of time that a baseline samples a $(u, v, \eta)$-mode (as it is limited by the timescale for Earth rotation to move the sampling of a baseline a distance of $\Omega^{-1/2}$ in the $u$-$v$ plane: $t_{\text{per,day}} \approx 1/\sqrt{\Omega} \omega_0 |\mathbf{b}|$, where $\omega_0$ is the angular speed of the Earth’s rotation), and sampling redundancy is captured by $f/f_0$, where $f_0$ is the sampling redundancy of a single baseline with a 1 s integration, and the ratio $f/f_0$ measures the increase in sensitivity for a redundant array over one in which there is no sampling redundancy.

The ‘redundancy boost’ factor, encapsulated by the last term of Equation 2.33, can be exploited to increase the sensitivity of the array. A compact redundant hexagonal configuration, employed by HERA (see Figure 2.4), has been shown to greatly reduce noise uncertainty and is, therefore, optimized for PS measurements; it has been shown that for a fixed number of elements, redundant hexagonal arrays provide about an order of magnitude improvement over non-redundant imaging arrays (Parsons et al. 2012; DeBoer et al. 2017). However, maximum-redundancy array configurations run directly opposite to the needs of imaging and can

---

$d$It was shown in Parsons et al. (2014) that the power-squared beam should enter as a normalizing factor, such that $\Omega' \equiv \Omega^2_{N}/\Omega_{pp}$ should be used instead of $\Omega$ in Equation 2.31, where $\Omega_N$ is the solid angle of the power primary beam and $\Omega_{pp}$ is that of the square of the power primary beam. Here, $\Omega$ is used for simplicity.
2.1. Radio interferometry

severely limit the resolution, unless outriggers are included. Moreover, minimum-redundancy configurations are best for characterizing foregrounds, an indispensable task required to detect the 21 cm signal.

We further note that the sensitivity can be improved by increasing the collecting area per element, although this comes at the cost of extending the shortest baselines: the 21 cm signal is expected to be a diffuse background with most of its power concentrated on large scales; therefore, the instrument is required to have short baselines to be sensitive to this signal. Longer core baselines would decrease the sensitivity to the EoR signal.

2.1.3 Foreground mitigation

In addition to observational sensitivity requirements, foreground contamination from astrophysical effects must also be accounted for to make a positive detection of the redshifted 21 cm line. In the relevant frequency range (~ 50–250 MHz), astrophysical foreground emission originates from sources including galactic diffuse synchrotron emission (~ 70% of the total intensity emission, at 150 MHz), extragalactic sources (~ 27%) and free-free emission (~ 1%) (Shaver et al. 1999).

The brightness temperature of foreground components is expected to be up to six orders of magnitude greater than theoretical expectations for the amplitude of the cosmological EoR signal. However, foregrounds have different and predictable spatial and spectral structures. They are expected to be spectrally smooth, while the cosmological EoR signal is expected to fluctuate rapidly with redshift (Oh & Mack 2003). Therefore, the PS of foregrounds will be distinct, and therefore separable.

Addressing foreground contamination to estimate the EoR signal is a great challenge, with different experiments employing different strategies to achieve a first detection. The two categories of strategies to mitigate foreground contamination are removal versus avoidance, which both rely on exploiting the spectral smoothness of the foregrounds relative to the EoR signal.

Foreground removal strategies rely on the smoothness of foregrounds to fit a set of basis functions to the total emission. The aim is to optimize the set of basis vectors that best differentiate between foregrounds and the 21 cm signal. Foreground emission can be isolated with the chosen basis and then subtracted from the overall signal, leaving the 21 cm signal plus thermal noise, which can be integrated down with increased observation time (Sims et al. 2016).

On the other hand, avoidance involves only working in the region of 3D cosmological k-space with minimal contamination by foreground emission: the ‘EoR window’. This is the strategy employed at HERA. I next discuss this foreground-free Fourier region in Section 2.1.3.1.
Chapter 2. Radio interferometry and HERA

2.1.3.1 The EoR window and the wedge

The PS is characterized by the magnitude of the wavevector $k$, which can be split into components parallel and perpendicular to the LoS, such that $k = k_{\perp} + k_{\parallel}\hat{z}$, where $|k_{\perp}| = \sqrt{k_x^2 + k_y^2}$ and $k_{\parallel} = k_z$. Recent observations by current EoR experiments (e.g. Pober et al. 2013b; Dillon et al. 2014), together with theoretical calculations and simulations (e.g. Datta et al. 2010; Liu et al. 2014a,b), have shown that foreground contamination of the EoR 21 cm signal is localized to a wedge-shaped region in cylindrical $k$-space, at low values of $\mu = k_{\parallel}/|k|$, leaving an ‘EoR window’ at high $\mu$ (Figure 2.2). This is a direct consequence of the spectral smoothness of the foregrounds, combined with instrument chromaticity, that results in ‘mode-mixing’ where power leaks from angular to frequency scales. Power is pushed upwards (to higher $k_{\parallel}$) from the foreground area in the window, with the effect increasing with $k_{\perp}$.

The EoR window is, thus, a region in Fourier space that is a priori expected to be foreground free, where the signal can be observed with minimal contamination (see Figure 2.2), provided the foregrounds are spectrally smooth. The existence of the EoR window, hence, provides a robust foreground avoidance strategy (in staying outside of the wedge), whence a detection of the 21 cm PS can be made. However, should further integrations reveal the existence of low-level unsmooth foregrounds, their influence will leak beyond the ‘edge’ of the wedge (Parsons et al. 2014).

Ensuring that measurements are contained within the EoR window obligates one to work at higher $|k|$. This is problematic because the ratio of the 21 cm signal to instrumental noise peaks at low $|k|$. Were it possible to work within the wedge, the significance of a potential detection of the 21 cm PS would be increased. It is shown in Pober et al. (2014) that working within the wedge can greatly increase the significance of such a detection, anywhere from a factor of 2–6 (depending on array configuration), with equivalent improvements in the errors of inferred astrophysical parameters. Moreover, avoiding the wedge altogether neglects the rigorous examination of the error statistics associated with this region. For these reasons, a more thorough analysis of the wedge is necessary.

Given the high payoff associated with working within the wedge, it is advantageous to statistically characterize the wedge in order to extract the useful data from this region of $k$-space that can be used towards making a significant 21 cm PS detection (e.g. Liu et al. 2014a,b). Ultimately, a rigorous Bayesian framework could encompass all such a priori information without having to reject any region of $k$-space (see e.g. Sims et al. 2019; Sims & Pober 2019).
2.1. Radio interferometry

The EoR Window and Foreground Wedge

![Diagram showing the Fourier region accessible to radio interferometers, the EoR window, in cylindrical \((k_\perp, k_\parallel)\) coordinates. The EoR window is a region of Fourier space with relatively low noise and foregrounds, and is thought to present the best opportunity for measuring the cosmological 21 cm PS during the EoR. At the lowest \(k_\perp\), errors increase because of limits on an instrument’s FoV. High \(k_\perp\) modes are probed by the longest baselines, and the sensitivity drops to zero beyond \(k_\perp\) scales corresponding to these baselines. Spectral resolution limits the sensitivity at large \(k_\parallel\). The lowest \(k_\parallel\) are in principle limited by cosmic variance, but in practice, of larger concern are the limits set by the bandwidth and foreground contamination that also reside at low \(k_\parallel\). At higher \(k_\perp\), however, the foregrounds leak out to higher \(k_\parallel\) in a characteristic shape known as the foreground wedge. The remaining parts of the Fourier plane are thermal-noise dominated, allowing (with large \(A_{\text{tot}}\) and \(t_{\text{int}}\)) a ‘clean’ measurement of the PS in this EoR window. Reprinted from Liu & Shaw (2020).

2.1.4 EoR experiments

Exploring the cosmic dawn with the 21 cm line has the potential to provide great insight into structure formation and the reheating of the IGM, as well as deepening our knowledge of the fundamental cosmology of the Universe. Observing the highly redshifted 21 cm line of \(\text{H}_1\) is a powerful probe of the EoR, as it can access a large portion of the observable Universe. Measuring the 21 cm signal will set important constraints on reionization and refine limits on cosmological parameters.

The rewards offered by studying the 21 cm spectral line have renewed interest in low-frequency radio arrays, with the main objective being detecting the signal during the EoR. Current experiments can be split into two categories: large radio interferometers and smaller-
scale experiments consisting of single dipole antennas or a small handful of elements. The
former seek to measure the PS of 21 cm brightness temperature fluctuations; such experiments
include: the Murchison Widefield Array (MWA) (Tingay et al. 2013), the Precision Array for
Probing the Epoch of Reionization (PAPER) (Parsons et al. 2010), the Low Frequency Array
(LOFAR) (van Haarlem et al. 2013), the Giant Metrewave Radio Telescope (GMRT) (Paciga et al. 2013), the Canadian Hydrogen Intensity Mapping Experiment (CHIME) (Bandura et al. 2014). The latter are designed to characterize the mean 21 cm (global) signal with redshift;
these include the Experiment to Detect the Global EoR Signature (EDGES) (Bowman & Rogers
2010), the Large-Aperture Experiment to Detect the Dark Ages (LEDA) (Greenhill & Bernardi
2012), the Shaped Antenna Measurement of the Background Radio Spectrum (SARAS) (Patra et al. 2013), the Sonda Cosmológica de las Islas para la Detección de Hidrógeno Neutro (SCI-HI) (Voytek et al. 2014), the Zero-spacing Interferometer Measurements of the Background Radio Spectrum (ZEBRA) (Mahesh et al. 2014), the Broadband Instrument for Global Hydrogen
Reionisation Signal (BIGHORNS) (Sokolowski et al. 2015) and the Radio Experiment for the
Analysis of Cosmic Hydrogen (REACH) (de Lera Acedo et al. 2022).

While the 21 cm signal has yet to be detected, limits on the timings (i.e. beginning and end)
and duration of reionization, as well as the 21 cm PS, have started to appear. Second-generation
instruments, such as the Square Kilometre Array (SKA) (Mellema et al. 2013) and the Hydrogen
Epoch of Reionization Array (HERA) (Pober et al. 2014) are forecast to make a first detection
of this signal. The large collecting areas and high S/N should enable these instruments to fully
constrain the 21 cm PS; the SKA also has the potential to image the H\textsc{i} distribution at these
early times.

2.2 HERA

The Hydrogen Epoch of Reionization Array (HERA)\textsuperscript{e} (DeBoer et al. 2017) is an international experiment to measure the 21 cm emission from the primordial IGM during the EoR (6 \(\leq z \leq 13\)) and to also explore earlier epochs of the cosmic dawn (up to \(z \sim 30\)). See Figure 2.3 for an
illustration of the cosmic evolution of the Universe and eras that HERA is able to access.

HERA is designed to measure and characterize the evolution of the 21 cm PS, induced by
brightness temperature fluctuations during and before the EoR due to the heating and ionization
of the IGM by the first stars and BHs (see Section 1.3.3). Determining the 21 cm PS will
constrain the timings and morphology of reionization, the properties of the first luminous
sources (stars, BHs, galaxies), the mechanism behind large-scale structure formation, and help
answer many other astrophysical and cosmological questions.

\textsuperscript{e}http://reionization.org/
Probing the Early Universe with HERA

<table>
<thead>
<tr>
<th>Dark Ages</th>
<th>Epoch of Reionization</th>
<th>Modern Universe</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z=1100$</td>
<td>$z=10.4$</td>
<td>Now</td>
</tr>
<tr>
<td>CMB</td>
<td>400 kyr</td>
<td></td>
</tr>
<tr>
<td>400 kyr</td>
<td>164 Myr</td>
<td>1.1 Gyr</td>
</tr>
<tr>
<td>$z=27$</td>
<td>50 MHz</td>
<td></td>
</tr>
<tr>
<td>$z=13$</td>
<td>100 MHz</td>
<td>225 MHz</td>
</tr>
<tr>
<td>$z=6$</td>
<td>200 MHz</td>
<td>1 Gyr</td>
</tr>
<tr>
<td>$z=5$</td>
<td>225 MHz</td>
<td>1.1 Gyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Figure 2.3: Illustration of the evolutionary period of the Universe probed by HERA. The central redshifts of the two bands used for the PS estimation of the H1C_IDR2.2 dataset (results of which are published in The HERA Collaboration et al. 2022c) are also shown; these are also the bands used for much of the analysis in this thesis. Background image credit: Loeb (2006).](image)

2.2.1 Array design

The full HERA instrument is a 350-element interferometer located in the Karoo reserve in South Africa, consisting of 14 m fixed, zenith pointing parabolic dishes that are set to observe from $50 \text{ MHz} < \nu < 250 \text{ MHz}$. The array is made up of 320 antennas that are concentrated in a dense hexagonal core and 30 outriggers (see Figure 2.4 for HERA’s antenna configuration). Previous experiments, such as PAPER (Parsons et al. 2010), have seen success in focusing on a foreground avoidance approach. This has influenced the design of HERA, which is built to maximize collecting area outside of the wedge (see Section 2.1.3.1) by filling the $u$-$v$ plane with short baselines. The highly redundant configuration is to have increased sensitivity in PS estimation (Parsons et al. 2012) and to exploit redundant baseline calibration (Liu et al. 2010). The outrigger antennas assist in imaging and foreground characterization, but do not contribute significantly to HERA’s sensitivity (which is predominantly determined from the short core baselines). See Table 2.1 for HERA design parameters and respective observational consequences.

Older low-frequency arrays also tasked with measuring the EoR have struggled with the inherent challenge of meeting the strict sensitivity requirements while also suppressing the
foregrounds that are $\sim 5 - 6$ orders of magnitude brighter than the 21 cm signal (Bernardi et al. 2009; Pober et al. 2013b; Dillon et al. 2014). HERA, with its highly compact and redundant configuration (of the array’s 61,075 baselines, only 6610 are unique; the ratio of baselines to unique baselines triples if only the core antennas are considered), improves on these predecessors by bringing more sensitivity on the angular and spectral scales where the EoR 21 cm PS is predicted to dominate over foregrounds. HERA is optimized to deliver high S/N measurements of redshifted 21 cm emission, and its highly redundant compact configuration is optimized for robust PS detection.

The strip observed by HERA is shown on top of the 150 MHz sky in Figure 2.5, with the map showing the diffuse foreground galactic emission.

2.2.1.1 Observing seasons

HERA is a staged experiment that has been in operation while also being under construction. As of the time of submission, the 350 dishes in the array are erected, but commissioning activities are still underway to make all antennas functional.

The HERA observing seasons are during the southern summer, and observation takes place at night when both the Sun and the Galactic Centre (GC) are below the horizon. The science
Table 2.1: HERA design specifications, as detailed in DeBoer et al. (2017). All angular scales are computed at 150 MHz. The predicted foreground avoidance and modelling S/N figures in the bottom row is for an EoR model with 50% ionization at $z = 9.5$, with 1080 h observation, integrated over $\Delta z = 0.8$. I abbreviate baseline to bl., maximum to max. and synthesized to synth.

<table>
<thead>
<tr>
<th>Instrument design specification</th>
<th>Observational performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Element diameter</td>
<td>14 m</td>
</tr>
<tr>
<td>Minimum bl</td>
<td>14.6 m</td>
</tr>
<tr>
<td>Max. core bl</td>
<td>292 m</td>
</tr>
<tr>
<td>Max. outrigger bl</td>
<td>867 m</td>
</tr>
<tr>
<td>EoR frequency band</td>
<td>100 – 200 MHz</td>
</tr>
<tr>
<td>Extended range</td>
<td>50 – 250 MHz</td>
</tr>
<tr>
<td>Frequency resolution</td>
<td>97.8 kHz</td>
</tr>
<tr>
<td>Survey area</td>
<td>$\sim 1440 , \text{deg}^2$</td>
</tr>
<tr>
<td>$T_{\text{sys}}$</td>
<td>$100 + T_{\text{sky}} , \text{K}$</td>
</tr>
<tr>
<td>Collecting area</td>
<td>54,000 m$^2$</td>
</tr>
<tr>
<td>FoV</td>
<td>9$^\circ$</td>
</tr>
<tr>
<td>Largest scale</td>
<td>7.8$^\circ$</td>
</tr>
<tr>
<td>Core synth. beam</td>
<td>25$'$</td>
</tr>
<tr>
<td>Outrigger synth. beam</td>
<td>11$'$</td>
</tr>
<tr>
<td>Redshift range</td>
<td>$6.1 &lt; z &lt; 13.2$</td>
</tr>
<tr>
<td>Redshift range</td>
<td>$4.7 &lt; z &lt; 27.4$</td>
</tr>
<tr>
<td>LoS comoving res</td>
<td>1.7 $\text{Mpc}$ (at $z = 8.5$)</td>
</tr>
<tr>
<td>Comoving survey vol</td>
<td>$\sim 150 , \text{Gpc}^3$</td>
</tr>
<tr>
<td>Sensitivity after 100 h</td>
<td>50 $\mu\text{Jy beam}^{-1}$</td>
</tr>
<tr>
<td>Foreground avoidance</td>
<td>23$\sigma$</td>
</tr>
<tr>
<td>Foreground modelling</td>
<td>91$\sigma$</td>
</tr>
</tbody>
</table>

Worthy observing seasons are H1C (2017–2018), H3C (2019–2020), H4C (2020–2021) and H5C (2021–2022), with each season having an increased number of antennas. H0C (2016–2017) consists of preparatory observations taken with the 19 element commissioning array that recycled the dipole feeds from PAPER, and H2C (2018–2019) did not have enough usable data.

Phase I of HERA (H1C and prior) reused much of the receiving system from the PAPER experiment (i.e. the feeds and correlator). The array was upgraded in Phase II and new Vivaldi feeds installed that extend the bandwidth to 50–250 MHz (Fagnoni et al. 2021a).

For the H3C season that ran from October 2019 to April 2020, I took part in 4 observing weeks, where I monitored the outputs from HERA to assist with the commissioning and data collection of the interferometer. My responsibilities included writing nightly and weekly observation reports, logging recurring issues with the telescope, flagging faulty antennas, as well as checking diagnostics and the status of other affiliated processes essential to observation.

2.2.2 Traditional calibration, deconvolution and imaging

To better understand and visualize the radio interferometric data outputted by HERA, I worked with visibilities from data releases observed during the H0C and H1C seasons to get calibrated and cleaned (see Section 2.2.2.2) visibilities and images. While imaging is not the primary aim of HERA, it is indicative of the quality of both the data and the reduction steps, which are important for PS estimation. Furthermore, the gain solutions found from gain, delay and bandpass calibration, when a bright point source is transiting (e.g. the GC or Fornax A), can
then be applied to the other datasets with the same Julian date (JD), since gain solutions are stable over an evening of observation (Carilli et al. 2017).

In this subsection, I outline the principal conventional calibration and deconvolution routines for imaging. The Common Astronomy Software Applications (casa) package (McMullin et al. 2007) is used for such routines. I then present the reduction of preliminary HERA data to demonstrate the interferometer’s (early) imaging capabilities.

### 2.2.2.1 Conventional calibration

Calibration is required to remove, insofar as possible, the effects of instrumental variability and atmospheric disturbances (amongst others), which can affect the amplitude and phase of the data. The measurement equation for the observed visibility of antennas $i$ and $j$ can be expressed as (Hamaker et al. 1996)

$$ V_{ij}^{\text{obs}} = J_{ij} V_{ij}^{\text{true}} = g_i g_j^* V_{ij}^{\text{true}} $$ (2.34)

where $J_{ij}$ is a generalized operator characterizing the net effect of the observing process, which includes complex gain, bandpass response, atmospheric and baseline-based effects, to name but a few. The second equality arises from the insight that the measured visibility $V_{ij}^{\text{obs}}$ is formed from a product of the uncorrupted visibility $V_{ij}^{\text{true}}$ with antenna-based gains $g_i g_j^*$. This representation ignores direction-dependent effects (that include primary beam and pointing errors).
In Section 2.2.2.3, I perform traditional calibration with \texttt{casa} on HERA data. In doing so, I invoke delay, gain, and bandpass calibration; these are described below. The \texttt{casa} task is explicitly designated in brackets after each procedure.

**Delay calibration** (\texttt{gaincal} with \texttt{gaintype = ‘K’}): accounts for impurities in the correlator model, such as errors in antenna position and propagation timings that cause deviations in the model that are manifested as a time-constant linear phase slope as a function of frequency, known as a delay, in the correlated data for a single baseline. This routine solves for the antenna-based delays via FT of spectra on baselines involving (only) a reference antenna. Delay calibration is best done by observing a strong unpolarized source.

**Gain calibration** (\texttt{gaincal} with \texttt{gaintype = ‘G’}): solves for temporal variations in the instrument, atmosphere and ionosphere (amongst others), with the latter being especially important at the low frequencies at which HERA operates. Here, we solve for the gain phases, which are taken to be the same across frequencies for each antenna. The phase offset from delay calibration is kept. This can be done with the observation of a strong (preferably point) source every few minutes.

**Bandpass calibration** (\texttt{bandpass}): primarily used to compensate for any complex gain variations in frequency (as well as in time) caused mainly by antenna electronics and optics (e.g. standing waves for the latter). Bandpass calibration handles frequency-dependent effects that vary on timescales much longer than the time-dependent effects handled by calibration with \texttt{gaincal}. Bandpass calibration can be done with the observation of a strong continuum source with a flat spectrum to generate a complex response function for each antenna.

### 2.2.2 \texttt{CLEAN} deconvolution algorithm

The incomplete coverage of the $u$-$v$ plane leads to a complicated synthesized beam with irregular sidelobes, thus, producing a ‘dirty image’. Deconvolution is required to interpolate/extrapolate visibility samples into unsampled regions of the $u$-$v$ plane, with the aim of finding a sensible model for the brightness distribution $T(\hat{s}, \nu)$ that is compatible with the data. The \texttt{CLEAN} deconvolution algorithm, introduced by Högbom (1974), iteratively deconvolves the PSF from the source map, with the a priori assumption that $T(\hat{s}, \nu)$ is a collection of point sources.

Högbom \texttt{CLEAN} proceeds through the following steps:

1) Initialize:
   a) a \texttt{CLEAN} component list to empty
   b) a residual image to the dirty image

2) Locate and measure the peak of the greatest absolute intensity, with position $r_b$ and brightness $D(r_b)$, in the dirty image
3) Subtract the scaled dirty beam \( \gamma D(r_b)B(r - r_b) \) from the dirty image at \( r_b \), where \( \gamma \) is the loop gain (i.e. damping factor; common value \( \sim 0.1 \)), and \( B \) is the PSF.

4) Add this point source location and amplitude to the clean component list.

5) Go to step 2 (hence forming an iteration) unless stopping criterion reached. What is left are the residuals.

Stopping criteria include:

- Residual map maximum < threshold = multiple of RMS (noise limited)
- Residual map maximum < fraction of dirty map maximum (dynamic range limited)
- Maximum number of clean components reached

Steps to create a final clean image:

- Create a model image with all point source clean components
- Convolve the accumulated point source model image with an idealized ‘clean beam’, i.e. an elliptical Gaussian fitted to the main lobe of the PSF
- Add back residuals of the dirty image (see step 5 of the clean procedure) to the clean image

Variations of this algorithm exist, such as Clark clean (Clark 1980), which is used to produce the results in this thesis, as it is faster since only a small fraction of the PSF is used for its residual image updates.

Such deconvolution algorithms are commonly used to fill in missing data in a flagged visibility dataset, even when imaging is not the primary aim.

### 2.2.2.3 Data reduction method

HERA data was accessed through the Librarian\(^f\), which is the HERA archive system where raw, flagged and calibrated datasets in Miriad or uvfits (which derives from HDF5) visibility file format are stored. Note that to enable \texttt{casa} functionality, files had to be imported into \texttt{MeasurementSet} (MS) structure. The data reduction steps undertaken when a strong calibrator source is in the FoV are as follows:

1) Find and flag bad data, due to e.g. radio-frequency interference (RFI), antenna autocorrelations, faulty antennas. This can be done by eye through the inspection of visibility plots. Such manual flagging can, however, be rather onerous as the number of baselines and time integrations becomes substantial (especially if there are per-baseline or per-time effects).

\(^f\)https://github.com/HERA-Team/librarian
2.2. **HERA**

2) Fringe-rotate so that the phase tracking centre is the calibrator (this is convenient for viewing visibilities and images)

3) Create a flat-spectrum point source model for the calibrator

4) Perform initial delay and overall gain calibration using the calibrator point source

5) Apply two cycles of self-calibration with the `clean` algorithm, consisting of:
   a) Imaging the partially calibrated data (with appropriate `clean` box)
   b) Complex bandpass calibration, which is applied to the next cycle

6) Do final imaging using multifrequency synthesis and multiscale `clean`

In particular, I look at the GC and Fornax A. These strong sources can be used to obtain calibration solutions, which can then be copied and applied to data from the same JD, since gains are expected to be stable over such timescales.

### 2.2.2.4 Preliminary calibration and imaging results

I look at data from H1C_IDR2.2 to demonstrate traditional calibration and `cleaning` methods by imaging Fornax A, which transits through the main beam of the telescope at 3.4 h in Local Sidereal Time (LST). The array configuration and its `u-v` sampling for this dataset is shown in Figure 2.6, with antennas flagged by either basic antenna metric considerations or during calibration. In the plots in this section, by HERA-52 I mean the HERA Phase I array during the H1C season, since the array comprised 52 antennas at that stage.

The flagging of raw visibilities is shown in Figure 2.7a. The flagging of RFI (thin vertical lines), autocorrelations (large amplitude curves) and bad antennas can all be seen, with other less evident but, nonetheless, still erroneous data points also flagged; these are all henceforth discounted in the following calibration steps. Applying steps 5–9, which include delay, gain and bandpass calibration, the amplitudes in Figure 2.7b are obtained.

These maps demonstrate not only the early HERA imaging capabilities but also that Fornax A (or any other bright source, such as e.g. the GC) can be used as calibrators. For PS estimation, we require visibilities from a portion of the sky devoid of any strong sources, which cannot accurately be calibrated. Hence, calibrated solutions of Fornax A can be applied to empty-sky visibilities.

In the absolute calibration stage of the HERA analysis pipeline, the sky-based calibration is done on three fields at LSTs 2.0, 5.2 and 14.4 h, on JDs 2458042, 2458116 and 2458207 respectively, such that the field transits near midway through each night. These fields are chosen as they have minimal diffuse foregrounds, and each have a bright point source, catalogued by the GaLactic and Extragalactic All-sky MWA (GLEAM) survey (Hurley-Walker et al. 2017).

---

6 Throughout this thesis, LST refers to local apparent sidereal time.
transiting near the centre of the FoV (this also establishes the flux scale). The absolute calibration is then conducted on 5 min of averaged data at the field transit using the \textsc{casa} calibration routines described in this section. These gains are then copied to all the other times for the respective JDs. The calibrated visibilities are averaged onto an LST grid that yields model visibilities for 20 h of LST. These visibilities are then passed through a low-pass filter to remove high-delay structure not associated with foregrounds. The filter threshold is set to be the baseline horizon delay $\eta_h = |b|/c$ plus a 50 ns buffer. This process also has the added benefit of filling in flagged slices with a best guess of the foreground.

### 2.2.2.5 Per-baseline power spectrum analysis

The full HERA analysis pipeline, as described in Section 2.2.3, is an involved process with many stages that use different calibration methods. At each of these steps, there is potential for the introduction of spurious artefacts to the final PS; a $\approx 1\mu s$ bump in delay space was observed for the resulting H1C\_IDR2.2 calibrated visibilities whose origin was initially unknown. Kern et al. (2019); Fagnoni et al. (2021b) examine the major systematics observed in data from the first phase of the array, and attribute this particular effect to coaxial cable reflections, which can be suppressed through further calibration.

Before this was realized, I examined intermediary visibility products of the HERA pipeline in both the frequency and delay domains to see if any particular calibration step may have intro-
2.2. HERA

(a) Visibility amplitudes for flagged (red) and unflagged raw visibilities.

(b) Gain and bandpass calibrated visibility amplitudes, using the standard `casa` software.

**Figure 2.7:** Visibility amplitudes versus channel number for a 10 s time integration on JD 2458098 at LST 3.40 h.

(a) Gain amplitude versus frequency channel for the final bandpass calibration solution.

(b) Gain phase versus frequency channel for the final bandpass calibration solution.

**Figure 2.8:** Effects of bandpass calibration on the antenna gains, coloured by antenna. These gains are for a 10 s time integration on JD 2458098 at LST 3.40 h.
(a) PSF of the HERA-52 array, showing the main lobe and the edge of the hexagonally spaced side-lobes, owing to the redundant hexagonal interferometer configuration.

(b) Raster map of the final calibrated image using multifrequency synthesis and multiscale CLEAN cycles. The restoring beam is shown at the bottom left.

Figure 2.9: HERA dirty beam (left) and the final Fornax A synthesized image from the reduction steps undertaken in this section (right).

<table>
<thead>
<tr>
<th>Delay [\mu s]</th>
<th>( k_{\parallel} (z = 10.38) ) (Band 1) [( h \text{ Mpc}^{-1} )]</th>
<th>( k_{\parallel} (z = 7.93) ) (Band 2) [( h \text{ Mpc}^{-1} )]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.247</td>
<td>0.279</td>
</tr>
<tr>
<td>1.0</td>
<td>0.495</td>
<td>0.559</td>
</tr>
<tr>
<td>2.0</td>
<td>0.989</td>
<td>1.118</td>
</tr>
<tr>
<td>5.0</td>
<td>2.473</td>
<td>2.794</td>
</tr>
<tr>
<td>( 1 \times 10^6 )</td>
<td>494,600</td>
<td>558,900</td>
</tr>
</tbody>
</table>

Table 2.2: Conversion between delay and \( k_{\parallel} \) for selected delays at \( z = 10.38 \) and \( z = 7.93 \), which are the central redshifts for the frequency bands specified in Table 2.3.

duced this error. I also re-analysed the data in \texttt{casa} using the traditional and well-established calibration steps detailed in Section 2.2.2.3 to ensure transparency and full understanding of how the visibilities have been altered. From these calibrated data, I looked at per-baseline cross-sections of the data to see if any baseline/antenna was more susceptible to exhibit this \( 1 \mu s \) bump; an example plot of this is shown in Figure 2.10.

The cosmological conversion factors \( \Upsilon' \) that transform from delay \( \eta \) to \( k_{\parallel} (z) = \Upsilon'(z) \eta \) (see Equation 2.19) at \( z = 10.38 \) and \( z = 7.93 \) (corresponding to the central redshifts of HERA Bands 1 and 2 in Table 2.3) are 494,600 and 558,900 \( h \text{ Mpc}^{-1} \text{ s}^{-1} \), respectively. A few conversions
are listed in Table 2.2, for reference.

The results for this work were inconclusive, and with the source of the error eventually identified, the focus of my research shifted to developing robust statistical methods for radio interferometric data reduction, which are presented in the following chapters.

2.2.3 Analysis pipeline

The HERA collaboration analyses visibility data with the more sophisticated calibration steps found in the hera_cal\(^b\) package. The pipelines for each data release are documented in the hera_pipelines\(^b\) repository.

Most of my research has focused on improving calibration and PS results for H1C_IDR2.2, which corresponds to the first HERA observing campaign (note that these are zero indexed, with H0C preceding H1C, as explained in Section 2.2.1.1), second internal data release. The full pipeline for this dataset is shown in Figure 2.11. The analysis pipeline with LST-binning can be condensed into the following steps:

---

\(^b\)https://github.com/HERA-Team/hera_cal

\(^b\)https://github.com/HERA-Team/hera_pipelines
1) Start with the raw visibilities and remove the faulty antennas for that observation period identified from commissioning and from earlier rounds of analysis
2) Run redundant calibration
3) Run sky-based absolute calibration
4) Apply RFI flagging
5) Smooth the gain solutions in frequency and time – visibilities are now fully calibrated
6) Aggregate visibilities from different JDs by binning and averaging in LST
   a) A delay filter can also be applied to the final visibilities to further remove foregrounds by subtracting the spectrally smooth part of the data, to ensure that the final product lives outside of the wedge

The effect of a selected few reduction steps on the visibility phases of a few sample 14 m EW baselines can be seen in Figure 2.12.

As the last step in the preprocessing pipeline, we form pseudo-Stokes\textsuperscript{1} polarization visibilities for the final visibility data products, which are defined to be a linear combination of the

\textsuperscript{1}We use ‘pseudo-Stokes’ to refer to these visibilities, which differ from the ‘true’ Stokes parameters that are defined in the image plane (van Straten et al. 2010). These pseudo-Stokes visibilities are approximately the FT of the Stoke visibilities from the image to the \(u\)-\(v\) domain. In the absence of direction-dependent effects after calibration, this approximation becomes an equality.
2.2. HERA

We then work with the pseudo-Stokes I visibilities, namely the Fornax A radio galaxy at LST 3.36 h and the Galactic anticentre at LST ∼ 7.5 h (see Figure 2.5). Two frequency bands are also chosen, which are mostly devoid of strong RFI features and have low flagging rates. These fields and bands are defined in Table 2.3.

instrumental polarization visibilities (Hamaker et al. 1996):

\[
\begin{bmatrix}
V_p^l \\
V_pQ \\
V_pU \\
V_pV
\end{bmatrix} = \frac{1}{2} \begin{bmatrix}
1 & 0 & 0 & 1 \\
1 & 0 & 0 & -1 \\
0 & 1 & 1 & 0 \\
0 & i & -i & 0
\end{bmatrix}
\begin{bmatrix}
V_{EE} \\
V_{EN} \\
V_{NE} \\
V_{NN}
\end{bmatrix}
\]

(2.35)

We then work with the pseudo-Stokes I visibilities \(V_p^l = (V_{EE} + V_{NN})/2\) because we expect the 21 cm signal to be unpolarized (Moore et al. 2013). Here, E and N denote the east and north alignment of the feeds, such that east-east (EE) instrumental polarization corresponds to an east-facing feed correlated with another east-facing feed (likewise for north-north (NN)).

Three main fields and bands were identified for the PS estimation of the H1C_IDR2.2 dataset. These occupy relatively quiescent parts of the sky and are free from strong radio sources, namely the Fornax A radio galaxy at LST 3.36 h and the Galactic anticentre at LST ∼ 7.5 h (see Figure 2.5). Two frequency bands are also chosen, which are mostly devoid of strong RFI features and have low flagging rates. These fields and bands are defined in Table 2.3.
<table>
<thead>
<tr>
<th>Channels</th>
<th>Frequency range</th>
<th>Bandwidth</th>
<th>Central redshift</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band 1</td>
<td>$175 - 334 \text{ MHz}$</td>
<td>$117.09 - 132.62 \text{ MHz}$</td>
<td>$15.53 \text{ MHz}$</td>
</tr>
<tr>
<td>Band 2</td>
<td>$515 - 694 \text{ MHz}$</td>
<td>$150.29 - 167.77 \text{ MHz}$</td>
<td>$17.48 \text{ MHz}$</td>
</tr>
</tbody>
</table>

LSTs Time span

| Field 1 | $1.25 - 2.7 \text{ h}$ | $1.45 \text{ h}$ |
| Field 2 | $4.5 - 6.5 \text{ h}$ | $2 \text{ h}$ |
| Field 3 | $8.5 - 10.75 \text{ h}$ | $2.25 \text{ h}$ |

Table 2.3: Bands and fields use for the PS estimation of H1C_IDR2.2.

### 2.2.3.1 Averaging axes

In the HERA data reduction pipelines, there are three major axes over which visibilities are averaged: JDs, baselines and time. At each stage, there is potential for RFI or other contaminants to corrupt the estimates. While several flagging and outlier detection stages are incorporated in the HERA pipelines, there is still potential for subtle effects to go unnoticed and not be rejected. Robust methods to address this issue are the focus of Chapters 3 and 4. The three principal averaging axes are explained below:

#### Days

Visibilities are coherently averaged over evenings of observation, since visibilities measured at the same LST should be equal. The LST alignment of data from multiple JDs and the subsequent averaging are further detailed in Section 2.2.3.2.

#### Baselines

Near-coherent averaging. A key factor in HERA’s overall sensitivity is its redundant array configuration: redundant baselines can be averaged over, which provides an additional boost in sensitivity by a factor of $\sqrt{N_{\text{bl}}}$, if averaged in the visibility domain. However, despite being computationally more expensive, the redundant averaging of baselines is instead performed in the PS domain through cross power spectra (CPS). This is done for a couple of reasons: firstly, with additional statistical tests occurring at the PS level, this gives more flexibility in rejecting dubious baselines. Secondly, baselines paired with themselves can be rejected, as
these can have more prominent systematics than baselines paired with a different baseline in the redundant group. As an example, consider the averaging of \( N \) baselines in the visibility \( V_{\text{red}} = (V_1 + V_2 + V_3 + \ldots) / N \) and PS domains \( P_{\text{red}} \sim V_{\text{red}}^2 = (V_1^2 + V_2^2 + V_3^2 + \ldots + V_1 V_2 + V_1 V_3 + \ldots) / N^2 \). Products in \( V_i^2 \) can be discarded while those in \( V_i V_j \) retained, for \( i \neq j \); this cannot be done if averaging visibilities directly. Empirically, this approach has shown to improve final PS results. Therefore, all baseline permutations in a redundant set (i.e. sharing the same orientation and length) are cross-multiplied (using adjacent time bins) as the first step in PS computation. These cross-multipliers are then averaged.

**Incoherent averaging.** PS are also then averaged over baselines of the same length (even if they have different orientations).

### Time

**Coherent averaging.** Following the inpainting and systematic subtraction steps in the pre-processing pipeline, visibilities are coherently time averaged over windows spanning 214 s after being rephased to a central pointing centre. This is chosen to be half of the 428 s window that is shown to limit decoherence of the EoR power to an average of \( \sim 1\% \) (Aguirre et al. 2022). Adjacent time bins are then cross-multiplied in CPS computations to further avoid any noise bias.

**Incoherent averaging.** PS are additionally averaged over any remaining time bins in each field following baseline averaging.

#### 2.2.3.2 LST-binning

I expand on the LST-binning stage, as this is central to much of the work done in later chapters.

Following the final calibration and flagging of individual datasets in the HERA analysis pipeline, complex visibilities are aggregated and coherently averaged across JDs. An LST grid with cadence of 21.4 s (double the integration time) is established to account for the sidereal drift between consecutive JDs. In LST-binning, each integration from individual datasets is assigned to the nearest LST bin but is also rephased to account for the slight LST difference between its centre and the bin’s centre. Every LST bin gets two data points for each input night.

In each time bin, a further round of outlier rejection is then performed using median absolute deviation (MAD)-clipping, which rejects samples for every frequency/time/baseline slice that has a modified Z-score \( |Z_{i \text{mod}}| > 5 \), as defined below:

\[
Z_{i \text{mod}} = \frac{x_i - \text{med}(x)}{\sigma_{\text{mad}}} \tag{2.36}
\]

\[
\sigma_{\text{mad}} = 1.4826 \times \text{med} |x - \text{med}(x)| \tag{2.37}
\]
where $\sigma_{\text{mad}}$ is the MAD, a robust measure of variability that isn’t skewed by outliers. The factor of 1.4826 in Equation 2.37 is a consistency correction that is required to reproduce the standard deviation in the case of white Gaussian noise.\(^k\)

The MAD-clipping is done to ensure that any missed RFI and other problems with the data that do not repeat night-to-night at the same LST are flagged. The binned complex visibilities are then mean averaged about their $\Re e$ and $\Im m$ components separately.

### 2.2.4 Latest upper limits on the 21 cm power spectrum

The results from the reduction and PS estimation of the H1C_IDR2.2 data, consisting of 18 nights of observation in December 2017 (combining for $\approx 36$ h of integration) from 39 working antennas, have set the most stringent limits on the 21 cm PS yet, as presented in The HERA Collaboration et al. (2022c). The requirements for such a feat are demanding, with huge amounts of effort going into the building of the array, commissioning and data processing involving hundreds of researchers across dozens of institutions.

The final visibilities for the (12, 13, pl) 14 m EW baseline, following the preprocessing pipeline, are shown in Figure 2.13 for the full LST and frequency range; this exemplifies the thorough calibration, averaging and reconstruction conducted in the visibility domain. The selected bands and fields for PS estimation (see Table 2.3) are also shown in this plot.

Such visibilities from all remaining baselines are pushed through the PS pipeline to finally get the PS estimates for the data release. The principal results are the dimensionless PS shown in Figure 2.14. From these, the 95% upper limits on the 21 cm PS are constrained at $\Delta_{21}^2(k = 0.192 h \text{Mpc}^{-1}, z = 7.9) \leq (30.76)^2 \text{mK}^2$ and $\Delta_{21}^2(k = 0.256 h \text{Mpc}^{-1}, z = 10.4) \leq (95.74)^2 \text{mK}^2$, which are calculated by taking the measured dimensionless PS and adding the $2\sigma$ error bars. These results are for the $\Re e$ component of the PS, since the 21 cm cosmological signal is asserted to be purely real (and positive) in the PS, by definition. The $\Im m$ component of the computed PS is a manifestation of the noise. In Figure 2.14, any negative bandpower is thus set to zero.

These limits are contextualized with respect to the wider field in Figure 2.15, where they are compared to other 21 cm PS experiments. This plot was made with the limits stored in the eor\_limits\(^l\) repository.

The astrophysical implications of the upper limits from The HERA Collaboration et al. (2022c) are quantified using a suite of theoretical models in The HERA Collaboration et al. (2022b); the principal takeaway from this research is that the IGM was heated above the

\(^k\) For normally distributed data, $\text{med}|x - \text{med}(x)|$ (see Equation 2.37) estimates not the standard deviation $\sigma$ but rather $\Phi^{-1}(3/4)\sigma = \sigma/1.4826$, where $\Phi^{-1}$ is the reciprocal of the quantile function for the standard normal distribution, with its argument set to $3/4$ because we want $\pm \sigma_{\text{mad}}$ to cover 50% (between $1/4$ and $3/4$) of $\Phi$.

\(^l\) https://github.com/EoRImaging/eor\_limits
Figure 2.13: Final H1C_IDR2.2 visibility amplitudes and phases for the 14 m EW (12, 13, pl) baseline following all the reduction steps in the analysis and preprocessing pipelines. The band and field combinations used for PS computation are also shown. All flagged frequencies have been inpainted; however, flagged times are left unfilled. The Galactic anticentre is the cause of the elevated visibility amplitude at LST \( \sim 7.5 \) h.
adiabatic cooling threshold by $z \sim 8$. These new upper limits do not confirm nor disprove the EDGES measurement in Bowman et al. (2018), since it observes at lower redshifts.

HERA Phase I observed throughout the southern summer between JDs 2458026 – 2458208 (29th September 2017 – 31st March 2018). The HERA Collaboration et al. (2022c) used data from the best 18 nights of observation, which was packaged in the H1C_IDR2.2 data release. There are, however, 94 usable nights of data from this season that culminate in the H1C_IDR3.2 dataset, which can be reduced to set even deeper limits. This widened dataset is less homogeneous than H1C_IDR2.2 due to different epochs within that period having different numbers of antennas (the array was actively constructed and commissioned throughout the season), and it also has JDs that require significant flagging. Together with other intricacies, this makes the analysis of H1C_IDR3.2 trickier. The new limits will be published soon (currently available as a preprint; The HERA Collaboration et al. 2022a), and are provisionally set to improve on the ones from The HERA Collaboration et al. (2022c) by a factor of $\approx 2$. 

Figure 2.14: Spherically averaged dimensionless PS for Bands 1 and 2, each corresponding to a mean redshift of 10.4 and 7.9, respectively, at each of their three fields. The most sensitive limits, coming from Band 1 Field 1 and Band 2 Field 1, are circled in blue. The theoretical noise floor is also drawn.
2.2. HERA

Figure 2.15: The present state of EoR limits, following the publication of The HERA Collaboration et al. (2022c). At $z = 7.9$, the HERA limits from H1C_IDR2.2 are the most sensitive to date by about an order of magnitude across the entire EoR radio interferometric literature. The solid and dashed lines show the theoretical 21 cm PS.
CHAPTER 3

NORMALITY AND ROBUST LOCATION ESTIMATES
OF HERA VISIBILITIES

HERA, and radio interferometers, more generally, suffer from data corruption due to RFI, systematics and instrumental defects. These may cause the underlying visibility noise to exhibit non-Gaussian behaviour, even after attempts to remove contaminants in the standard pipelines. With the number of usable baselines and evenings of observations ever increasing with HERA as it nears completion, there is increased risk for averaging estimates to be skewed or even invalidated by outliers or other non-Gaussian effects. HERA has also been shown to exhibit non-redundancies in its baselines (Dillon et al. 2020), which may affect the distribution of redundant baselines. As the sensitivity of interferometers improves and the data volumes handled become ever larger, robust multivariate estimates of location, as well as more sophisticated outlier detection algorithms, are increasingly crucial to adequately and thoroughly reduce interferometric data.

This work is also motivated by the anomalous Band 2 Field 2 results in The HERA Collaboration et al. (2022c), which sees a systematic excess in power compared to thermal noise expectation, hypothesized to be due to low-level RFI. This signifies that the current reduction mechanism is not good enough. A robust approach to the averaging of visibilities across JDs has the potential to further remove RFI and improve results.

In this chapter, I examine the normality of visibility distributions aggregated across redundant groups and in LST in Section 3.1, through visualization and statistical methods. With outliers being the common reason for violating the assumption of normality, in Section 3.3...
I present an alternative multivariate outlier detection method based on robust Mahalanobis distances, which wholly considers the complex data rather than treating the \( \Re \)e and \( \Im \)m parts separately. In Section 3.4, I then introduce robust multivariate location estimators and how they can be used to better represent non-Gaussian and contaminated data than standard statistical techniques. I apply the geometric median to the standard LST-binning pipeline to compute visibility estimates across JDs. These results are compared in both the visibility and PS domains, with a tentative new limit on the 21 cm PS established. Lastly, complex multidimensional Gaussian process regression is used in Section 3.5 as a data inpainting method, in contrast to interpolation with \texttt{clean} or other delay filtering techniques.

The work in this chapter is mainly applied to LST-binning (see Section 2.2.3.2), but these techniques can also be used, with slight modification, to the other averaging dimensions in Section 2.2.3.1; this is further discussed in Section 3.6. Unless stated otherwise, the dataset examined is a subset of the Band 2 Field 2 data from fully calibrated LST-binned H1C_IDR2.2 visibilities. I consider LSTs 5.39 – 5.74 h, EE polarization and only look at the 19 short 14 m EW baselines redundant to (12, 13). The resulting data array, therefore, has dimensions (JDs, frequencies, times, baselines) = (36, 180, 60, 19), since each new time bin takes 2 data points from each of the 18 nights of observation.

The content of this chapter includes research presented in Molnar & Nikolic (2021a, 2022). Much of the code used can also be found in the \texttt{robstat}\textsuperscript{a} \texttt{Python} package.

### 3.1 Normality of HERA visibilities

In this section, I first visualize visibility data that is aggregated both/either in LST and/or across redundant baselines to get a qualitative intuition for the data. I then test for the normality of these distributions.

#### 3.1.1 Data visualization

I employ kernel density estimate (KDE) plots to visualize the distribution of redundantly grouped and LST-binned HERA visibilities. For a sample of \( d \)-variate random vectors \( x_1, x_2, \ldots, x_n \) drawn from a common distribution, the KDE is defined to be

\[
\hat{f}(x, H) = \frac{1}{n} \sum_{i=1}^{n} K_H(x-x_i)
\]

where

\[
K_H(u) = |H|^{-\frac{1}{2}} K(H^{-\frac{1}{2}}u)
\]

\textsuperscript{a}https://github.com/matyasmolnar/robstat
3.1. Normality of HERA visibilities

where \( \mathbf{H} \) is the bandwidth matrix, a \( d \times d \) positive-definite and symmetric matrix (with determinant \(|\mathbf{H}|\)) that sets the smoothing, and \( K \) is the kernel function; we use the standard multivariate normal kernel:

\[
K(\mathbf{u}) = \frac{1}{\sqrt{(2\pi)^d}} e^{-\frac{1}{2} \mathbf{u}^\top \mathbf{u}} \tag{3.3}
\]

such that

\[
K_{\mathbf{H}}(\mathbf{u}) = \frac{1}{\sqrt{(2\pi)^d |\mathbf{H}|}} e^{-\frac{1}{2} \mathbf{u}^\top \mathbf{H}^{-1} \mathbf{u}} \tag{3.4}
\]

where \( \mathbf{H} \) takes on the role of the covariance matrix.

The bandwidth matrix \( \mathbf{H} \) is a free parameter that will strongly affect the resulting estimate. Bandwidth selection is important, and misspecification of \( \mathbf{H} \) can result in a distorted representation of the data (see e.g. Jones et al. 1996 for a review). For this work, I employ Scott’s ‘rule of thumb’ (Scott 2015), where the square root of the diagonal entries of the bandwidth matrix are given by

\[
\sqrt{H_{ii}} = n^{-\frac{1}{d}} \sigma_i \tag{3.5}
\]

where \( n \) is the number of data points, \( \sigma_i \) is the standard deviation of the \( i \)th variable. I use the \texttt{kdeplot} function in \texttt{seaborn} to plot KDEs.

I show a scatter plot in Figure 3.1 and its corresponding KDE plot in Figure 3.2, with various bivariate location estimators also marked, for some sample data that is quite typical of the H1C_IDR2.2 dataset. There are up to \( 36 \times 19 = 684 \) data points that can be considered, although a significant fraction are flagged in practice. The geometric mean is given by \((\prod_{i=1}^{n} x_i)^{\frac{1}{n}}\), the Mardia median (Mardia 1972) minimizes the circular mean absolute deviation \( d(\tilde{\theta}) = \pi - \frac{1}{n} \sum_{i=1}^{n} |\pi - |\theta_i - \tilde{\theta}|| \) for candidate angular median \( \tilde{\theta} \) (the amplitude for this estimator is given by the univariate median amplitude of the bivariate quantity), the marginal median is the vector whose components are univariate medians, the HERA mean consists of a MAD-clipping routine at the \( 5\sigma \) level followed by mean averaging, and the geometric and Tukey medians are robust multivariate location estimators that are described in Sections 3.4.1.2 and 3.4.1.3.

Four further examples of KDE plots are shown in Figure 3.3 for visibilities redundant to (1, 12, EE). This baseline group corresponds to 14 m baselines that have a short projected EW separation and a comparatively larger NS separation (see Figure 2.6a for the array layout). This means that the sky does not move as rapidly through the fringe pattern produced by these baselines compared to others with longer projected EW component, and therefore slowly varies in LST, which makes this group useful to study as additional statistics can be run on them by including the time axis. Baselines with projected EW separation < 14 m are eventually rejected

\(^{b}\text{https://seaborn.pydata.org/}\)
\(^{c}\text{The median angular direction can be non-unique, as there can be instances where two or more diameters can equally divide the data and give the same circular mean deviation; we take the mean of these solutions in such cases.}\)
Chapter 3. Normality and robust location estimates of HERA visibilities

Figure 3.1: Scatter plot of aggregated visibility data for baseline group (12, 13, EE) across JDs at frequency channel 640 and LST 5.40 h.

Figure 3.2: KDE of aggregated visibility data for baselines redundant to (12, 13, EE) across JDs at frequency channel 640 and LST 5.40 h. The central distribution of visibilities does not appear to be Gaussian, with the geometric and Tukey medians appearing to better represent the location of the distribution than the other estimators.
at the PS averaging stage as cross-coupling systematics present in their data, which occupy high delays regions (at both positive and negative delays) at low fringe-rate modes, cannot be effectively removed (Kern et al. 2019). These baselines are, nonetheless, used throughout the calibration pipeline, so it is still instructive to understand their distributions.

Such KDE plots can be evolved in frequency or time to get videos that depict the movement of the visibility distributions across the complex plane. While Figures 3.1 to 3.3 consider data aggregated both across JDs and redundant baselines, I also split up this aggregation and look at data distributed across JDs for each baseline redundant to (12, 13, EE). A snapshot at channel 569 and LST 5.67 h is shown in Figure 3.4.

Additionally, I examine the distributions of visibilities for each JD separately for all baselines in redundant group (12, 13, EE) at fixed LST 5.67 h. I show a frame at channel 549 in Figure 3.5.

These KDE plots and the associated videos reveal a considerable amount of structure in the distribution of visibilities aggregated together and separately across JDs and redundant baselines. The spread of the visibilities is quite large, and many distributions are multimodal, with a minority appearing to be completely Gaussian. There also seems to be some possible leakage from neighbouring frequency channels, which is more evident from the videos. What is observed is far from what is expected from an ideal interferometer with bouts of RFI.

We, however, notice that the data across JDs is more consistent than it is across redundant baselines, with the latter showing higher discrepancies between baselines. This indicates that these baselines are seeing different skies and that the calibration pipeline cannot fully reconcile redundant visibilities. While non-redundancy in HERA is a known effect, its extent could be more severe than anticipated. Per-baseline flagging could help remedy such discrepancies. This kind of flagging is not part of the standard pipeline, with per-antenna flagging preferred. From this visualization, averaging first across JDs and then performing some outlier detection on redundant baselines appears to be a promising approach.

### 3.1.2 Multivariate normality

As a more quantitative approach, I present the Henze-Zirkler statistic (Henze & Zirkler 1990) to test and quantify the normality of multivariate data, and apply it to complex redundant HERA visibilities across JDs. I cautiously advise that this test, while commonly used in the literature, should not be solely relied on to check the normality of data; instead, multiple statistics should be used and the data itself should be visualized, if possible.

The Henze-Zirkler is a statistic based on the following non-negative function that measures the distance between the hypothesized function (which is the multivariate normal) and the

---

\[ \text{While there is the ability to exclude specific baselines at the PS level, this is not usually done in practice.} \]
(a) Redundant visibility distribution for baselines redundant to (1, 12, EE) at frequency channel 519 and LST 5.54 h.

(b) Redundant visibility distribution for baselines redundant to (1, 12, EE) at frequency channel 520 and LST 5.62 h.

(c) Visibility distribution with the lowest HZ test p-value (see Section 3.1.2) for baselines redundant to (1, 12, EE), which occurs at frequency channel 515 and LST 5.40 h.

(d) Visibility distribution with the largest distance between its geometric median and HERA mean for baselines redundant to (1, 12, EE), which occurs at frequency channel 531 and LST 5.65 h.

Figure 3.3: KDE plots for redundant and LST aggregated visibilities at various frequencies and LSTs, to illustrate the distribution of typical HERA data.
3.1. Normality of HERA visibilities

![KDE plots for the visibility distribution](image)

**Figure 3.4:** KDE plots for the visibility distribution for each baseline in redundant group (12, 13, EE) aggregated across all JDs at frequency channel 569 and LST 5.67 h.

The observed function:

$$HZ = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} e^{-\frac{d_i^2}{2}} - 2(1 + \beta^2)^{-\frac{d_i}{2}} \sum_{i=1}^{n} e^{-\frac{n^2}{2(1+\beta^2)}} D_i^2 + n(1 + 2\beta^2)^{-\frac{d}{2}}$$  \hspace{1cm} (3.6)

where

$$d = \# \text{ variables}$$  \hspace{1cm} (3.7)

$$\beta = \frac{1}{\sqrt{2}} \left( \frac{n(2d+1)}{4} \right)^{\frac{1}{n+1}}$$  \hspace{1cm} (3.8)

$$D_i^2 = (x_i - \bar{x})^T \Sigma^{-1} (x_i - \bar{x})$$  \hspace{1cm} (3.9)

$$D_{ij}^2 = (x_i - x_j)^T \Sigma^{-1} (x_i - x_j)$$  \hspace{1cm} (3.10)

with $D_i$ giving the Mahalanobis distance of the $i^{th}$ observation to the centroid and $D_{ij}$ the Mahalanobis distance between the $i^{th}$ and $j^{th}$ observations, as $\Sigma$ is the covariance matrix (see Equation 3.17 for the definition of the Mahalanobis distance, which is further explored...
as a method for outlier rejection in Section 3.3). If the data is multivariate normal, $HZ$ is approximately log-normally distributed.

I show the $HZ$ statistic, its corresponding $p$-value and whether the distribution is deemed Gaussian or not for our selected HERA visibilities in Figure 3.6. I also repeat the $HZ$ test but this time using MAD-clipped data; this is also plotted in the bottom row of Figure 3.6. Even with this further outlier rejection step, most of the data still does not appear to be Gaussian.

This process is repeated to test for the multivariate normality of visibilities aggregated either in LST or over redundant baselines to quantify the non-Gaussianity of these distributions separately. I compute the $HZ$ statistic for LST-binned visibilities, and mean average the $p$-value and test outcome (to get a percentage of normal outcomes) over the baseline axis. I repeat the same for visibilities grouped over the redundant baseline axis, with the $p$-value and test outcome averaged over JDs. This is done before any MAD-clipping; the results are largely the same (although we do see a slight improvement) after any clipping. These findings are plotted in Figure 3.7.
3.1. Normality of HERA visibilities

Figure 3.6: $H_Z$ statistic, $p$-value and test outcome for LST-binned + redundantly aggregated HERA visibility data for baseline group (12, 13, EE), before and after MAD-clipping. The $p$-value threshold is set at 0.05, below which the dataset is concluded to deviate from multivariate normality. For the considered data, 59.1% of grouped visibility distributions are deemed normal from the $H_Z$ statistic, with MAD-clipping contributing an additional 0.7% to the tally. For baseline group (1, 12, EE) (which is eventually rejected at the PS estimation stage), only 15% of visibilities pass the multivariate normality test.
Figure 3.7: Mean $p$-value and percentage of normal test outcome for LST-binned (left column; values averaged of the baseline axis) and redundantly aggregated (right column; values averaged of the JD axis) HERA visibility data for baseline group (12, 13, EE). We observe drops in the mean $p$-value for certain time integrations when the $HZ$ statistic is run over visibilities aggregated in LST, while for visibilities aggregated over redundant baselines, we see regions of slightly decreased mean $p$-values towards the lower end of the band, indicating a deviation towards non-normality.

I remark that the $HZ$ test $p$-values tend to be lower for visibilities aggregated over redundant baselines than those for visibilities aggregated in LST, further confirming that redundant visibilities are more divergent than visibilities observed at the same LST for different JDs.

I also looked at the Shapiro-Wilk test (Shapiro & Wilk 1965; used for univariate data) across the $\Re e$ and $\Im m$ components and found similar results to those described in this section, with a significant proportion of visibility slices testing non-Gaussian, even when the components are treated separately.

Along with the many metrics described in the rest of this thesis, the $HZ$ statistic, particularly when computed to probe the normality of visibilities aggregated over redundant baselines, can be used for the identification of faulty baselines.

## 3.2 Overview of principles of robust statistics

A robust statistic attempts to fit a model imposed by the majority of the observed data points, regardless of contamination. Such a statistic should return a fit that is similar to that found without any outliers in the data.
A couple of key features that characterize the robustness of a multivariate estimator are breakdown value and equivariance. Furthermore, a robust estimator should perform well for non-contaminated observations, too; its ability to yield good estimates in such conditions compared to classical estimators is quantified by its efficiency. These concepts are stated below.

### 3.2.1 Breakdown value

The breakdown value of an estimator $T$ for a sample $X_n = \{x_1, \ldots, x_n\}$ is defined to be

$$\text{BD}(T, X_n) = \inf \left\{ \frac{m}{n} : \text{Bias}(m; T, X_n) = \infty \right\}$$

where

$$\text{Bias}(m; T, X_n) = \sup_{X'_n} \|T(X'_n) - T(X_n)\|$$

where $X'_n$ denotes a corrupted dataset that has had $m$ datapoints arbitrarily replaced. Intuitively, the breakdown value is the smallest amount of fractional contamination necessary to entirely upset an estimator. The highest possible breakdown value is 0.5, which is achieved, for example, by the univariate median. The univariate mean, on the other hand, has a breakdown point of 0.

### 3.2.2 Affine equivariance

In the location or scatter estimation of multivariate data, we usually wish to be agnostic as to our choice of axes, such that if we were to shift, shear or rotate our data, we would still get the same estimate. We call this property equivariance, and we ideally want our robust multivariate estimator to obey this.

For location estimator $\Lambda$ to be affine equivariant, it must transform properly under rotation of the data and changes in location and scale. That is to say, for a $d$-by-$d$ nonsingular matrix $A$ and vector $b$ with length $d$:

$$\Lambda(x_1 A + b, \ldots, x_n A + b) = \Lambda(x_1, \ldots, x_n) A + b$$

where $x_1, \ldots, x_n$ is a sample from a $d$-variate distribution and each $x_i$ is a vector of length $d$.

From Donoho & Gasko (1992), the maximum breakdown point of any multivariate affine equivariant estimator is

$$\text{BD}_{\text{max}} = \frac{n - d + 1}{2n - d + 1}$$

Note that if Equation 3.13 is only satisfied when $A$ is orthogonal, such that $A^T = A^{-1}$, then $\Lambda$ is said to be orthogonal equivariant, which means it is equivariant with respect to all Euclidean

---

*Also known as invertible: for a matrix $A$ there exists an inverse $A^{-1}$ such that $AA^{-1} = A^{-1}A = I$.\footnote{Also known as invertible: for a matrix $A$ there exists an inverse $A^{-1}$ such that $AA^{-1} = A^{-1}A = I$.}
Chapter 3. Normality and robust location estimates of HERA visibilities

similarity transformations (rotation, reflection and translation). Orthogonality is not as rigid as affinity but is ample for most situations.

For a measure of dispersion $V$, affine equivariance is established if

$$V(AX + b) = AV(X)A^\top$$  \hspace{1cm} (3.15)

### 3.2.3 Efficiency

We define the relative efficiency between two estimators $\hat{\theta}$ and $\tilde{\theta}$, for a fixed underlying distribution, as

$$RE(\tilde{\theta}; \hat{\theta}) = \frac{\text{Var}(\hat{\theta})}{\text{Var}(\tilde{\theta})}$$  \hspace{1cm} (3.16)

with $\hat{\theta}$ needing approximately $RE$ times as many observations as $\tilde{\theta}$ to reproduce an estimate with the same precision. In the limit of $n \to \infty$, $RE \to ARE$: the asymptotic relative efficiency. As an example, $ARE$ (median; mean) $= 2/\pi \approx 0.64$.

For location estimates, the term efficiency is usually taken to be $ARE(\tilde{\theta}; \text{mean})$.

### 3.3 Multivariate robust outlier detection

LST-binning in the analysis of H1C_IDR2.2 uses MAD-clipping at the $5\sigma$ level (see Equation 2.37) followed by mean averaging to estimate the location of visibilities aggregated across JDs. This is done for each frequency/time/baseline slice.

MAD-clipping, while robust, is not equivariant, as it is performed on the $\Re e$ and $\Im m$ components of the visibilities separately. Data points are flagged if either component is clipped. Through this method, it is implied that the location of the visibility distribution is given by the marginal median. The non-equivariance of MAD-clipping can also be seen by considering the region in the complex plan that it rejects. $Z_{\Re e}^{\text{mod}}$ and $Z_{\Im m}^{\text{mod}}$ only represent the spread of the data along the $\Re e$ and $\Im m$ axes, thus, resulting in a rectangular boundary that does not account for the covariance of the visibility distribution.

I present an improved outlier rejection routine that considers robust Mahalanobis distances calculated with Minimum Covariance Determinant (MCD) location and covariance estimates, which overcomes some of the shortfalls of MAD-clipping, as the location and scatter from the MCD estimator are both robust and equivariant. The aim is not only to present a theoretically better method of outlier rejection but to potentially also address issues in the H1C_IDR2.2 results that may be due to low-level RFI (cf. the Band 2 Field 2 results in The HERA Collaboration et al. 2022c).

The flagging capabilities of this improved method are compared to that of MAD-clipping, and preliminary averaged visibility and PS results are also shown.
3.3. Multivariate robust outlier detection

3.3.1 Outlier detection with robust Mahalanobis distances

The Mahalanobis distance (Mahalanobis 1936) is a multivariate distance metric that measures the distance between a point $x_i$ and a distribution. It is given by

$$\text{MD}(x_i) = \left((x_i - \hat{\mu})^\top \hat{\Sigma}^{-1} (x_i - \hat{\mu})\right)^{\frac{1}{2}}$$

(3.17)

where $\hat{\mu}$ is the sample multivariate mean and $\hat{\Sigma}$ is the sample covariance matrix of the distribution. Unlike Euclidean distances, it accounts for any correlation between variables. It is commonly used to find outliers in multivariate sets.

Naturally, the mean and covariance will be heavily influenced by the presence of outliers; obtaining good robust estimates of $\hat{\mu}$ and $\hat{\Sigma}$ are necessary to measure the outlyingness of data points and to have a proper distance-based outlier detection procedure. Therefore, we modify Equation 3.17 to get robust Mahalanobis distances:

$$\text{RMD}(x_i) = \left((x_i - \hat{\mu}_r)^\top \hat{\Sigma}_r^{-1} (x_i - \hat{\mu}_r)\right)^{\frac{1}{2}}$$

(3.18)

where $\hat{\mu}$ and $\hat{\Sigma}$ have been replaced with $\hat{\mu}_r$ and $\hat{\Sigma}_r$, which are robust estimates of centrality and covariance matrix.

In practice, the most frequently used covariance estimator is the MCD estimator (Rousseeuw 1984), which is based on the computation of the ellipsoid with the smallest volume or with the smallest covariance determinant that would encompass at least half of the data points.

3.3.1.1 Minimum Covariance Determinant estimator

This MCD estimator is a high-breakdown and affine equivariant estimator of both location and scatter. It consists of determining the subset $J$ of observations of size $h$ that minimizes the determinant of the sample covariance matrix, computed from only these $h$ good observations, which are not considered to be outliers. The choice of $h$ (also called the tuning constant) determines the robustness of the estimator; it is a compromise between robustness and efficiency. Once this subset of size $h$ is found, it is possible to estimate the location and covariance matrix based only on that subset.

More formally, $J$ is defined as

$$J = \left\{ h : |\hat{\Sigma}_J| < |\hat{\Sigma}_K| \forall K \text{ s.t. } \#K = h \right\}$$

(3.19)

with $(n + d + 1)/2 \leq h \leq n$ for an $n \times d$ data matrix, and where $|A|$ denotes the determinant of a matrix $A$, and $\#K$ denotes the cardinality of the subset $K$. The location and scatter are then
Chapter 3. Normality and robust location estimates of HERA visibilities

estimated to be

$$\hat{\mu}_{\text{MCD}} = \frac{1}{h} \sum_{i \in J} x_i$$  \hspace{1cm} (3.20)

$$\hat{\Sigma}_{\text{MCD}} = \frac{1}{h} \sum_{i \in J} (x_i - \hat{\mu}_{\text{MCD}})(x_i - \hat{\mu}_{\text{MCD}})^\top$$  \hspace{1cm} (3.21)

The tuning constant $h$ is generally taken to be its minimum value of $(n + d + 1)/2$ to maximize the robustness of the MCD estimator.

The computation of the exact MCD estimator is very demanding, as it requires the evaluation of $\binom{n}{h}$ subsets of size $h$. The fast-mcd algorithm (Rousseeuw & Driessen 1999) is computationally efficient and allows the MCD estimator to be applied to large datasets. The basic premise of the fast-mcd algorithm revolves around the key C-step, which considers $m$ random subsets of size $h$: the location and covariance of each subset is estimated (using standard mean and covariance estimation, as described in Equations 3.20 and 3.21), which are then used to calculate the Mahalanobis distances for all $n$ points (these distances are separate for each subset). Each subset is then re-populated with the $h$ points with smallest Mahalanobis distances, which concludes an iteration. This process ceases for a given subset when the determinants of the estimated covariances for consecutive C-steps are equal or if the determinant is zero. The C-step procedure does not necessarily find the global minimum of the MCD objective function for a given initial subset, hence why many initial subsets are required. C-steps are applied to each of these subsets until convergence; the solution with the lowest determinant is kept. The intricacies of initial subset generation is discussed in Rousseeuw & Driessen (1999).

I use the sklearn\textsuperscript{1} MinCovDet class for fast-mcd covariance estimation and consequent Mahalanobis distance computation for the research in this thesis.

If the outlier detection algorithm is applied to data for the same baselines but on different JDs (as is done in the LST-binning pipeline), then the data should mostly be near-normal enough for MCD to work adequately. However, MCD should not be used for multimodal distributions or those that deviate too far from Gaussianity; there is evidence that distinct but redundant HERA baselines may fall under this category.

3.3.1.2 Robust Mahalanobis distance clipping

For robust outlier detection, we require $\text{RMD}(x_i) > c_d$ for some threshold $c_d$. It follows that $\text{RMD}(x_i)$ approximately follows a $\chi^2$ distribution (Rousseeuw & van Zomeren 1990), hence, we can use

$$c_d = \sqrt{\chi^2_{d, q}}$$  \hspace{1cm} (3.22)

\textsuperscript{1}https://scikit-learn.org/stable/
that corresponds to the square root of the \( q \)-quantile of the \( \chi^2 \) distribution with \( d \) degrees of freedom.

We, therefore, outline steps for finding outliers through these MCD Mahalanobis distances, which we call robust Mahalanobis distance (RMD)-clipping:

1) Compute RMD(\( \mathbf{x}_i \)) (see Equation 3.18) using fast-mcd with \( h = (n + d + 1)/2 \) (thus maximizing the breakdown point of the estimator)

2) Compute the \( q \)-quantile of the chi-squared distribution \( \chi^2_{d, q} \); its square root is the clipping threshold \( c_d \) (\( q \) is usually taken to be 0.975, 0.99, 0.999 etc.; I show in Section 3.3.1.3 how this quantile can be related to multiples of \( \sigma \))

3) Declare RMD(\( \mathbf{x}_i \)) > \( c_d \) as possible outliers

We note that the threshold \( c_d \) for this method can be modified such that it is adjusted to the sample size; an adjusted quantile can be used instead (see e.g. Filzmoser et al. 2005) in step 2. Proceeding with the adjusted quantile option generally improves the false classification rates while maintaining the same correct classification rates (Cabana et al. 2021). The adjusted threshold is computed by comparing the theoretical cumulative \( \chi^2 \) distribution function and the empirical cumulative distribution function (CDF) of the squared robust distance samples, and finding the supremum of the difference between the two tails of these distributions.

### 3.3.1.3 Relating quantiles between the \( N \) and \( \chi^2 \) distributions

When dealing with outlier detection procedures, we commonly set the rejection threshold in terms of multiples of \( \sigma \), the standard deviation of the normal distribution \( N \). From Equation 3.22, the threshold for outlier rejection in RMD-clipping is given in terms of \( \chi^2 \) quantiles. Hence, we need to relate the quantiles of the chi-squared distribution \( \chi^2_{d, q} \) to those of a Gaussian distribution \( z_q \).

Given a number of standard deviations \( n \), the probability \( p \) that a normal deviate lies in the range between \( \mu - n\sigma \) and \( \mu + n\sigma \) is given by

\[
p = F_N(\mu + n\sigma) - F_N(\mu - n\sigma) = \Phi(n) - \Phi(-n) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{n}{\sqrt{2}} \right) \right]
\]

where \( F_N \) is the CDF for a generic normal distribution, \( \Phi \) the CDF for the standard normal distribution and erf is the error function, which are all related through

\[
F_N(x) = \Phi \left( \frac{x - \mu}{\sigma} \right) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{x}{\sqrt{2}} \right) \right]
\]

To retrieve \( \chi^2_{d, q} \) we apply the quantile function \( F^{-1} \chi^2 \) (i.e. the inverse of the CDF) to \( p \):

\[
\chi^2_{d, q} = F^{-1} \chi^2 (p; d)
\]
Chapter 3. Normality and robust location estimates of HERA visibilities

$F_{\chi^2}^{-1}$ does not have a simple, closed-form representation. Its CDF can, however, be given in terms of complete $\Gamma$ and lower incomplete $\gamma$ gamma functions:

$$F_{\chi^2}(x; d) = \frac{\gamma\left(\frac{d}{2}, \frac{x}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}$$ \hspace{1cm} (3.26)

Using Equation 3.25, we numerically compute the $\chi^2_{d, q}$ quantile equivalents for $z_q = 4$ and $z_q = 5$ (corresponding to a $4\sigma$ and $5\sigma$ threshold) to be 19.334 and 28.744, respectively.

We note the following approximation (Fisher 1992) to Equation 3.25, which bypasses the complexities of working with $F_{\chi^2}^{-1}$ and works well for large $d$:

$$\chi^2_{d, q} \approx \frac{1}{2} \left(z_q + \sqrt{2d - 1}\right)^2$$ \hspace{1cm} (3.27)

3.3.2 LST-binning results and comparison to MAD-clipping

3.3.2.1 Illustrative example

To better visualize how MAD- and RMD-clipping work and contrast, I look at sample visibilities across the 18 nights of H1C_IDR2.2 for the 14 m baseline (55, 71, EE) at frequency channel 514 that fall into the LST bin centred at 5.59 h with cadence 21.4 s (meaning 2 data points for each JD). I then perform outlier rejection with both the MAD- and RMD-clipping procedures, with clipping at the $5\sigma$ threshold. I also only perform the clipping on each slice if there are at least 5 unflagged data points.

On the left of Figure 3.8, I show the scatter of selected data points as well as the MAD-clipping boundary and outlier region. I show the same for RMD-clipping on the right, with concentric Mahalanobis distance contours also marked.

I compare these MAD and RMD boundaries by overlaying them in Figure 3.9a; I also draw an indicative ellipse with width $2 \times 5\sigma_{\text{mad}}^\text{Re}$ and height $2 \times 5\sigma_{\text{mad}}^\text{Im}$ on the sample plot to show a somewhat halfway house between the two methods. The RMD boundary is tighter and more closely confines the distribution of the data compared to MAD-clipping. This is typical for data slices at this stage of the reduction pipeline. The RMD-clipping method is seen to generally reject more data points.

The estimated MCD covariance computed for the RMD-clipping in this example is given by the following matrix:

$$C = \begin{bmatrix} 6.5727 & -2.0061 \\ -2.0061 & 3.5462 \end{bmatrix}$$ \hspace{1cm} (3.28)
3.3. Multivariate robust outlier detection

Figure 3.8: Left: outlier detection with MAD-clipping, where the red rectangle shows the boundary with width $2 \times 5\sigma_{\text{mad}}^{\Re}$ and height $2 \times 5\sigma_{\text{mad}}^{\Im}$. No data points are flagged as outliers. The marginal median is also shown, which represents the location of the distribution according to this particular method. Right: Outlier detection with RMD-clipping, with Mahalanobis distance contours shown and the red ellipse demarcating the inlier/outlier boundary, corresponding to the contour with $RMD(x) = \chi_{\text{thresh}} = 5.361$. Two points are flagged in this case. The location returned from the MCD estimator is also marked. These plots have equal axis scaling.

(a) MAD- and RMD-clipping comparison for the data and boundaries in Figure 3.8.  
(b) A second example where the RMD algorithm discounts 13 points that would otherwise not come close to being clipped with the MAD method.

Figure 3.9: Comparison of the outlier boundaries and locations set by the MAD (green) and RMD (orange) anomaly detection methods. The ellipse with equation $\left(\frac{x - \text{mmed}_{\Re}}{5\sigma_{\Re}}\right)^2 + \left(\frac{y - \text{mmed}_{\Im}}{5\sigma_{\Im}}\right)^2 = 1$ is drawn (purple) as a potentially smoother version of MAD-clipping that could be implemented with little added computational cost. Figures 3.9a and 3.9b both have equal axis scaling.
From the eigendecomposition of Equation 3.28, we find the eigenvectors and eigenvalues to be

\[
\mathbf{v}_1 = \begin{bmatrix} 0.8950 \\ 0.4460 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -0.4460 \\ 0.8950 \end{bmatrix} \tag{3.29}
\]

\[
\lambda_1 = 7.5723, \quad \lambda_2 = 2.5466 \tag{3.30}
\]

These can be used to draw the covariance error ellipse that sets the boundary for outliers. The width and height of the ellipse are given by

\[
w = \chi_q \sqrt{\lambda_1}, \quad h = \chi_q \sqrt{\lambda_2},
\]

with the orientation given by

\[
\alpha = \arctan2(\lambda_1 - C_{0,0}, C_{0,1}),
\]

where \( C_{i,j} \) denotes the entry of the covariance matrix at the \( i \)th row and \( j \)th column.

In Figure 3.9b, I show a situation where RMD-clipping rejects a significant number of data points that would otherwise be well-within the MAD-clipping boundary. This data slice is for baseline (124, 143, EE) at channel 201 and LST 5.63 h. Even by eye, it is difficult to judge if the points on the outside of the RMD boundary are outliers. As RMD-clipping already rejects a much higher proportion of data points compared to MAD-clipping (for equivalent quantile thresholds), the threshold for RMD-clipping could be lowered to reduce the average number of outliers it picks out. As mentioned in Section 3.3.1.1, the MCD estimator is likely to perform poorly for non-Gaussian data, which will affect a small proportion of data slices. However, even in such circumstances, it seems that the resulting location of the clipped distribution still represents the central tendency of the data.

The MCD-clipping procedure implies a covariance to the distribution of data, whereas, for perfect white noise, we do not expect any off-diagonal elements to the covariance matrix. The asymmetry and skew seen throughout visibility distributions is likely to be caused by the contribution of sky emission to the noise. Fixed sources will have fluctuating amplitudes that will add a noise contribution in the complex radial direction, which explains the tilted covariance ellipses seen in the data. These fluctuating amplitudes are due to the intrinsic self-noise of sources; for a source at a fixed position on the sky, there is no change in visibility phase, and thus only in amplitude. This is particularly relevant at lower frequencies, where the power contributed by astronomical sources becomes an increasingly significant portion of the total received power. Furthermore, certain calibration steps from the analysis pipeline may stretch the data along particular axes.

### 3.3.2.2 Clip flags for a 20 min LST-binned file

I test the clipping routines on the LST-binning of 20 min of H1C_IDR2.2 visibility data between LSTs 5.39 – 5.74 h for all the unflagged 14 m baselines (all orientations are considered). I look at the EE polarization only. Again, I use a 5\( \sigma \) threshold and only flag data slices with more than 5 data points. I summarize the results in Table 3.1.
3.3. Multivariate robust outlier detection

<table>
<thead>
<tr>
<th>Clipping routine</th>
<th>Compute time</th>
<th>Number of additional flags</th>
<th>% of total flags</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAD</td>
<td>≈ 30 s</td>
<td>146,141</td>
<td>0.192%</td>
</tr>
<tr>
<td>RMD</td>
<td>≈ 3 h</td>
<td>1,250,164</td>
<td>1.619%</td>
</tr>
</tbody>
</table>

Table 3.1: Sample flagging capabilities for the MAD- and RMD-clipping routines. Compute times are from using 1 node, 8 cores on the NRAO cluster. For the equivalent quantile thresholds, RMD-clipping flags 8.5 times more data points than MAD-clipping; this is far more than would be expected if the data were Gaussian distributed. I note that the sklearn.covariance.MinCovDet MCD estimator used for these results has a (pseudo) randomness element that is used for shuffling the data; the random state should be fixed for reproducibility. RMD-clipping is computationally expensive, being over 360 times slower than MAD-clipping; this is because the MCD estimator and subsequent Mahalanobis distance calculations need to be run separately for each frequency/time/baseline slice, while for MAD-clipping the operations are simple and vectorized in NumPy.

I then apply these new flags to the dataset, and, in line with the LST-binning pipeline, I take the mean of each LST bin to get JD-averaged visibility products.

Looking more closely at the visibilities redundant (both in length and orientation) to (12, 13, EE), I further take the mean over baselines to get the redundantly averaged visibility estimates shown in Figure 3.10 (amplitude) and Figure 3.11 (phase). I only plot the MAD-clipped results as the heatmaps appear nearly identical to those from RMD-clipping: the difference in the mean between unflagged MAD- and RMD-clipped points is minute; any difference is further suppressed from averaging along the baselines axis. We, however, also show residual plots, which have some horizontal structure indicating that MAD- and RMD-clipping result in slightly different estimates at certain times.

In Figure 3.12, I show the percentage of data points flagged for each frequency/time slice with the MAD and RMD outlier rejection algorithms (with clipping done across JDs only and flags summed over the baseline axis). Both processes seem to pick out the same problematic time integrations. RMD-clipping does, however, seem to be able to better pick out anomalous features confined to particular frequencies (i.e. it has some light vertical lines not seen in the plot for MAD-clipping, even if the number of flags is normalized), with some of these channels known to be bad. This suggests that RMD-clipping is more effective at finding genuine problematic issues with the data.

For a simple PS comparison of these clipping results, I compute the CPS for Band 1 (channels 175 – 515) across all baseline permutations, where we are still only looking at baselines redundant to (12, 13, EE). I then mean average these CPS in time, producing Figure 3.13. The resulting PS are very similar and achieve essentially the same noise-floor, with the mean residual...
Figure 3.10: Top: visibility amplitude after mean averaging across JDs and baselines post MAD-clipping (RMD-clipped visibilities look very similar). Bottom: residual between MAD- and RMD-clipped visibility amplitude averages. We see some time integrations showing increased divergence between the two clipping routines.

Figure 3.11: Top: visibility phase after mean averaging across JDs and baselines post MAD-clipping. Bottom: residual between MAD- and RMD-clipped visibility phase averages.
3.3. *Multivariate robust outlier detection*

**Figure 3.12:** Percentage of flagged data points across JDs and baselines returned by the MAD (top) and RMD (bottom) clipping routines, for the same $5\sigma$ threshold. Some problematic time integrations are picked up by both methods, however RMD-clipping also identifies some further time integrations, as well as some frequency channels that have more outliers, indicating that it can find real issues at a finer level.

**Figure 3.13:** Absolute value of the CPS averaged across all baseline permutations for the MAD- and RMD-clipped visibilities that have been mean averaged across JDs. The absolute value of the mean of all CPS over time is also shown. These final mean-averaged CPS are isolated and compared in the rightmost plot; the CPS result for unclipped data is also shown for comparison.
between the two PS for delays $|\eta| \geq 1.25 \mu s$ (chosen to avoid the $\pm 1 \mu s$ bump) $\sim 10^{-5}$ Jy$^2$ Hz$^2$. The mean residual between the PS from the unclipped and MAD-clipped data is lower, at $\sim 10^{-6}$ Jy$^2$ Hz$^2$, indicating that these two results are much the same for this dataset. As a reminder, the Blackman-Harris window is used for all PS computations in this thesis, and the conversion from $\eta$ to $k_\parallel$ is shown in Table 2.2. A deeper analysis, with RMD-clipping extended to the full H1C_IDR2.2 dataset, would offer more definitive results.

3.3.3 The role of outlier rejection

The robust method of outlier rejection presented in Section 3.3.1 that employs MCD-based Mahalanobis distances, which we call RMD-clipping, is a statistically rigorous and, in this case, more appropriate flagging routine than MAD-clipping as it considers both the robust location and covariance of the data, and thus delineates reasonable tilted elliptical outlier boundaries. Comparatively, MAD-clipping is rudimentary as it draws a boundary box in the complex plane from treating the $\Re e$ and $\Im m$ components separately.

It is found that the $\Re e$ and $\Im m$ components of LST-binned visibilities not only have unequal variance but also have non-zero covariance. The RMD-clipping method should, therefore, perform better than MAD-clipping, with preliminary results showing equal performance in the PS; extending this procedure to the full H1C_IDR2.2 analysis could further improve the limits set in The HERA Collaboration et al. (2022c). The clipping observed for equivalent quantiles is more aggressive for RMD-clipping (by a factor of around 8). With existing concerns of overflagging, the threshold could be increased if RMD-clipping is to be used in production.

Despite being the preferred method of outlier rejection, RMD-clipping is computationally expensive. Using RMD-clipping instead of MAD-clipping at the LST averaging stage would be very costly with the current implementation. Acceleration with software and hardware of the robust Mahalanobis distance computation using the MCD covariance estimator (currently done with sklearn) should be investigated if RMD-clipping is to be deployed in production.

Ultimately, the final location estimate in the averaging procedure is the important result; thus, we look to robust multivariate estimators such as the geometric median to get location estimates directly. These would better represent contaminated/non-normal data, and can be used outright without having to conduct any prior outlier rejection.

There is still application to use RMD-clipping for per-baseline flagging at the PS computation stage, as it was shown in Section 3.1 that baseline-to-baseline variation is greater and deviates further from normality than fluctuations across JDs.
3.4 Robust location estimates of HERA visibilities

3.4.1 Multivariate median estimators

I introduce several robust estimators of location, which can be used to better find the central tendency of non-normal distributions compared to classical methods.

3.4.1.1 Medians in bivariate and higher dimensional data

The median of a univariate dataset is a natural robust estimator for the centre of a distribution. It can be defined as the order statistic of rank \((n + 1)/2\) when \(n\) is odd and as the mean of the order statistics of rank \(n/2\) and \((n + 1)/2\) when \(n\) is even (other definitions also exist, see e.g. Oja 2013). It has different characteristics from the mean, chiefly through its breakdown properties. A single infinite point contaminating a dataset will send the mean to infinity. By comparison, at least 50% of the data must be moved to infinity to reproduce the same effect in the median; it is said to have a breakdown point of 1/2 (see Section 3.2.1 for the mathematical formalism of breakdown value). Thus, the median is robust to outliers and is commonly used in nonparametric problems.

Generalizing the median to higher dimensions is not straightforward. While it can be tempting to take the marginal median of multivariate data, that is, to take the univariate median across each coordinate separately, this leads to a result that is not necessarily representative of the central tendency of the distribution. The concept of the median relies on ordering. In higher dimensions, there is, however, no natural concept of rank and a marginal median estimate will heavily depend on the choice of axes. For these reasons, alternative equivariant methods that wholly consider all coordinate information together are required to produce a better embodiment of the median in higher dimensions.

In this section, I present higher-dimensional analogues of the median that are used in robust and nonparametric data analysis and inference. Such robust statistical techniques, while less influenced by abnormal observations, have higher computational complexity, making them historically less appealing in practice. With the increased computational power now available, together with the dawn of machine learning (ML) libraries, these methods can be efficiently implemented and their use more tractable. Such robust location estimators are particularly relevant in the averaging of large datasets riddled with outliers; this is especially pertinent to radio interferometry where RFI and other non-Gaussian effects are prevalent. With limited data, good estimates of visibilities are required that are not skewed by anomalous effects.
3.4.1.2 The geometric median

The geometric median, also known as the $L_1$-median, is defined as the value of the argument $y \in \mathbb{R}^d$ that minimizes the sum of Euclidean distances between $y$ and all points $x_i$:

$$GM = \arg \min_{y \in \mathbb{R}^d} \sum_{i=1}^{n} \|x_i - y\|$$

(3.31)

The geometric median can also be weighted:

$$GM_w = \arg \min_{y \in \mathbb{R}^d} \sum_{i=1}^{n} \eta_i \|x_i - y\|$$

(3.32)

with this latter minimization also known as the famous Weber problem (Weber 1909) in location theory, which was initially posed as a transportation cost minimization problem: the best location for a warehouse that services $n$ customers each at position $x_i$ is to be found, with different customers being associated with different transportation costs $\eta_i$.

When $x_i$ are not collinear, the cost function $\sum_{i=1}^{n} \|x_i - y\|$ is positive and strictly convex in $\mathbb{R}^d$, and hence the minimum is unique.

An iterative solution technique to Equation 3.32 was first proposed by Weiszfeld (1937), which is given by

$$y \rightarrow y^{t+1} = T(y^t)$$

(3.33)

where

$$T(y) = \left( \sum_{x_i \neq y} \frac{\eta_i}{\|x_i - y\|} \right)^{-1} \sum_{x_i \neq y} \frac{\eta_i x_i}{\|x_i - y\|}$$

(3.34)

for steps $t = 0, 1, \ldots$ with $y^0 \in \mathbb{R}^d \setminus X_n$, where $X_n = \{x_1, \ldots, x_n\}$ denotes the set of data points. This algorithm converges to the geometric median and can be stopped when a suitable termination criterion is reached. If $y$ lands on $x_k \in X_n$, then the algorithm also ceases; this does not, however, necessarily mean that the geometric median is found. In practice, to deal with the null denominators $d_k = \|x_k - y\| \neq 0$ in Equation 3.34 and avoid early termination, $d_k$ can be shifted by a small amount (nominally +1) to ensure that the algorithm does not get stuck.

To get an intuition as to the basis of the Weiszfeld algorithm (this is not a proof), consider taking the derivative of the sum $f(y) = \sum_{i=1}^{n} \eta_i \|x_i - y\|$ in the definition of the weighted geometric median (Equation 3.32):

$$\nabla f(y) = \sum_{x_i \neq y} \frac{\eta_i}{\|x_i - y\|} (y - x_i)$$

(3.35)
Now, assume that \( x_i \) are not collinear and that \( y^* \) is the unique optimal solution. For \( y^* \notin X_n \), we therefore have
\[
\nabla f(y^*) = \sum_{i=1}^{n} \eta_i \frac{y^* - x_i}{\|x_i - y^*\|} = 0
\]
we next extract \( y^* \) by rearrangement to obtain
\[
y^* = \left\{ \sum_{i=1}^{n} \frac{\eta_i}{\|x_i - y^*\|} \right\}^{-1} \sum_{i=1}^{n} \frac{\eta_i x_i}{\|x_i - y^*\|}
\]  
which is similar to the relation in Equation 3.34. The Weiszfeld algorithm is a fixed point method for solving Equation 3.37.

Alternatively and more rigorously, the Weiszfeld algorithm can be modified (Vardi & Zhang 2000) such that if \( y = x_k \), then we consider a weighted average of \( T(x_k) \) and \( x_k \) for our next step in the iteration. We first define the multiplicity to be
\[
\eta(y) = \begin{cases} 
\eta_k, & \text{if } y = x_k \in X_n \\
0, & \text{otherwise}
\end{cases}
\]
The modified Weiszfeld algorithm is then given by
\[
y \rightarrow T_M(y) = \left( 1 - \frac{\eta(y)}{r(y)} \right)^+ T(y) + \min \left( 1, \frac{\eta(y)}{r(y)} \right) y
\]
where
\[
r(y) = \|R(y)\|, \quad R(y) = \sum_{x_i \neq y}^{n} \eta_i \frac{x_i - y}{\|x_i - y\|} = (T(y) - y) \sum_{x_i \neq y}^{n} \frac{\eta_i}{\|x_i - y\|}
\]
where the convention \( 0/0 = 0 \) is used when computing \( \eta(y)/r(y) \) in Equation 3.39. For \( y \notin X_n \), \( T(y) = T_M(y) \eta(y) = 0 \), thus, returning to the original algorithm. The modified Weiszfeld algorithm is slightly more expensive but better deals with the situation when \( y = x_k \).

Various other algorithms have since been proposed that improved on computational speed (see e.g. Cohen et al. 2016). Even without such methods, we can now easily solve Equation 3.32 numerically with common optimization software, which can also be accelerated with ML libraries and GPUs.

I implement the Weiszfeld algorithm in Python, which is publicly available as part of the robstat package. A CPython version of the code was also experimented with, which provides a \( \sim 10\% \) speedup. There is also the option to use the modified Weiszfeld algorithm or to run a BFGS minimization of the Euclidean distances with SciPy.

\*The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm is a quasi-Newton iterative method for solving unconstrained non-linear optimizations. Here, the gradient that determines the descent direction is preconditioned with curvature information, namely the Hessian, which is approximated and updated at each iteration. See e.g. Nocedal & Wright (2006) for a detailed description of the BFGS algorithm. This is the method used for most of the minimizations in this thesis.

\*https://scipy.org/
The geometric median has a maximally high breakdown point of 1/2 and is orthogonally equivariant: this means that it is equivariant for scale, translation and orthogonal transformations, i.e. \( A \) must be an orthogonal \( d \times d \) matrix in Equation 3.13, which is sufficient in most situations (Fritz et al. 2012). Furthermore, the geometric median has good efficiency. As shown in Brown (1983), the asymptotic efficiencies for the geometric median components in two dimensions reach 0.785 and slowly degrade as the distribution of points becomes narrower ellipses. The efficiency of the estimate in the direction of the largest principal component tends to \( 2/\pi = 0.637 \) as the ellipse flattens and the distribution becomes one-dimensional in appearance. The efficiency in the direction of the smaller principal component degrades slightly more rapidly, but only deteriorates beyond 0.637 for \( \lambda < 0.2 \), where \( \lambda^2 \) is the ratio of the smaller to the larger variance of the principal components. We note that the efficiency of the geometric median further improves as the number of dimensions increases. The distribution of complex visibilities encountered in this research is near circular, with \( \lambda \gtrsim 0.5 \) meaning efficiencies \( \gtrsim 0.76 \).

### 3.4.1.3 The Tukey median

Tukey (1975) proposed the **halfspace depth** as a tool to visually describe multivariate datasets. He suggested this depth could be used to define multivariate analogues of univariate rank and order statistics via depth-induced **contours**.

For a finite set of data points \( \mathcal{X}_n = \{x_1, \ldots, x_n\} \) in \( \mathbb{R}^d \), the Tukey depth, or halfspace depth, of any point \( y \in \mathbb{R}^d \) determines how central the point is inside the data cloud; it is defined as the minimum fraction of data points in any closed halfspace determined by a hyperplane through \( y \):

\[
d_T(y; \mathcal{X}_n) = \inf_{\|u\|=1} \frac{1}{n} \# \left\{ i \in \{1, \ldots, n\} : u^\top x_i \geq u^\top y \right\}
\]

(3.41)

where \( \|u\| = 1 \) denotes the unit sphere in \( \mathbb{R}^d \).

The set of all points with depth \( \geq \kappa \) is called the Tukey \( \kappa \)-depth region \( \mathcal{D}(\kappa) \):

\[
\mathcal{D}(\kappa) = \left\{ x \in \mathbb{R}^d : d_T(x) \geq \kappa \right\}
\]

(3.42)

with \( 0 < \kappa \leq 1 \). The halfspace depth regions form a sequence of closed convex polyhedra, with each polyhedron included in \( \text{conv}(\mathcal{X}_n) \), making it compact. Tukey regions are also nested: they shrink with increasing \( \kappa \). For practical purposes, it can be more convenient to work with the number of deepest points taken into account \( n_p = \kappa n \), where we scale the depth \( \kappa \) by \( n \) (thus undoing the \( 1/n \) factor in Equation 3.41, with \( n_p \in \mathbb{N}^+ : 1 \leq n_p \leq n \)). For data in
3.4. Robust location estimates of HERA visibilities

A d-variate dataset is said to be in general position if there are no more than d points in any (d−1)-dimensional hyperplane.

3https://cran.r-project.org/web/packages/TukeyRegion/
I apply the geometric median estimator to HERA visibilities to find robust location estimates of the complex data when averaging over both/either the time and/or baseline axis. I also compute simple PS of the resulting location estimates for spectral comparison.

The geometric median estimates are compared to $5\sigma$ MAD-clipping followed by mean averaging, which I shall henceforth refer to as the HERA mean. This is the estimator used for...
3.4. Robust location estimates of HERA visibilities

(a) \( \mathcal{D}(10) \).

(b) \( \mathcal{D}(50) \).

(c) \( \mathcal{D}(100) \).

(d) \( \mathcal{D}(200) \).

Figure 3.15: Tukey regions \( \mathcal{D}(n_p) \) for \( n_p = 10, 50, 100, 200 \) (left to right, top to bottom) for three-dimensional Gaussian simulated data with mean \( \mu = [0, 0, 0] \) and covariance matrix \( \Sigma = [[1, 1, 1], [1, 2, 2], [1, 2, 4]] \); \( n = 500, d = 3 \). The Tukey median \( \bar{T} = [-0.0032, 0.0106, -0.0396] \) is reached at \( n_p = 234 \) and is the barycentre of a region with 8 vertices and 12 hypertriangle defining facets.

averaging in LST for the results in The HERA Collaboration et al. (2022c). I use this as the benchmark for comparison with other location estimates.

3.4.2.1 LST + redundant averaging

The visibility estimates from averaging both in LST and across redundant baselines with various robust and standard estimators are shown in Figure 3.16. By eye, all these estimates
Chapter 3. Normality and robust location estimates of HERA visibilities

<table>
<thead>
<tr>
<th>Median estimator</th>
<th>Breakdown value</th>
<th>Affine equivariance</th>
<th>Time Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal</td>
<td>1/2</td>
<td>no</td>
<td>$\sim O(n)$</td>
</tr>
<tr>
<td>Geometric</td>
<td>1/2</td>
<td>no*</td>
<td>$\sim O(n)$</td>
</tr>
<tr>
<td>Tukey</td>
<td>worst-case 1/(d + 1) typically 1/3</td>
<td>yes</td>
<td>$O(n \log^3 n)$</td>
</tr>
<tr>
<td>Simplicial</td>
<td>$\leq 1/(d + 2)$</td>
<td>yes</td>
<td>$O(n^d)$</td>
</tr>
<tr>
<td>Oja</td>
<td>$2/n \approx 0$</td>
<td>yes</td>
<td>$O(n \log^3 n)$</td>
</tr>
<tr>
<td>Projection</td>
<td>1/2</td>
<td>yes</td>
<td>-</td>
</tr>
</tbody>
</table>

*Table 3.2: Properties of various multivariate median estimators. *The geometric median is orthogonal equivariant.

seem consistent with each other and have the same overall shape; the statistics of these visibility estimates will enable us to better compare them.

As an aside, I note that the estimates for the marginal median are considerably worse when the data is not rephased, with some adjacent time or frequency channels becoming disjointed for the amplitude, phase and $\Im$ plots, whereas the other estimators perform adequately (I do not show this here).

Location estimate smoothness

I look at the smoothness for the location estimates shown in Figure 3.16 by calculating the standard deviation of the distances between successive points in either frequency or time:

$$S = \sqrt{\frac{\sum_{i}^{D} (d_i - \bar{d})^2}{N_D}}$$  \hspace{1cm} (3.44)

where

$$d_i = x_{i+1} - x_i$$  \hspace{1cm} (3.45)

I then take the mean of Equation 3.44 over the other dimension to obtain a single quantity.

These smoothness proxies are shown in Table 3.3. The HERA mean provides the smoothest visibility estimates, with the geometric median in second place. With some distributions experiencing multimodality, a median estimator will choose a peak and possibly jump to another peak if the distribution for the next frequency/time slice is similar (depending on the fluctuation in magnitude of the peaks). A mean estimator will tend to lie in the centre of the peaks and is more stable.
### 3.4. Robust location estimates of HERA visibilities

#### 3.4.2.2 Power spectrum estimation

Before resorting to a full PS analysis using the HERA PS pipeline, as an intermediate result, I compute the simple PS using a periodogram at each time bin for the geometric median and HERA mean location estimates of the LST-binned and redundantly grouped visibilities shown in Section 3.4.2.1. I plot the respective PS, as well as the comparison of their means (incoherent average in time), in Figure 3.17.

From Figure 3.17, the HERA mean approach performs better than its robust statistical counterpart. This may be because the HERA mean smoothes out the signal too much, especially considering that the visibility distribution is non-Gaussian; a clipping + mean estimator will tend to compute a location that is more smooth between frequencies (and other axes), whereas a robust estimator will choose between peaks of a multimodal distribution that slightly vary in strength, thus, resulting in a more amplified and varied signal.

---

**Figure 3.16:** Complex visibility location estimates with the HERA mean method (right column), with residuals compared to the geometric median (GM), Tukey median (TM) and marginal median (MM) also shown (first three columns). The amplitude, phase and $\Re$ and $\Im$ components are all shown. The visibility estimates are largely similar across all methods, although the displayed residuals for the robust estimators appear to differ in their results for certain broad regions (seen by the darker colours) in frequency/time space, with also some time bins being highlighted as slightly granting dissonant results, possibly due to them experiencing some low-level RFI.
Figure 3.17: PS of geometric median and HERA mean estimates of LST-binned + redundantly grouped visibilities for the 30 time bins between LSTs 5.28–5.46 h. The mean PS over time is shown and compared, with the HERA mean PS having a lower noise floor than its geometric median counterpart.

Figure 3.18: Absolute value of the CPS averaged across all baseline permutations for the geometric median and HERA mean estimates of LST-binned visibilities. The absolute value of the mean of all CPS over time is also shown and compared.
### Table 3.3: The mean smoothness for consecutive location estimates of redundant visibilities in time $\hat{S}_t$ and frequency $\hat{S}_\nu$, with the difference measure being the absolute value of consecutive values in the first two rows, while the last two rows compute Equation 3.44 while keeping the complex differences between consecutive values. A reminder that $\text{Var}[z] = \text{Var}[\Re(z)] + \text{Var}[\Im(z)], z \in \mathbb{C}$; therefore, strictly speaking, the last two rows are the more correct smoothness values. It is still, however, interesting to compute the smoothness of the distances (absolute values) to see if they are consistent with the ordering in the last two rows.

<table>
<thead>
<tr>
<th>Difference measure</th>
<th>Geometric Median</th>
<th>Tukey Median</th>
<th>Marginal Median</th>
<th>HERA Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{S}_t$</td>
<td>0.4161</td>
<td>0.42209</td>
<td>0.4630</td>
<td>0.3700</td>
</tr>
<tr>
<td>$\bar{S}_\nu$</td>
<td>0.2494</td>
<td>0.2558</td>
<td>0.2919</td>
<td>0.2124</td>
</tr>
<tr>
<td>$\bar{S}_t$</td>
<td>2.3160</td>
<td>2.3183</td>
<td>2.3348</td>
<td>2.3063</td>
</tr>
<tr>
<td>$\bar{S}_\nu$</td>
<td>0.5268</td>
<td>0.5358</td>
<td>0.6077</td>
<td>0.4528</td>
</tr>
</tbody>
</table>

I perform the double averaging across JDs and baselines in the visibility domain to demonstrate that 1) redundant baseline averaging works best through CPS and that 2) per-baseline systematics become more prominent using a robust estimator if forming auto-PS.

As an alternative approach that follows the steps of the HERA PS pipeline more closely, I run location estimates on the visibilities aggregated in LST. As fewer data points are combined for this averaging, there are (very few) gaps in the data that I fill using cubic interpolation. I then compute the (complex) CPS across all baseline permutations and incoherently mean average these in time, producing Figure 3.18. The PS in Figure 3.18 have lower noise floors by an order of magnitude than those in Figure 3.17, which is expected as the CPS suppresses noise that is specific to individual baselines. These results are more encouraging, and the geometric median performs just as well as the HERA mean. As more baselines and times are considered, the robust estimator will likely outperform the HERA mean if low levels of contamination are present.

I note that this entire analysis was repeated for frequency channels 175–333 (Band 1) and LSTs 9.59–9.76 h (a subset of Field 3) as a consistency check, with very similar results obtained.

### 3.4.3 H1C_IDR2.2 re-analysis

With the encouraging results of Section 3.4.2, I decided to fully implement the geometric median computation in the LST-binning pipeline. Re-running the LST-binning was expensive,
and multiprocessing was employed to speed up the computation. The preprocessing and PS steps were then executed, with the final spherically averaged dimensionless PS plotted in Figure 3.19, with the limits from The HERA Collaboration et al. (2022c) (see Figure 2.14 and Section 2.2.4) also shown for comparison.

As the geometric median is a less efficient estimator than the mean, we expect the final PS results to be slightly degraded, which is indeed the case for the majority of data points in Figure 3.19. Looking at the Band 2 Field 2 plot, we do not see a drop in $\Delta^2(k)$ to bring the PS closer to its estimated noise floor. This indicates that this anomalous result is not due to low-level transient RFI that is not repeated across JDs; we can rule out this particular class of RFI. Other types of RFI or systematics could be the cause of this effect, with more work needed to deduce its origin.

I find that the geometric median result sets a slightly better upper limit of $\Delta^2_{21}(k = 0.192 h \text{Mpc}^{-1}, z = 7.9) \leq (26.24)^2 \text{mK}^2$, compared to the limits from The HERA Collaboration et al. (2022c) (see Section 2.2.4). This is encouraging but not statistically significant; the original result finds an upper limit of $(30.76)^2 \text{mK}^2$ for the same $k$ and $z$ and the associated 1σ error is $(19.17)^2 \text{mK}^2$.

In the preprocessing pipeline, before inpainting, some additional LSTs and channels are flagged from further inspection in delay space. Wideband delay clean is used to filter redundantly averaged data by modelling sources from ±2000 ns. The residual visibilities are inspected
by eye, and recurring narrowband features appearing across baseline groups are flagged. Of these flagged channels, those in the range 570–573 coincide with Band 2. I re-ran the preprocessing and PS pipelines with channels 570–573 unflagged to see if this additional flagging could have had an effect on the PS; the comparison with the published results is shown in Figure 3.20.

We do not see any significant difference in the PS if channels 570–573 are flagged or not. Interestingly, though, the flagging of these channels impacts the clean model used for inpainting, which results in slight variations in some of the $\Delta^2(k)$ in Figure 3.20.

### 3.4.4 High-pass filtering

As an extension to averaging visibilities with robust location estimators, high-pass filters (HPFs) in delay space could be applied prior to the averaging to remove the low order modes of the data. Such delay-filtered data becomes Gaussian distributed (as evidenced by a low $HZ$ statistic), and the multimodality seen in some instances is removed. This aids in the averaging of visibilities. Any RFI or other types of contaminants are likely to still be present in the higher modes of the filtered visibilities, so a robust estimator is still advantageous in performing the averaging.

Delay filtering is classically done with the clean (Högbom 1974) algorithm. Other more recent methods, such as e.g. dayenu (Ewall-Wice et al. 2021), are also available, but are also more expensive. As a worked example, in Figure 3.21, I show sample data and clean model
visibilities for (37, 38, EE), a 14 m EW baseline, at LST 5.39 h on JD 2458098, using a filtered region $|\eta| \leq 0.5 \mu s$. Expanding to more times gives the heatmaps shown in Figure 3.22. The data – model = residuals are then taken forward for PS analysis. The Band 2 PS for the residual and original visibilities are shown in Figure 3.23. Any flagged channel is replaced by the visibility model; therefore, the residual visibility of a flagged channel is $0 + 0j$.

In removing the lower delay modes (i.e. foregrounds) from the visibilities, the distributions of visibilities aggregated both over JDS and baselines (and separately as well) become closer to being Gaussian, as demonstrated in Figure 3.24.

Repeating the PS computation from Figure 3.18 with the delay filtered visibilities, I obtain the spectra in Figure 3.25. The PS for the filtered data match those for the unfiltered visibilities.

A re-analysis of H1C IDR2.2 was done, with clean delay filtering at threshold $|\eta| \leq 0.5 \mu s$ conducted before the LST-binning stage, and geometric median used for LST averaging. The PS computation through the standard PS pipeline cannot, however, be conducted. This is because the cable reflection fitting and subsequent removal requires foreground components in the visibilities to be effective. More recent HERA data releases, which were observed with an upgraded array, do not have this feature, so this systematic removal step is not required. These newer datasets do, however, suffer from other issues. I also note that the filter threshold of $|\eta| \leq 0.5 \mu s$ is fairly strong and ends up filtering out the region of delay space with a lot of astronomical signal.

Delay filtering is a powerful tool, and, if used correctly in combination with a robust statistical estimator for averaging in the visibility domain, has potential for greatly improved results that would also not require the same level of manual adjustments to the data. HPF visibilities are closer to being Gaussian distributed, and multimodality is less present, meaning that robust estimators perform better and can improve on the results of classical estimators, especially as anomalous high-delay artefacts are expected to pervade HERA visibilities. Delay filtering requires further consideration and should be explored for future HERA data releases.

### 3.5 Gaussian process interpolation

After the LST-binning of visibilities, certain frequencies and times remain flagged due to RFI or other issues. Such flags are problematic in PS computation since the FT assumes cyclic data that is regularly sampled. If missing values in the discontinuous data are set to zero, they corrupt otherwise foreground-free modes by inducing strong sidelobes. In the HERA preprocessing pipeline, delay-based clean (Parsons & Backer 2009) is used to fill in the flagged data in a step called inpainting. Here, delay-space visibility models are created for each time integration and baseline separately, and flagged frequency/time slices are then substituted for the best guess of
3.5. Gaussian process interpolation

Figure 3.21: Observed data (dots) and clean model (lines; calculated for $|\eta| \leq 0.5 \mu s$) for visibilities from baseline (37, 38, EE) at LST 5.39 h on JD 2458098.

Figure 3.22: Heatmaps showing the delay filtering process for visibilities from baseline (37, 38, EE) between LSTs 5.39 – 5.74 h on JD 2458098.

Figure 3.23: PS of the observed and delay filtered visibilities, as plotted in Figure 3.21) for channels 513 – 696 (corresponding to frequencies 150 – 168 MHz). We see that the clean filtered visibilities match the original data, with $|PS_{\text{data}} - PS_{\text{HPF}}| \approx 10^{-6}$ Jy$^2$ Hz$^2$ outside of the filtered region.
Normality and robust location estimates of HERA visibilities

(a) HZ p-value and test outcome for LST-binned and redundantly aggregated HERA visibility data for baseline group (12, 13, EE), before MAD-clipping. The p-value threshold is set at 0.05. 90.0% of grouped visibility distributions are deemed normal from this statistic, a vast improvement on unfiltered data.

(b) Visibility distribution after delay filtering for baselines redundant to (12, 13, EE) at frequency channel 549 and LST 5.67 h on JD 2458116, which is the same data in the last plot of Figure 3.5. The distribution is closer to being normal, and there is no clear sign of multimodality.

Figure 3.24: Normality of HPF visibilities.

the signal that should have been in that slice. Any time integration that is fully flagged before inpainting remains fully flagged. We note that the accuracy of inpainting is inevitably degraded for wide-channel gaps, as studied in Aguirre et al. (2022).

As an alternative to deconvolution to fill in the missing data, I suggest that these data can be interpolated with a Gaussian process model, which is nonparametric and highly flexible. While computationally expensive, such an approach can wholly consider the complex visibilities across both frequency and time. See Rasmussen & Williams (2006) for a detailed review of Gaussian processes.

Gaussian process models have been used for gain smoothing and foreground modelling (Mertens et al. 2018; Ghosh et al. 2020; Kern & Liu 2021), and their use for inpainting have also been explored (Trott et al. 2020; Kern & Liu 2021); the HERA pipeline does not, however, presently use Gaussian process regression for inpainting. Current methods for visibility interpolation mostly rely on clean-based approaches or delay filtering techniques that impose smooth foregrounds. The clean algorithm assumes that the sky consists of a number of point sources that are modelled as delta functions, with each of these independently solved for each time and frequency. This approach, however, does not explicitly allow for continuity across times and frequencies. Gaussian process regression with input dimensions of both time and frequency can, however, ensure smoothness across these fields.

A Gaussian process distribution is defined by its mean function \( m(x) \) and its kernel function
### 3.5. Gaussian process interpolation

**Figure 3.25:** Absolute values of the CPS averages for both the filtered and unfiltered H1C_IDR2.2 14 m EW visibilities redundant to baseline (12, 13, EE), between LSTs 5.39 – 5.74 h. The averaging method across JDs is shown at the top of the first two plots.

\[ k(x, x') : \]

\[ y \sim \mathcal{GP}(m(x), k(x, x')) \]  

(3.46)

The mean function gives the mean at any position of the input space. The kernel function describes the covariance between points, i.e. how correlated the fluctuations are about the mean trend; it should be positive definite.

I choose a Gaussian process with a combined kernel taken to be the sum of the radial basis
function and white noise kernels, each given by

\[ k_{\text{RBF}}(x_i, x_j) = \sigma_A^2 \exp\left(-\frac{\|x_j - x_i\|^2}{2l^2}\right) \]  \hspace{1cm} (3.47)

\[ k_{\text{W}}(x_i, x_j) = \sigma_N^2 I_n \]  \hspace{1cm} (3.48)

where \( \sigma_A^2 \) is the overall variance (or amplitude squared), \( l \) is the length scale, \( \sigma_N^2 \) is the noise variance, \( I_n \) is the identity matrix (\( \sigma_N^2 I_n \) is, thus, a diagonal matrix with each element giving the variance of the individual random variable).

As an example dataset, I examine visibilities for baseline (1, 12, EE) (taken as it has more structure than 14 m EW baselines) for LSTs 5.39 – 5.74 h on JD 2458098 and run a 2D complex Gaussian process regression. A time and frequency slice of the regression solution are plotted in Figure 3.26.

To further test the aptness of Gaussian process interpolation, I take LST-averaged visibilities (post 5\( \sigma \) MAD-clipping) for baseline (1, 12, EE) and for Band 2 over LSTs 5.39 – 5.74 h, and artificially perforate the data with random flags (some full time integrations and channels are flagged, to mimic RFI and other common effects). I then use a Gaussian process model to fit for those flags, as shown in Figure 3.27. I further compare these Gaussian process inpainted visibilities to those that are filled in using a per-frequency \textsc{clean} method, which is the inpainting method used for the preprocessing of H1C_IDR2.2 data, in the PS domain in Figure 3.28.

All of these Gaussian process models require some hyperparameter refinement that is determined by minimizing the negative log-marginal-likelihood. In Equation 3.48, the length scale \( l \) controls the smoothness of the model function, while \( \sigma_N^2 \) determines the noise level of the data. In practice, hyperparameter exploration is done to find the optimal set. For a 1D complex Gaussian process for a given time slice, the validity of hyperparameter combinations by log-marginal-likelihood can be seen in Figure 3.29. Such an analysis can be beneficial in finding the characteristic properties of the hypersurface under investigation.

### 3.6 Discussion

I have visualized and shown that the aggregated underlying HERA H1C_IDR2.2 visibility data is mostly non-Gaussian, even after calibration, with various unconfirmed effects often causing the distribution to be multimodal; these are likely to originate from manufacturing variance and disturbance from neighbouring antennas, as well as low-level RFI that has eluded the anomaly detection algorithms. I introduced an outlier rejection method based on robust Mahalanobis distances calculated with MCD estimates, which is more rigorous than MAD-clipping. I also presented the geometric median, a robust multivariate location estimator, and applied it to
the aggregated visibilities in order to improve on the location estimates for these non-normal data. I found that, in general, the geometric median better finds the central location of the data compared to the clipping + mean approach used in the H1C_IDR2.2 LST-binning pipeline.

In the visibility KDE plots and videos, I notice that the spread of data between nominally redundant baselines is higher than that between days, possibly due to non-redundancies of the array. Direction-dependent effects could be at play here. It appears that even after the full analysis pipeline, different redundant baselines are still seeing different skies; this is an issue that ultimately needs to be fixed in the hardware. For this reason, robust estimators may not
Figure 3.27: Gaussian process inpainting of sample fully calibrated and LST-averaged (with the HERA mean) HERA visibilities for baseline (1, 12, EE) over Band 2 between LSTs 5.39 – 5.74 h.

Figure 3.28: CPS of Gaussian process and clean inpainted mock visibilities, where all 60 time integrations between LSTs 5.39 – 5.74 h for baseline (1, 12, EE) have been cross-multiplied and mean averaged, for Band 2. The clean model is calculated for delays $|\eta| \leq 2000$ ns, which is used for the inpainting in the standard preprocessing pipeline. The Gaussian process result matches that from clean well; Gaussian process inpainting, however, has the advantage of also producing good results in the visibility domain, whereas clean inpainting only focuses on PS compatibility.
work quite as desired due to the multimodality observed in this case, since such statistics tend to choose a peak rather than settle in between them. I refer to the robust redundant calibration detailed in Section 4.1 as a method of potentially better reconciling redundant baselines.

When looking at preliminary PS of the visibility estimates for a subset of Band 2 Field 2 for the 14 m baselines, estimates using the geometric median perform just as well as those from the standard pipeline that uses the HERA mean, in the CPS of Figure 3.18. Further extending this analysis to the full H1C_IDR2.2 dataset, as summarized in Figure 3.19, shows that the geometric median finds very similar results to the ones presented in The HERA Collaboration et al. (2022c), despite its marginally lower efficiency.

In order to ameliorate the robust averaging procedure, HPFing of the data was investigated. Such a step removes any multimodality in the data and seems to improve PS. Due to the promising results here, it would be advisable that further consideration is put into a full delay-filtered analysis of HERA visibility data, and for systematic removal methods (that currently require foregrounds) to be adapted.

As an alternative method to clean-based inpainting, multidimensional complex Gaussian process regression is studied to fill in flagged data. While expensive, this provides a flexible nonparametric inpainting method that achieves good performance in both the visibility and spectral domain.

As newer data releases are reduced that consist of many more visibilities (due to the growing number of antennas) measured with upgraded receivers, robust methods will be required to
mitigate the effect of RFI and other subtle effects. At these limit-pushing sensitivities, even
the finest of outliers can have a noticeably damaging effect; a robust analysis will ensure that
results are not limited by this.
Redundant calibration is a key step in the processing of HERA data. HERA’s array configuration is highly regular, and many baselines have the same length and orientation, thus forming redundant groups of baselines.\footnote{This is no coincidence, as dense and highly redundant arrays are optimal for the high sensitivity requirements of 21 cm detection (Morales 2005).} This arrangement can be exploited for calibration, as each baseline in a redundant group should, in theory, measure the same visibility. In practice, however, the visibilities from these baselines will differ due to the antenna-based gains. By applying this redundancy constraint, we can further constrain these antenna gains using this mathematical consistency (Wieringa 1992; Liu et al. 2010).

In the HERA calibration pipeline, redundant calibration is performed by assuming Gaussian noise statistics; existing algorithms based on this principle include logcal (Wieringa 1992), linca1 (Liu et al. 2010) and omnical (Zheng et al. 2014), which can all be found in the HERA calibration pipeline in the redca1 module in hera_cal\footnote{https://github.com/HERA-Team/hera_cal}. These techniques either require linearization of the measurement equation or use some approximation. See Dillon et al. (2020) for a comprehensive survey of these methods and their application to HERA. This chapter presents a generalized maximum likelihood approach to redundant calibration that can use non-Gaussian models, without the need for approximation. I show how this can be achieved at good computational performance with very little programming effort by repurposing open-source libraries intended for ML, in this case, JAX\footnote{https://github.com/google/jax}.

Miscalibration can easily overwhelm a putative true 21 cm signal that is expected to be $\sim 10^5$
times weaker than astrophysical foregrounds, especially because the chain of calibration steps is not local, i.e. one error can propagate more widely across frequencies, times or antennas, as constrained equations are solved and solution smoothing is applied. Employing robust statistics in calibration routines has potential to mitigate such ripple effects. Therefore, throughout this chapter, I present results from redundant calibration of H1C_IDR2.2 HERA visibilities with Gaussian and Cauchy assumed noise distributions. The latter is a more robust distribution: it is median-centred and has fatter tails, and is, thus, expected to have increased resilience to RFI or other outliers.

The outline of this chapter is as follows: in Section 4.1, I review redundant calibration and extend its maximum likelihood estimation (MLE) to a Cauchy model. In Section 4.2, I show how the generalized redundant calibration minimization problem presented can be significantly and easily sped up with JAX. To test the stability of the interferometer, in Section 4.4 I show how redundant calibration visibility solutions can be compared across JDs by solving for the degenerate parameter offsets between them. Finally, as an extension of single-day redundant calibration, in Section 4.5, I present a unified solver that redundantly calibrates across all JDs.

Some of the work covered in this chapter is presented in Molnar & Nikolic (2020, 2021b). The code for the work described in this section can also be found in the simpleredcal\textsuperscript{d} PYTHON package.

Unless otherwise specified, example results in this chapter use HERA data observed on JD 2458098 between LSTs 5.28 – 5.46 h from the H1C_IDR2.2 dataset. Only the EE polarization is looked at.

### 4.1 Relative redundant calibration

An array with regularly spaced antennas has many redundant visibilities that are sensitive to the same modes on the sky. Redundant calibration uses the fact that the true visibilities from redundant baselines are equal. Supposing that there are no direction-dependent calibration effects, we hence have a system of equations for all antenna pairs \( i \) and \( j \):

\[
V_{ij}^{\text{obs}}(\nu) = g_i(\nu)g_j^*(\nu)U_\alpha(\nu) + n_{ij}(\nu) \tag{4.1}
\]

where \( U_\alpha(\nu) = V(r_i - r_j) \), the visibility for the baseline vector \( b_{ij} = r_i - r_j \), corresponds to a redundant baseline set that we index by \( \alpha \), and \( n_{ij} \) is the noise.

The full HERA array consists of 320 elements in the hexagonal core, equating to \( N_{bl} = 320 \times (320 - 1)/2 = 51,040 \) baselines. The core only has 1501 unique baselines, which means

\textsuperscript{d}https://github.com/bnikolic/simpleredcal
that we have a non-linear system of 51,040 equations to determine the 1501 true visibilities and 320 antenna gains.

With the redundant calibration prior, an MLE for the gains and true visibilities can be constructed by assuming a distribution for the observed visibility noise.

4.1.1 MLE under Gaussian noise

Assuming Gaussian uncorrelated noise with variance $\sigma_{ij}^2$, which is the expected noise from the receivers and the sky, through MLE considerations, the gains and true visibilities can be found by minimizing the following negative log-likelihood function:

$$-\ln(\mathcal{L}_{\text{rel}}^G(\nu)) = \frac{1}{2} \sum_{\alpha} \sum_{\{i,j\}_\alpha} \ln(2\pi\sigma_{ij}^2(\nu)) + \frac{|V_{ij}^{\text{obs}}(\nu) - g_i(\nu)g_j^*(\nu)U_{\alpha}(\nu)|^2}{\sigma_{ij}^2(\nu)}$$

where $\{i, j\}_\alpha$ are sets of antennas that belong to baseline group $\alpha$. This minimization is equivalent to that of the chi-squared:

$$\chi^2_{\text{rel}}(\nu) = \sum_{\alpha} \sum_{\{i,j\}_\alpha} \frac{|V_{ij}^{\text{obs}}(\nu) - g_i(\nu)g_j^*(\nu)U_{\alpha}(\nu)|^2}{\sigma_{ij}^2(\nu)}$$

This non-linear least-squares optimization can be done independently between frequencies and time. The $\chi^2$ minimization (Equation 4.3) forms the basis of most redundant calibration formalisms (see e.g. Byrne et al. 2019), with most current efforts opting to linearize Equation 4.1 (Dillon et al. 2020). With the computational techniques presented in this chapter, I will perform the full negative log-likelihood minimization of Equation 4.2 without the need for approximations.

I find that, in practice, as with the rest of the minimizations encountered in this chapter, the solutions of the previous iteration can be used to initialize the next solve, which greatly speeds up the computation.

4.1.2 MLE under Cauchy noise

We can extend the MLE analysis to different distributions for the visibility noise. As an example, we can assume a Cauchy (also known as a Lorentz) distribution for the visibility noise; it is
Chapter 4. Robust redundant calibration

given by

$$f(x; x_0, \gamma) = \frac{1}{\pi \gamma \left[ 1 + \left( \frac{x - x_0}{\gamma} \right)^2 \right]}$$  \hspace{1cm} (4.4)$$

where $x_0$ and $\gamma$ are the Cauchy distribution’s location and scale parameters. The peak and median of the distribution are at $x_0$, and $\gamma$ specifies the half-width at half-maximum (HWHM). These are both robust measures of central tendency and statistical dispersion, respectively, that are not unduly affected by outliers. The mean of the Cauchy distribution is undefined, since it does not have finite moments of any order. Using such a distribution reduces the effect of outliers, for instance those caused by RFI.

In Figure 4.1, I fit both the Cauchy and Gaussian distributions to raw (uncalibrated) visibility amplitudes from sample redundant baselines taken from a narrow frequency cross-section, for comparison. A case with outliers is also plotted in Figure 4.2. In the presence of RFI, a Cauchy model is superior as its location and scale parameters are not as swayed by outliers as those for the Gaussian. However, in the majority of cases, a Gaussian model better represents the data, as expected.

If we assume Cauchy distributed data, the negative log-likelihood when solving for redundant baseline sets is given by

$$- \ln(L^C_{\text{rel}})(\nu) = \sum_{\alpha} \sum_{\{i,j\}_\alpha} \ln(\pi \gamma_{ij}(\nu)) + \ln \left( 1 + \left( \frac{V_{\text{obs}}_{ij}(\nu) - g_i(\nu)g^*_j(\nu)U_\alpha(\nu)}{\gamma_{ij}(\nu)} \right)^2 \right)$$  \hspace{1cm} (4.5)$$

This MLE with Cauchy-distributed noise fully encapsulates the distribution of the data without being distorted by outliers and is the best median estimator of the data.

The advantage of the Cauchy distribution in fitting for redundant baselines is not clear-cut. While it does reduce the impact of outliers, it only performs better than the Gaussian in cases that deviate slightly from normality. This is expected, as the median is a less efficient estimator than the mean (see Section 3.2.3). I quantify this discrepancy through the negative log-likelihood when fitting the redundant visibility amplitudes at a given time for both distributions, as can be seen in Figure 4.3. However, the use of the Cauchy could still be appreciable, especially if weak RFI is not picked up by flagging. This would corrupt estimates as the observations are integrated down.

In reality, we expect the ‘true’ probability distribution of visibilities to be a sum of Gaussian and Cauchy distributions, which, combined, would describe the thermal and outlier elements observed by the receiver.
4.1. Relative redundant calibration

Figure 4.1: Histograms of visibility amplitudes for baselines redundant to (12, 13, EE) for channels 620 – 624 at LST 5.34 h, with Gaussian (N) and Cauchy (C) probability density functions fitted to these data. The Cauchy fits are sharper and have fatter tails than their Gaussian counterparts. The sample size is fairly small, with only 24 visibilities in this redundant set.

Figure 4.2: Same as in Figure 4.1, but with frequency channels 700 and 701, the former of which contains outliers that greatly and inordinately affect the Gaussian fitting. While we expect that such extreme outliers would be flagged in the HERA analysis pipeline, this plot still demonstrates how (even marginal) outliers can significantly affect any Gaussian statistic. The Cauchy distribution is robust to such aberrations.
Chapter 4. Robust redundant calibration

![Figure 4.3](image)

Figure 4.3: Negative log-likelihood from MLE fitting of Gaussian, Cauchy and \( t \)-distributions, for the amplitudes of visibilities redundant to (12, 13, EE) at LST 5.34 h on JD 2458098. It is clear that at channel 700, which contains outliers, the Cauchy distribution provides a better fit to the data, as is also seen in Figure 4.2.

4.1.3 Constraining degeneracies

Relative calibration yields solutions with degeneracies that can be parameterized as four terms per frequency: an overall amplitude \( A(\nu) \), an overall phase \( \Delta(\nu) \), and two tip-tilts (or phase gradient components) \( \Delta_x(\nu) \) and \( \Delta_y(\nu) \). This can be seen by considering the below transformations of these degenerate parameters, which leaves \( -\ln(L_{\text{rel}}) \) (for both the Gaussian and Cauchy distributions) unchanged:

1) Overall amplitude
   - \( g_t \rightarrow A g_t \)
   - \( U_\alpha \rightarrow A^{-2} U_\alpha \)

2) Overall phase
   - \( g_k = |g_k| e^{i \phi_k} \rightarrow |g_k| e^{i \phi_k + \Delta} \)
   - such that \( g_k g_l^* = |g_k||g_l| e^{i(\phi_k - \phi_l)} \rightarrow |g_k||g_l| e^{i(\phi_k + \Delta - \phi_l - \Delta)} = g_k g_l^* \)

3) Phase gradients
   - \( g_k = |g_k| e^{i \phi_k} \rightarrow |g_k| e^{i(\phi_k + \Delta_x x_k + \Delta_y y_k)} \)
4.1. Relative redundant calibration

\[ U_\alpha = |U_\alpha|e^{i\phi_\alpha} \rightarrow |U_\alpha|e^{i(\phi_\alpha - \Delta_x x_\alpha - \Delta_y y_\alpha)} \]

where we assume the array to be co-planar, with \((x_k, y_k)\) the positional coordinates of antenna \(k\), and \((x_\alpha, y_\alpha)\) the separations of the antennas that form baselines in redundant set \(\alpha\). This degeneracy applies to each solution interval separately, i.e. for each time, frequency and day slice of data. Hence, relative redundant calibration without constraints is likely to converge to different parts of this space for each run.

The degenerate parameters can be calculated from a sky model in an absolute calibration step. Alternatively, we can fix these degenerate parameters when minimizing \(-\ln(L_{rel})\) by applying a few conditions (Byrne 2019). This method, however, still ultimately needs to reference the sky to set the flux scale and phase centre.

Fixing these degenerate parameters becomes a constrained minimization problem, which can be done with common optimization software (e.g. minimization in SciPy with scipy.optimize.minimize). Imposing constraints, however, tend to slow down the solver, and can sometimes produce erroneous results; classically, the parameterization of the problem can be reduced, and the constraints implicitly added into the function being minimized. Below, I review the constrained minimization that redundantly calibrates while also fixing the degeneracies, and in Section 4.2 I show how this computation can be significantly sped-up in JAX, without having to re-parameterize.

We first define a set of parameters \(h_i\) to be the gains that obey the following constraints:

\[
\frac{1}{N} \sum_{i=1}^{N} |h_i| = 1 \quad \rightarrow \quad \text{mean gain amplitude of 1} \quad (4.6)
\]

\[
\frac{1}{N} \sum_{i=1}^{N} \text{Arg}(h_i) = 0 \quad \rightarrow \quad \text{mean gain phase of 0} \quad (4.7)
\]

\[
\begin{aligned}
\sum_{i=1}^{N} x_i \text{Arg}(h_i) &= 0 \\
\sum_{i=1}^{N} y_i \text{Arg}(h_i) &= 0 \\
\end{aligned}
\]

\[
\quad \rightarrow \quad \text{phase gradients of 0} \quad (4.8)
\]

such that the antenna gains can be written as

\[
g_i(\nu) = A(\nu)e^{i\left[\Delta(\nu) + \Delta_x (\nu)x_i + \Delta_y (\nu)y_i\right]}h_i(\nu) \quad (4.9)
\]

where \((x_i, y_i)\) is the position of antenna \(i\), so that all degenerate dependencies are removed from \(h_i\). We note that these constraints (RHS of Equations 4.6 to 4.8) are arbitrary; however, the ones imposed here follow convention and are easy to work with. Non-degenerate formulations of Equations 4.2 and 4.5 are, therefore, given by

\[
-\ln(L_{\text{constr}}^G)(\nu) = \frac{1}{2} \sum_{\alpha} \sum_{(i,j),\alpha} \ln(2\pi\sigma_{ij}^2(\nu)) + \left| \frac{V_{i,j}\text{obs}(\nu) - h_i(\nu)h_j^*(\nu)W_{\nu}(\nu)}{\sigma_{ij}^2(\nu)} \right|^2 \quad (4.10)
\]
Chapter 4. Robust redundant calibration

\[-\ln(L^C_{\text{constr}})(v) = \sum_{\alpha} \sum_{\{i,j\}_\alpha} \ln(\pi \gamma_{ij}(v)) + \ln \left( 1 + \left( \frac{V_{ij}^{\text{obs}}(v) - h_i(v) h_j^*(v) W_\alpha(v)}{\gamma_{ij}(v)} \right)^2 \right)\] (4.11)

where

\[W_\alpha(v) = A^2(v) e^{\Delta_x(x_\alpha, y_\alpha)} U_\alpha\] (4.12)

and \((x_\alpha, y_\alpha)\) are the baseline coordinates of redundant set \(\alpha\).

The overall phase is also degenerate and is set by requiring that the phase of the gain of a reference antenna is null: \(\text{Arg}(h_{\text{ref}}) = 0\); it is an arbitrary convention with no physical significance.

Applying such constraints from the outset can be beneficial since it ensures that all visibility solutions (from different frequencies or times) coexist in the same degenerate space, so they can be directly compared. Practically, minimization of Equations 4.10 and 4.11 can be done with the trust-constr SciPy minimization method, which uses a trust-region algorithm to conduct nonlinear constrained optimization (see e.g. Conn et al. 2000 for a comprehensive overview of trust-region methods). This is, however, an expensive computation, and the algorithm can sometimes struggle with the rigidity of all these constraints and fail (this is usually observed for problematic data slices).

4.1.4 Student’s \(t\)-distribution

Other examples of robust distributions that could be used to fit redundant sets of visibilities include the Student’s \(t\)-distribution, which has probability density function given by

\[f(t; \nu, x_0, \gamma) = \frac{\Gamma \left( \frac{\nu + 1}{2} \right)}{\sqrt{\pi} \nu \Gamma \left( \frac{\nu}{2} \right) \gamma} \left( 1 + \frac{1}{\nu} \left( \frac{t - x_0}{\gamma} \right)^2 \right)^{-\frac{\nu + 1}{2}}\] (4.13)

where \(\Gamma\) is the gamma function and \(\nu\) is the number of degrees of freedom (also known as the normality parameter, since \(f(t)\) approaches a Gaussian as \(\nu\) increases). For \(\nu = 1\), the \(t\)-distribution coincides with the Cauchy distribution.
4.1. Relative redundant calibration

Figure 4.4: Student’s t-distribution for degrees of freedom $1 \leq \nu \leq 5$ and $\nu = 20$, with location parameter $x_0 = 0$ and scale parameter $\gamma = 2$. A normal distribution with $\mu = 0$ and $\sigma = 2$ is also plotted for comparison. From this plot, it is clear that the tails of the Cauchy distribution ($\nu = 1$) are fatter than those of the Gaussian, with the tails being tapered as $\nu$ increases.

For the first few degrees of freedom, the t-distribution is given by

$$ f_{\nu=2}(t; x_0, \gamma) = \frac{1}{2\sqrt{2\gamma}} \left( 1 + \frac{1}{2} \left( \frac{t-x_0}{\gamma} \right)^2 \right)^{-\frac{3}{2}} $$

(4.14)

$$ f_{\nu=3}(t; x_0, \gamma) = \frac{2}{\sqrt{3\pi\gamma}} \left( 1 + \frac{1}{3} \left( \frac{t-x_0}{\gamma} \right)^2 \right)^{-\frac{5}{2}} $$

(4.15)

$$ f_{\nu=4}(t; x_0, \gamma) = \frac{3}{8\gamma} \left( 1 + \frac{1}{4} \left( \frac{t-x_0}{\gamma} \right)^2 \right)^{-\frac{7}{2}} $$

(4.16)

$$ f_{\nu=5}(t; x_0, \gamma) = \frac{8}{3\sqrt{5\pi\gamma}} \left( 1 + \frac{1}{5} \left( \frac{t-x_0}{\gamma} \right)^2 \right)^{-\frac{9}{2}} $$

(4.17)

As $\nu$ increases, the t-distribution’s tails become thinner and the peak sharper, with the t-distribution becoming a Gaussian as $\nu \to \infty$. The probability density functions for the first few degrees of freedom at a set location and scale parameter are shown in Figure 4.4.

We note that for $\nu > 1$, $x_0$ is both the mean and the median; the mean is otherwise undefined.
For $1 < \nu \leq 2$ and $\nu > 2$, the variance is infinite and $\frac{\nu}{\nu-2}$, respectively, and otherwise undefined.

For an MLE of the parameters of the $t$-distribution when fitting to a dataset, we need to minimize

$$-\ln(L^t)(\nu, x_0, \gamma) = -N \ln \left( \Gamma \left( \frac{\nu + 1}{2} \right) \right) + \frac{N}{2} \ln \left( \pi \nu \right) + N \ln \left( \Gamma \left( \frac{\nu}{2} \right) \right) + N \ln (\gamma) + \frac{\nu + 1}{2} \sum_i \ln \left( 1 + \frac{1}{\nu} \left( \frac{t_i - x_0}{\gamma} \right)^2 \right)$$

(4.18)

In practice, we would already fix the degrees of freedom parameter $\nu$, when fitting for the $t$-distribution.

The negative log-likelihood for the MLE fitting of visibility amplitudes for a redundant set of baselines using the $t$-distribution with $\nu = 3$ is also shown in Figure 4.3. It is clear that the $t$-distribution is a trade-off (that can be modulated by varying $\nu$) between the Cauchy and the Gaussian distributions, i.e. between robustness and efficiency.

Robust interferometric calibration using a complex Student’s $t$-distribution is also presented in Sob et al. (2020), where it is shown that such a robust framework can mitigate the effects of unmodelled sources and RFI during calibration. The focus in this research is on sky-based calibration for the purpose of imaging, which is in contrast to the redundant calibration in this chapter that has 21 cm PS computation as its ultimate objective.

### 4.2 JAX

Redundant calibration is computationally expensive. As we depart from Gaussianity, we can no longer use the Levenberg–Marquardt algorithm for non-linear least squares problems (as is done in Grobler et al. 2018), which further slows down calculations. In addition, if we wish to programmatically add constraints, such computations become even slower as to make them unfeasible.

I introduce JAX, a Google project that provides a framework for high-performance array computations on accelerated hardware by bringing together automatic differentiation (AD) and Accelerated Linear Algebra (XLA) to augment NumPy and Python code, therefore, greatly speeding up operations common in ML. JAX has its own version of NumPy, which is (for most purposes) directly interchangeable with the original version. Since JAX implements the NumPy API, it makes it straightforward to start working with JAX and to already significantly accelerate existing code with very few changes.

---

8https://github.com/google/jax  
9https://numpy.org/  
10JAX has its own version of SciPy too for that matter, although this is less complete than its NumPy peer.
4.2.1 Automatic differentiation

AD is a powerful method to calculate derivatives used in many ML libraries (e.g. TensorFlow\(^1\) and PyTorch\(^2\)). It is preferable to more traditional methods, such as finite and symbolic differentiation, which perform poorly when applied to complex mathematical functions and when evaluating high derivatives (Baydin et al. 2018). AD refers to a general set of techniques that compute the derivatives of functions by exploiting the fact that computer programs ultimately execute a series of elementary arithmetic operations on elementary functions that can be combined with the chain rule to compute complicated derivatives of arbitrary order.

The established computational methods for differentiation are:

**Symbolic differentiation:**

Here, algebraic expressions of functions can be differentiated by applying the rules for combined functions (e.g. linearity, product rule, chain rule etc.) to a database of basic functions (polynomials, trigonometric functions, exponentials and logarithms etc.) for which their derivatives are known. The resulting symbolic expressions, however, may grow exponentially large, producing unwieldy formulae that become inefficient to evaluate, in a problem known as expression swell.

**Numerical differentiation:**

Also known as finite differences (FD), this method is based on the limit definition of the derivative. FD is computationally expensive, as it requires $O(n)$ evaluations of the function for a gradient in $n$ dimensions. Moreover, truncation errors and round-off errors due to machine precision make FD ill-conditioned and unstable. This is what SciPy and other common libraries use by default.

**Automatic differentiation:**

Much like symbolic differentiation, AD decomposes complex expressions through the systematic application of the chain rule into a finite sequence of primitive operations and elementary functions for which derivatives are known. The combination of these constituent derivatives gives the derivative of the overall expression (Verma 2000). Unlike symbolic differentiation, however, the chain rule is applied to actual numerical values instead of symbolic expressions. With symbolic differentiation suffering from expression swell and FD from precision errors, AD emerges as a powerful technique that yields exact results, unlike FD, at greater speed and

\(^1\)https://www.tensorflow.org/
\(^2\)https://pytorch.org/
memory efficiency than symbolic differentiation. See Griewank & Walther (2008); Margossian (2019) for detailed reviews of AD.

When computing the derivative of a function, a computer program will execute a sequence of internal operations, which are recorded in an evaluation trace. These form the foundation of AD. Reused subexpressions are algorithmically exploited, which leads to a more efficient evaluation of functions and their derivatives.

**Forward-mode:**

To compute \( \frac{\partial f}{\partial x_1} \), we associate each intermediate variable \( v_i \) with
\[
\dot{v}_i = \frac{\partial v_i}{\partial x_1}
\] (4.19)

Applying the chain rule to each elementary operation in the primal trace (which evaluates \( f \)), we then generate the corresponding tangent trace. Evaluating the primals \( v_i \) and their corresponding derivatives simultaneously and in order gives us the required derivative in the output variable.

**Reverse-mode:**

Derivatives are propagated backwards from a given output. This is a two-phase process whereby all intermediate variables and their dependencies are initially populated, with their values stored in memory, in a forward phase. A second phase follows where derivatives are calculated by propagating adjoints \( \bar{v}_i \) in reverse, from the outputs to the inputs, with each adjoint complementing their respective intermediate variable like so:
\[
\bar{v}_i = \frac{\partial y_j}{\partial v_i}
\] (4.20)

Reverse-mode has the advantage that it is considerably less expensive to evaluate than the forward-mode for functions with more inputs than outputs, whereas forward-mode is more efficient for a function with fewer inputs than outputs. For most purposes, reverse-mode is preferred; however, forward-mode is useful as it can be used to compute Hessians when combined with reverse-mode.

To get a better understanding of AD, I look at an example computational graph of JAX’s intermediate representation (IR) to visualize its program tracing. I consider taking the derivative of the function \( f(x_1, x_2) = \ln(x_1) + x_1x_2 - \sin(x_2) \):\(^1\)

\(^1\)This is the same example used in Section 3 of Baydin et al. (2018), in which I refer to Figure 4 for a computational graph of the primal trace and Table 3 for the evaluations of the forward primal and reverse adjoint traces.
```python
import jax.numpy as jnp

from jax import grad  # take derivatives with AD; built on reverse-mode

def f(x1, x2):
    return jnp.log(x1) + x1*x2 - jnp.sin(x2)

grad_f1 = grad(f, argnums=0)  # derivative wrt x1
print(grad_f1(2.0, 5.0))  # 5.5

JAX’s vectorization map `jax.vmap` can also be used to vectorize the above function, so that the output of `grad_f1` can be evaluated in parallel (over some axis) for some input array:

```python
from jax import vmap

def gradv(x1, x2, argnum):
    return vmap(grad(f, argnums=argnum))(x1, x2)

grad_vf1 = gradv(jnp.arange(0, 10, dtype=float), jnp.arange(10, 20, dtype=float), 0)
print(grad_vf1)
# [ inf  12.  12.5 13.33333333 14.25
#  15.2  16.16666667 17.14285714 18.125 19.11111111 ]
```

The computational graph for the above code for vectorized gradient evaluation is shown in Figure 4.5. These graphs, however, quickly swell up as the program becomes more complicated, and sifting through them becomes rapidly intractable.

### 4.2.2 JIT and XLA

JAX offers just-in-time (JIT) compilation via `jax.jit` (used as a decorator or explicit function), such that standard Python and NumPy functions run efficiently with minimal coding overhead. While code performance using JIT is greatly improved when running on accelerators, there is also a noticeable difference on CPUs. With JIT, JAX will compile the function when it is called for the first time on the fly (or just-in-time) into a faster form, and will use the optimized version from the second call onwards.

JAX compiles functions with XLA, which is a domain-specific compiler for linear algebra created by TensorFlow. XLA analyses the computational graph for a given program, specializes
Figure 4.5: Computational graph from JAX's High-Level Optimizer (HLO) IR. The forward evaluation of the primals is in the left branch of the graph, while the reverse adjoint tangent operations are on the right. For a more complicated derivative, the computational graph would be deeper, and the intermediate results from the primal trace would link into the reverse adjoint trace, thus, further speeding up the computation.

it for the actual runtime dimensions and types, fuses multiple operations into a single kernel, and outputs efficient native machine code, without having to write intermediate results to memory.

JIT compilation and AD in JAX can be composed arbitrarily,\(^m\) thus, allowing for the creation of sophisticated programs that run at maximum performance, all while remaining in Python.

4.3 Results for generalized redundant calibration

I conduct relative redundant calibration with both Gaussian and Cauchy assumed noise (see Sections 4.1.1 and 4.1.2) to a) compare the two methods in the visibility domain and b) show that with JAX acceleration, we can handle such MLE with considerable speed-up and ease (compared to pure NumPy and SciPy), even when moving away from Gaussianity and without the need to linearize or approximate.

In the results of Sections 4.3.1 and 4.3.2, I assume that the noise in the negative log-likelihoods for relative redundant calibration is uniform across baselines. Therefore, I set \(\sigma_{ij}\) and \(\gamma_{ij}\) to unity in Equations 4.2 and 4.5, respectively. The inclusion of a non-constant noise term comes at additional computational cost; this is later investigated in Section 4.3.3.

\(^m\)As well as vectorization (jax.vmap) and parallelization (jax.pmap).
4.3. Results for generalized redundant calibration

4.3.1 Redundant visibility solutions

I consider raw visibilities from JDs 2458098 and 2458099 over LSTs 5.28 – 5.46 h,\(^n\) and run relative redundant calibration on all baseline groups that have at least two baselines in them. The redundant calibration is done on these two JDs, as they will eventually be compared through degenerate transformation in Section 4.4. See Appendix B.2 for the mock implementation of this redundant calibration process for a single JD, frequency and time bin. By default, I normalize the relative redundant calibration solutions after they are found such that the average gain amplitude is 1, to make them more digestible; this is allowed since the overall amplitude is a degenerate parameter in Section 4.1.3.

The minimum negative log-likelihoods \(-\ln(L_{rel})\) for the relative redundant calibration of the selected visibilities, with Gaussian assumed noise, are plotted in the left column of Figure 4.6. I also use the median (over baselines) absolute normalized residual for these solutions, given by

\[
R_{\text{man}} = \text{med}_{\text{bls}} \left( \frac{|V_{\text{meas}} - V_{\text{pred}}|}{\sqrt{|V_{\text{meas}}| |V_{\text{pred}}|}} \right)
\]

which proxies as a goodness of fit measure for this minimization (this is the median of the absolute values of Figure B.3, but for each frequency/time slice). These are also shown in the right column of Figure 4.6.

In Figure 4.6, the dark vertical lines correspond to RFI from known bad frequency channels. The negative log-likelihoods in these plots also decrease as the frequency channel increases; this is due to the natural shape of radio visibilities whose amplitudes decrease at higher frequencies. Results with Cauchy assumed noise are much the same for these heatmaps.

The corresponding visibility solutions \(U_{\alpha}\) for the 14 m EW baseline on JD 2458098 for Gaussian assumed noise is shown in the top plot of Figure 4.7. Comparison with the solutions from a solve with Cauchy assumed noise is shown in the bottom plot. MLE with the two noise distributions grant consistent results, and the difference in the visibilities (in this particular instance) is marginal, especially in the bands used for PS estimation. Any difference in the calibration solutions is also lower bounded by the stopping criterion in the redundant calibration minimization.

The gain solutions can also be examined as another diagnostic. In Figure 4.8, the gain amplitudes are plotted, as well as the absolute difference in the gains between the two distributions.

\(^n\)Datasets from H1C_IDR2.2 are stored in HDF5 file format (https://github.com/HDFGroup/hdf5), with each file containing 60 \(\times\) 10.7 s time integrations. These datasets are recorded nightly at the same fractional JD, which means that some alignment is required if they are to match in LST. In this example case, I calibrate datasets labelled by 2458098.43869, 2458099.43124 and 2458099.43869; the latter two datasets are required to cover the LST range spanned by the former.
Chapter 4. Robust redundant calibration

(a) Minimum $-\ln(L_{\text{rel}})$ for the redundant calibration of HERA data on JD 2458098 between LSTs 5.28 – 5.46 h.

(b) $R_{\text{man}}$ for the redundant calibration of HERA data on JD 2458098 between LSTs 5.28 – 5.46 h.

(c) Minimum $-\ln(L_{\text{rel}})$ for the redundant calibration of HERA data on JD 2458099 between LSTs 5.28 – 5.46 h.

(d) $R_{\text{man}}$ for the redundant calibration of HERA data on JD 2458099 between LSTs 5.28 – 5.46 h.

Figure 4.6: Minimum $-\ln(L_{\text{rel}})$ (left) and $R_{\text{man}}$ (right) for the redundant calibration of visibilities on JDs 2458098 (Figures 4.6a and 4.6b) and 2458099 (Figures 4.6c and 4.6d) between LSTs 5.28 – 5.46 h.

for frequency channel 250; any recurrent and significant deviation in gains could be used to identify faulty antennas. Gain solutions here are consistent.

There are many ways of slicing and visualizing the data. Even then, the difference between the redundant calibration results (i.e. the visibility solutions $U_{\alpha}$, the gains $g_{i}$ and the calibrated visibilities $V_{ij}/(g_{i}g_{j}^\star)$) using Gaussian and Cauchy assumed noise is very slight in the overwhelming majority of instances, especially when the data is narrowed down to frequencies and times that are not flagged by the end of the HERA analysis pipeline. As an example, I consider visibilities for frequency channel 282 at LST 5.34 h (this is a data slice that is eventually flagged). I show the raw data and redundantly calibrated data assuming Gaussian noise, by baseline group, in Figures 4.9 and 4.10. The raw data does not seem anomalous, but the calibration solutions push certain baselines with common antennas away from the distribution
4.3. Results for generalized redundant calibration

![Figure 4.7](image)

Figure 4.7: Top: redundant visibility solutions for the 14 m EW baseline for Gaussian assumed noise on JD 2458098 between LSTs 5.28 – 5.46 h. The results are very similar for Cauchy assumed noise. Bottom: absolute residual between the redundant visibility solutions for Gaussian and Cauchy assumed noise. Such a comparison is more complicated than simply subtracting one dataset from the other, as it requires one dataset to first be transformed into the degenerate space of the other; this is discussed later in Section 4.4. As a reminder, the two frequency bands used for H1C_IDR2.2 PS estimation are also shown.

centre of their respective groups. For example, visibilities with antenna 124 are the farthest from the distribution centres in baseline groups 0, 2, 3, 11, 13, 17, 19, 24, 32, 34, 50, 66.

I zoom in on baseline group 0, i.e. baselines redundant to (1, 12, EE), and also plot the calibration solutions for a Cauchy redundant solve in Figure 4.11. The Gaussian solutions were transformed to match the degenerate parameters of those from the Cauchy solve (see Section 4.4 for an explanation as to why this is required). This particular data slice is chosen because there are a couple of significant outlier points where the difference in the calibration solutions between the two distributions is more noticeable. The $-\ln(L_{\text{rel}})$ in both redundant solves is also very high at $\approx 1.36$, while it is $\approx 0.09$ for other healthy slices at that frequency for neighbouring times (see Figure 4.6a); this slice is flagged in the HERA analysis pipeline, but I use it here for illustrative purposes. The calibrated visibilities in both solves are consistent to $< 0.5\%$. 


Figure 4.8: Gains in this plot are for frequency channel 250 on JD 2458098 and have been mean averaged in time over $10 \times 10.7$ s integrations between LSTs $5.37 - 5.30$ h, with all times having good data that is successfully redundantly calibrated. Note that this averaging can only be done if the gains live in the same degenerate space (see Section 4.4). Gains are stable over such time scales: there is evidence for intrinsic gain variation on 6 h and 10 MHz scales (Dillon et al. 2020; Kern et al. 2020b). Left: amplitude of gain solutions from redundant calibration with Gaussian assumed noise. Right: absolute difference in the complex gains solutions from redundant calibration with Gaussian and Cauchy assumed noise distributions.

Location estimates for the calibrated visibilities are plotted, to demonstrate that they differ from the solved visibility solutions $U_{\alpha}$ for the redundant group; this is because antenna gains, and consequently the calibrated visibilities, are influenced by all the other baseline groups, too. This is also the reason as to why the calibration has cast some visibilities in this baseline group as outliers, despite the raw data having little scatter.

Solving the non-linear system of equations in Equation 4.1 with MLE under Gaussian or Cauchy noise is a complicated process that yields very similar solutions for both distributions, possibly because the system is so overdetermined. For the dataset on JD 2458098 that spans LSTs $5.28 - 5.46$ h, there are few data slices where the difference in the calibration solutions is appreciable (if we only consider data from frequency channels that are not fully flagged following the analysis pipeline). There are still situations where the use of the Cauchy should help, the impact of which could be recognized if more data was calibrated. As we push to deeper integrations to set more stringent limits on the 21 cm PS, our calibration solutions need to be as good as possible, and the use of a robust noise distribution in such a calibration process avoids ‘baking in’ problems into the data by absorbing an RFI glitch or other flaw into the calibration solutions; such issues are very difficult to identify and remove, so it is best to reduce their impact as much as possible at the outset. This robust approach may work more effectively once
antennas with high $\chi^2/\ln(L_{rel})$ for each data slice is removed, and the redundant calibration is redone with the anomalous antennas excised. In Section 4.5, I explore a unified framework that considers data for all JDs at once, which is the best way of dealing with such anomalies; this way, bad antennas for a particular JD are automatically discarded.

4.3.2 Performance

The timings for relative redundant calibration that uses the unbounded BFGS minimization method, with and without JAX acceleration that employs both JIT and AD, are summarized in Table 4.1. As another test, the redundant calibration with the constraining of degeneracies (see Section 4.1.3) is also timed; this employs the trust-region minimization method where both
bounds and constraints on the solution can be set. The trust-region method in SciPy calls upon Hessian computation, whereas BFGS only requires evaluation of the Jacobian (the Hessian is then iteratively approximated).

Caution is needed so as not to mix and match JAX and pure NumPy, and instead to fully commit to one implementation; otherwise, this may significantly increase runtime. The memory layout for NumPy and JAX arrays are different, with the latter using asynchronous dispatch to hide Python overheads. The back and forth transfer to Python slows down a JAX program, since converting back to a numpy.ndarray will make JAX force the Python code to await the completion of the computation.

The speed-up attained through JAX is remarkable and only requires the ever-so-slight tweak in code; with these accelerated calculations, we can perform full redundant calibration without

![Figure 4.10: Redundantly calibrated visibilities by redundant baseline group on JD 2458098 at frequency channel 282 and LST 5.34 h. As in Figure 4.9, the redundant baseline group index and the first baseline in each group is also shown at the bottom of each plot. The 5σ MCD Mahalanobis distance contours are also drawn.](image)
4.3. Results for generalized redundant calibration

having to make any approximations. Distributions other than a Gaussian can also be used at no extra cost. As the problems scale up, we expect JAX to perform even better. We also expect even greater acceleration on GPUs or TPUs; however, this may need a bit of thought when coding up to ensure that the hardware is best utilized.

With the current implementation, a single dataset (1024 frequencies \(\times\) 60 time integrations \(\times\) 721 baselines; 1 polarization only) takes \(\approx\) 40 hours using 4 CPU cores. If solutions from a separate JD are reused as initial parameters, then the compute time is reduced to \(\approx\) 10 hours. Strong scaling is observed as speed-up also scales with the number of CPUs used. I found
Table 4.1: Speed comparison for relative redundant calibration (RedCal; Section 4.1) and the optimal redundant calibration (OptCal; where degeneracies are also constrained, see Section 4.1.3) with Cauchy and Gaussian noise assumptions, when run on a CPU with 4 cores.

<table>
<thead>
<tr>
<th>Noise Assumption</th>
<th>Minimization Method</th>
<th>Derivative Information</th>
<th>JAX Time</th>
<th>NumPy Time</th>
<th>Speed-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>RedCal Cauchy</td>
<td>BFGS</td>
<td>Jacobian</td>
<td>2.57 s</td>
<td>9.27 s</td>
<td>×3.6</td>
</tr>
<tr>
<td>RedCal Gaussian</td>
<td>BFGS</td>
<td>Jacobian</td>
<td>3.55 s</td>
<td>10.3 s</td>
<td>×2.9</td>
</tr>
<tr>
<td>OptCal Cauchy</td>
<td>trust-region</td>
<td>Hessian</td>
<td>9.65 s</td>
<td>110 s</td>
<td>×11</td>
</tr>
<tr>
<td>OptCal Gaussian</td>
<td>trust-region</td>
<td>Hessian</td>
<td>9.53 s</td>
<td>156 s</td>
<td>×16</td>
</tr>
</tbody>
</table>

that multithreading uses all the cores for this computation, and further multiprocessing was not found to help for this particular situation; simpler optimizations, like the ones in Section 3.4.1.2 to solve for the geometric median or Section 4.4.1 to degenerately translate between redundant visibility solutions, are found to benefit from multiprocessing.

4.3.3 Noise in MLE calculations for redundant calibration

In the hitherto MLE results, I neglected the noise contribution in the negative log-likelihoods of Equations 4.2 and 4.5, effectively assuming it is uniform across baselines. For a more complete solve, the MLE routines are modified to also take noise as an input in the calibration. The variance $\sigma_{ij}^2$ in Gaussian MLE for a particular baseline between antennas $i$ and $j$ can be predicted using the product of antenna autocorrelations:

$$\sigma_{ij}^2 = \frac{V_{ii}V_{jj}}{\Delta \nu t_{int}} \tag{4.22}$$

where $\Delta \nu$ is the channel bandwidth and $t_{int}$ the integration time. This equality follows from the radiometer equation (Equation 2.24). See Dillon (2019) for the different noise products from the HERA pipeline.

The Cauchy distribution has undefined finite moments, so its scale parameter $\gamma$ cannot be related to the predicted variance. For a normal distribution, the sample standard deviation $\sigma$ is expected to decrease as $1/\sqrt{N}$ for $N$ independent measurements. This behaviour is not seen when drawing from a Cauchy distribution, as samples from the heavier tails are constantly at odds with any $N^{-1/\alpha}$ behaviour. For Cauchy assumed noise, I therefore consider uniform $\gamma$ across baselines. In any case, an increase in noise counteracts a large delta in the negative log-likelihoods; we wish to be sensitive to such discordances, so the omission of the noise parameter in MLE is not problematic and is even preferred for flagging purposes.

Ideally, we would like the variance $\sigma_{ij}^2$ in the MLE under Gaussian noise to be purely thermal-like. The predicted variance from autocorrelations in Equation 4.22 may have contributions from other effects, and $\sigma_{ij}^2(\nu, t)$ could rapidly fluctuate in frequency and time. For
### 4.3. Results for generalized redundant calibration

#### Figure 4.12: Minimum $-\ln(L_{\text{rel}})$ for the redundant calibration of visibilities on JD 2458098 between LSTs 5.28 – 5.46 h with (N) and without ($\bar{N}$) noise input. Figure 4.12a is the same as Figure 4.6a, and is reprinted here for easier comparison with Figure 4.12b.

![Figure 4.12](image)

(a) Minimum $-\ln(L_{\text{rel}})$ (without noise input).

(b) Minimum $-\ln(L_{\text{rel}}^N)$ (with noise input).

#### Figure 4.13: Solved redundant visibility amplitudes for calibration with noise (left) and residual between calibrations with and without noise (right) for the redundant calibration of visibilities on JD 2458098 at the 30th time integration of Figure 4.12 (i.e. at LST 5.37 h). Note that degenerate comparison is required for the residual comparison in Figure 4.13b.

![Figure 4.13](image)

(a) Amplitude of redundant visibility solutions per baseline group for calibration with noise input. The solutions for calibration without noise input are very similar.

(b) Absolute normalized residual between the redundant visibility solutions solved with or without noise input for each baseline group.

Improved results, a smoothed noise surface across time and frequency should be created for each JD, which could be used for input in the MLE.

I show $-\ln(L_{\text{rel}})$ and a time slice of visibility amplitude solutions for the redundant calibration with and without Gaussian noise parameter input of raw H1C_IDR2.2 visibilities between LSTs 5.28 – 5.46 h (EE polarization) on JD 2458098 in Figures 4.12 and 4.13.

The heatmaps for $-\ln(L_{\text{rel}})$ in Figure 4.12, while inverted in colour across the frequency axis compared to each other, highlight the same problematic regions. The visibility amplitudes for sample time integration 0 (LST 5.28 h) in Figure 4.13 are also practically equal, bar for a few
Chapter 4. Robust redundant calibration

problematic regions (band edges and RFI contaminated channels). The calibration solutions calculated with noise input have higher fidelity as they are the full solution of Equation 4.2. For flagging purposes, the results without noise input are preferred.

The compute times for redundant calibration with and without variance input are discussed in Figure 4.27. It is found that minimization with variance input is just as fast as without variance input in the single-day case. If multiple JDs are simultaneously considered in a unified solve (see Section 4.5), then the compute time for minimization with variance input grows faster from $\approx 10$ days.

4.4 Comparing visibility solutions from redundant calibration

In this section, I develop routines to make sets of redundantly calibrated visibilities from different initializations directly comparable by solving for the degenerate parameter offsets between them. This allows for comparison between pairs of datasets, e.g. adjacent frequencies, times, or between different days. Building on this, we can also compute statistics of visibilities across different days without the need for absolute calibration by ensuring that all visibilities on separate JDs live in the same degenerate space. This can be used to test the consistency of visibility solutions, and provides a new approach of identifying and flagging data corrupted by e.g. RFI. This comparison requires MLE of the degenerate parameters, and is again generalizable such that different noise distributions can be assumed. I show that this technique is a useful quality assessment step and, in initial application, appears to identify plausibly bad data not previously flagged in the HERA reduction pipeline.

4.4.1 MLE to translate between degeneracies

Observed visibilities are affected by the $A(\nu)$, $\Delta_x(\nu)$ and $\Delta_y(\nu)$ degenerate parameters (see Section 4.1.3). The overall phase $\Delta(\nu)$, however, has no influence on the visibilities at all. When true visibilities are expected to be the same or very similar in a set of slots (e.g. looking at the same frequency and LST on different days), then the observed visibilities after relative redundant calibration must be consistent up to a transformation in the $A(\nu)$, $\Delta_x(\nu)$ and $\Delta_y(\nu)$ space.

Solving for, or potentially marginalizing, these parameters with an MLE framework enables us to check if these solutions are consistent, as well as allowing for direct comparison of the visibility solutions. To compare datasets that have been redundantly calibrated, we minimize:
4.4. Comparing visibility solutions from redundant calibration

Gaussian assumed noise

\[- \ln(L^G_{\text{deg}}(\nu)) = \frac{1}{2} \sum_\alpha \sum_{(i,j)\alpha} \ln(2\pi \sigma_{ij}^2(\nu)) + \frac{|U'_\alpha(\nu) - W_\alpha(\nu)|^2}{\sigma_{ij}^2(\nu)}\]

(4.23)

Cauchy assumed noise

\[- \ln(L^C_{\text{deg}}(\nu)) = \sum_\alpha \sum_{(i,j)\alpha} \ln(\pi \gamma_{ij}(\nu)) + \ln \left(1 + \frac{|U'_\alpha(\nu) - W_\alpha(\nu)|^2}{\gamma_{ij}(\nu)}\right)^2\]

(4.24)

where

\[W_\alpha(\nu) = A^2(\nu)e^{i[\Lambda_\alpha(\nu) x_\alpha + \Delta_\alpha(\nu) y_\alpha]} U_\alpha\]

(4.25)

and \(U_\alpha\) is the first set of redundant calibration visibility solutions for the redundant baseline group indexed by \(\alpha\), and \(U'_\alpha\) the second set of visibility solutions (this can be e.g. an adjacent time integration or frequency channel, or can be visibility solutions on a different JD but at the same LST and frequency).

The degenerate parameters found from minimizing either Equation 4.23 or Equation 4.24 can then be used to translate the visibility solutions of one dataset into the degenerate space of the other so that both datasets are degenerately consistent. They can then be directly compared, and statistical diagnostics may be run on them.

4.4.2 Phase gradients

In relative redundant calibration, the degenerate parameters can take any value. This does not affect the convergence of the calibration, but it can prevent convergence of the solver that we use to reconcile visibilities observed on different days. A particular problem encountered is that the phase gradient terms can take on very high values and can also vastly differ between adjacent times and frequencies as the minimizer jumps around the degenerate space when redundantly calibrating a given dataset. I find that such disparate phase gradients sometimes prohibits the comparison of datasets, with the minimizer not working as expected (probably due to subtleties induced by phase wrapping). To deal with this, when redundantly calibrating the second dataset \(U'_\alpha\), I reuse the solutions from the first redundant calibration as initial parameters, and introduce the following penalty term:

\[P_{\text{tilt}} = \sum_i (\phi'_i - \phi_i)^2\]

(4.26)

where \(\phi_i\) are the solved gain phases from the first redundant calibration and \(\phi'_i\) are the gain phases that minimize \(- \ln(L_{\text{cal}})\) for the redundant calibration of the second dataset. This penalty term ensures that the phases, and therefore the tilts, of the second calibration solutions are as
close as possible to those from the first. I also ensure that $\phi'_i - \phi_i$ is wrapped from $-\pi$ to $\pi$ when computing Equation 4.26.

The final cost function to minimize in the relative redundant calibration of the second dataset, to ensure consistent tilts, is, therefore:

$$C = -\ln(L_{rel}) + P_{\text{tilt}}$$

(4.27)

Computationally, it is found that constraining tilts in this manner, as well as using cartesian coordinates for the input parameters ($\Re e$ and $\Im m$ components for both visibilities and gains, instead of considering polar coordinates) in relative redundant calibration, is the fastest and most reliable method (its implementation is also the simplest).

Other attempts at solving this problem were made, including adding a tip-tilt regularization term when running the redundant calibration, such that all solutions have a set (arbitrary) tip-tilt; however, these did not succeed, most likely because of phase wrapping issues that cause the cost function to be discontinuous.

The addition of the phase constraint in the cost function should only be used if the aim is to compare datasets. If the redundant calibration for a particular data slice is off due to bad raw data, minimizing Equation 4.27 could produce degraded calibration results since $P_{\text{tilt}}$ will push the solver towards gains with erroneous phases that may be at odds with the ones computed if only $-\ln(L_{rel})$ was considered. This is fine for the comparison of datasets, since this tells us that the datasets are inconsistent (as will $-\ln(L_{\text{deg}}^G)$). For healthy data slices $P_{\text{tilt}} \approx 0$; still, for the best redundant calibration, it should not be included in the total cost function as it restricts the movement of gain solutions to the radial direction.

### 4.4.3 Pairwise comparison results

I focus on quantifying the stability of redundant calibration visibility solutions at the same LST for different JDs, as an alternative angle to inspect the quality of the calibration and the data itself. Comparison between adjacent frequency channels and time integrations can be similarly made with the presented method and code. I first look at the pairwise comparison of visibilities from JDs 2458098 and 2458099 between LSTs 5.28 – 5.46 h that were redundantly calibrated in Section 4.3.1. Example code for such a comparison is shown in Appendix B.2.2 for a single time integration and frequency. I note that the redundant calibration of JD 2458098 had to be redone with the $P_{\text{tilt}}$ penalty term added to the cost function (Equation 4.27).

The minimum negative log-likelihoods $-\ln(L_{\text{deg}})$ found by degenerate comparison of the redundant visibility solutions, with Gaussian assumed noise, are shown in Figure 4.14a. I further show $R_{\text{man}}$ for this minimization in Figure 4.14b. These heatmaps follow similar structure to those in Figure 4.6. The darker regions show where the visibility solutions between
4.4. Comparing visibility solutions from redundant calibration

Figure 4.14: $-\ln(L_{\text{deg}})$ and $R_{\text{man}}$ for the degenerate comparison of redundant visibility solutions on JDs 2458098 and 2458099 between LSTs 5.28 – 5.46 h.

Figure 4.15: Final calibration flags from the HERA analysis pipeline for JDs 2458098 and 2458099 between LSTs 5.28 – 5.46 h.

JDs disagree; these can mainly be attributed to RFI present in either or both JDs. With this method, we cannot tell which JD is anomalous purely by looking at $-\ln(L_{\text{deg}})$.

For reference, I show in Figure 4.15 the final flagging that is applied on the very same visibility datasets if calibration is done through the HERA analysis pipeline; Figures 4.6 and 4.14 identify many of the same erroneous regions.

The $-\ln(L_{\text{deg}})$ and $R_{\text{man}}$ results from the degenerate comparison can be used for further flagging: we can plot histograms for both of these quantities, categorized by the flagging from Figure 4.15; in doing so, we get a feel for the distribution of the data, which consequently indicates a reasonable cut-off value, after which frequency/time slices should be cast as outliers and flagged (if they haven’t been already). There is, however, some spectral structure in $-\ln(L_{\text{deg}})$ (see Figure 4.14a) due to the shape of the bandpass. Therefore, we divide $-\ln(L_{\text{deg}})$ by the smoothed variance for the 14 m EW baselines $\sigma_{14m}^2$ calculated from autocorrelations to obtain a dimensionless quantity that is comparable across frequencies. To compute $\sigma_{14m}^2$, we
first consider the variance of the visibilities for the 14 m EW baselines at each frequency/time slice for a given dataset, calculated using Equation 4.22. We then take the mean of this variance across baselines and perform a round of sigma-clipping along the time axis, where we use the median as the centre value for the clipping and set the lower and upper clipping limits to 4\sigma. Any flags are then filled through linear interpolation. A Savitzky-Golay filter (Savitzky & Golay 1964) with window length of 17 time integrations and fitting polynomial of order 3 is finally applied to smooth the noise signal along the time axis. Any negative values at this stage are set to $1 \times 10^{-8}$, to ensure a positive $-\ln(L_{\text{deg}})/\sigma^2_{14\text{ m}}$ ratio. This smoothing is done so that any intermittent high noise does not cancel out high $-\ln(L_{\text{deg}})/\sigma^2_{14\text{ m}}$ values. We do not do any smoothing along the frequency dimension. The histograms for $-\ln(L_{\text{deg}})/\sigma^2_{14\text{ m}}$ and $R_{\text{man}}$ are plotted in Figure 4.16.

From Figure 4.16, I find several frequency/time slices that lie above the indicative outlier cut-offs, but that are not flagged through the HERA pipeline (cf. Figure 4.15). I present these slices in Table 4.2.

I further show the visibility amplitudes and phases for the slices in Table 4.2 for the redundantly solved 14 m baseline in Figure 4.17. These show that the identified potentially erroneous slices are either near known faulty frequency channels or near band edges. While these frequencies are not used for PS estimation (see Table 2.3), accurate flagging of visibilities...
4.4. Comparing visibility solutions from redundant calibration

Comparing visibility solutions from redundant calibration involves analyzing the consistency and reliability of visibility data across different calibration procedures. This section discusses methods for evaluating and improving the quality of visibility measurements, particularly in the context of redundant calibration.

### Statistics across degenerately transformed redundant visibilities

#### Table 4.2: Potentially bad frequency/time slices found from Figure 4.16.

<table>
<thead>
<tr>
<th>Channel</th>
<th>Time integration</th>
<th>$-\ln(L_{\text{deg}})/\sigma_{14\text{m}}^2$</th>
<th>$R_{\text{man}}/10^{-2}$</th>
<th>$-\ln(L_{\text{deg}})/10^{-3}$</th>
<th>$\sigma_{14\text{m}}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>311</td>
<td>1</td>
<td>9.823</td>
<td>7.557</td>
<td>5.919</td>
<td>60.258</td>
</tr>
<tr>
<td>322</td>
<td>51</td>
<td>9.933</td>
<td>9.661</td>
<td>5.807</td>
<td>58.455</td>
</tr>
<tr>
<td>328</td>
<td>19</td>
<td>10.663</td>
<td>8.481</td>
<td>6.165</td>
<td>57.814</td>
</tr>
<tr>
<td>334</td>
<td>37</td>
<td>9.870</td>
<td>10.154</td>
<td>5.630</td>
<td>57.036</td>
</tr>
<tr>
<td>339</td>
<td>36</td>
<td>9.963</td>
<td>10.111</td>
<td>5.595</td>
<td>56.157</td>
</tr>
<tr>
<td>354</td>
<td>4</td>
<td>9.988</td>
<td>9.100</td>
<td>5.527</td>
<td>55.340</td>
</tr>
<tr>
<td>908</td>
<td>31</td>
<td>1.244</td>
<td>20.055</td>
<td>0.049</td>
<td>39.733</td>
</tr>
<tr>
<td>910</td>
<td>44</td>
<td>1.407</td>
<td>20.712</td>
<td>0.054</td>
<td>38.447</td>
</tr>
</tbody>
</table>

The degenerate comparison between two datasets (each 1024 frequencies × 60 time integrations × 721 baselines; 1 polarization only) takes ≈ 6 hours using 4 cores.

#### 4.4.4 Statistics across degenerately transformed redundant visibilities

Pairwise comparison of redundantly calibrated visibility solutions is a limited diagnostic tool, as we usually wish to compare and subsequently reduce visibilities observed on many JDs (in the case of H1C_IDR2.2, there are 18 evenings of observation). Comparing pairs of days in such a manner quickly grows into an increasingly large combinatorial problem that becomes difficult to manage. We can, however, compare all days to a single anchor day by transforming each day into the degenerate space of that anchor day. We can then run statistics on the entire multi-day dataset that is now degenerately consistent. While simpler, this approach has the caveat that this anchor day must be well behaved and relatively uncontaminated by RFI, as any anomalies present in the anchor day will be present in the degenerate comparison negative log-likelihoods (which we wish to reduce insofar as possible).

As discussed in Section 4.1.3, degeneracies in redundant calibration can also be constrained...
straight away through an ‘optimal’ calibration procedure whereby constraints (Equations 4.6 to 4.8) are fed directly into the minimizer. This is, however, an expensive computation that is not always reliable, so we prefer the degenerate comparison explained here.

4.4.4.1 Statistics and outlier detection

I first redundantly calibrated datasets for all H1C_IDR2.2 JDs (except for JD 2458109 [21st December 2017]) that fall between LSTs 5.28 – 5.46 h. I then transform all visibilities to the degenerate space of JD 2458099 (our anchor day), and align the visibilities in LST.

Owing to the high dimensionality of the data, only cross-sections of this data can be

Figure 4.17: Redundantly calibrated visibilities for the solved 14 m EW baseline group for JDs 2458098 and 2458099 between LSTs 5.28 – 5.46 h. The outliers from Table 4.2 are circled. Note that the jump in visibility phases in Figures 4.17b and 4.17d at time integration 22 corresponds to the point at which redundantly calibrated visibilities from datasets labelled by 2458099.43124 and 2458099.43869 have been concatenated. The discontinuity in phase for certain channels is due to the solver finding different (but equally valid) solutions, due to the degeneracies explained in Section 4.1.3.
4.4. Comparing visibility solutions from redundant calibration

visualized at once. In Figure 4.20, I plot heatmaps for the visibility amplitudes and standard deviations for cross-sections of the data in time and for a given redundant baseline. I also plot the median and standard deviation of $-\ln(L_{\text{rel}})$ (also across JDs).

There is structure in both time and frequency for the standard deviation heatmaps in Figure 4.20, indicating that at least one JD is not in agreement with the others. I next discuss the different ways of identifying faulty data from these quantities.

Statistics can be run across the high-dimensional dataframe of redundantly calibrated, degenerately transformed, and LST-aligned visibilities over different dimensions to find possible outliers. I aim to identify corrupted frequency/time slices that have not been flagged through the HERA analysis pipeline (final flags can be recovered by looking at the calibration files following gain smoothing in Section 2.2.3).

One possible way of look at outliers is to filter by the $Z$-score ($\frac{(x - \mu)}{\sigma}$, i.e. the number of standard deviations from the mean) or modified $Z$-score (see Equation 2.36) of the solved $-\ln(L_{\text{rel}})$ for each JD/frequency/time slice compared to the mean for that given frequency/time slice across JDs, therefore checking if any particular JD is anomalous. Here, I reject all baselines for an entire frequency/time slice, which is in contrast to the per-baseline flagging in some of the outlier detection methods outlined in Chapter 3. In Table 4.3, I present a list of potentially bad JD/frequency/time slices that have not been flagged in the analysis pipeline, and that were found by applying $|Z| > 3.3$ or $|Z^{\text{mod}}| > 11$ filters.

This method of outlier rejection differs from the MAD-clipping done at the LST-binning stage of the main HERA pipeline (see Section 2.2.3.2): in the clipping routine, each baseline is considered separately, and days for each baseline can be clipped. Here, we reject all baselines for a given JD/frequency/time; we do, however, expect there to be some overlap in the flagged data from both methods.

There are many ways to dissect the dataframe; outlier detection can be done by e.g. looking at visibilities directly, instead of $-\ln(L_{\text{rel}})$, to obtain per-baseline flags. Alternatively, one could also look at $-\ln(L_{\text{deg}})$ for all JDs comparisons to the anchor day. Outliers can then be identified through histogram considerations, like in Figure 4.16.

The methods of outlier detection outlined in this section are somewhat intricate, as they require many steps to finally get visibilities that can be compared like-for-like (redundant calibration followed by degenerate transformation, with the latter proving delicate to get right as it requires a gain phase constraint from an anchor day). Ideally, all redundantly calibrated visibility solutions would end up in the same degenerate space straight out of the calibration, but adding these constraints through regularization or through constrained optimization is not always successful. In the next section, I introduce a more sophisticated redundant calibration mechanism that takes in data from multiple JDs; there are many advantages to this unified
Chapter 4. Robust redundant calibration

Figure 4.18: Redundantly calibrated and degenerately transformed median visibility amplitude and associated visibility standard deviation across JDs at LST 5.44 h. Each row indexes a redundant baseline group, with baseline length increasing with index number. The 14 m and 29 m EW baselines have indices 2 and 6, respectively.

Figure 4.19: Redundantly calibrated and degenerately transformed median visibility amplitude and associated standard deviation across JDs for the 14 m EW baselines over LSTs 5.28 – 5.46 h.

Figure 4.20: Median and standard deviation of the minimum $-\ln(L_{\text{rel}})$ found in the redundant calibration minimizations across JDs, between LSTs 5.28 – 5.46 h.
4.5 Unified generalized redundant calibration

Much like the more widely used ‘self-calibration’ (that uses a source model with e.g. \texttt{clean}), redundant calibration solves for both the visibilities and the gains of the receiving elements, including that the redundant visibility solutions are degenerately consistent, making them easier to work with for outlier detection purposes.

### Table 4.3: Potentially bad JD/frequency/time slices that are not flagged by the HERA analysis pipeline, which exceed the $|Z| > 3.3$ or $|Z^{\text{mod}}| > 11$ thresholds. These thresholds were chosen arbitrarily, but mainly to limit the number of slices shown.

<table>
<thead>
<tr>
<th>JD</th>
<th>Channel</th>
<th>Time integration</th>
<th>$-\ln(L_{\text{rel}})$</th>
<th>$\text{med} - \ln(L_{\text{rel}})$</th>
<th>$Z$</th>
<th>$Z^{\text{mod}}$</th>
<th>$\sigma^{\text{mad}}/10^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2458098</td>
<td>177</td>
<td>46</td>
<td>0.1584</td>
<td>0.1325</td>
<td>1.9986</td>
<td>12.2887</td>
<td>1.4243</td>
</tr>
<tr>
<td>2458101</td>
<td>274</td>
<td>59</td>
<td>0.1026</td>
<td>0.0822</td>
<td>3.3712</td>
<td>4.9429</td>
<td>2.7893</td>
</tr>
<tr>
<td>2458101</td>
<td>292</td>
<td>20</td>
<td>0.1209</td>
<td>0.0899</td>
<td>3.3787</td>
<td>10.7163</td>
<td>1.9544</td>
</tr>
<tr>
<td>2458101</td>
<td>311</td>
<td>17</td>
<td>0.1423</td>
<td>0.1089</td>
<td>3.4826</td>
<td>6.1406</td>
<td>3.6649</td>
</tr>
<tr>
<td>2458101</td>
<td>442</td>
<td>36</td>
<td>0.0516</td>
<td>0.0409</td>
<td>3.3019</td>
<td>4.6670</td>
<td>1.5537</td>
</tr>
<tr>
<td>2458101</td>
<td>730</td>
<td>31</td>
<td>0.0106</td>
<td>0.0083</td>
<td>3.4344</td>
<td>7.6664</td>
<td>0.2032</td>
</tr>
<tr>
<td>2458101</td>
<td>395</td>
<td>53</td>
<td>0.0905</td>
<td>0.0737</td>
<td>0.8880</td>
<td>11.8645</td>
<td>0.9564</td>
</tr>
<tr>
<td>2458110</td>
<td>599</td>
<td>50</td>
<td>0.0318</td>
<td>0.0216</td>
<td>3.3433</td>
<td>9.3110</td>
<td>0.7379</td>
</tr>
<tr>
<td>2458110</td>
<td>606</td>
<td>39</td>
<td>0.0285</td>
<td>0.0213</td>
<td>3.4352</td>
<td>5.8054</td>
<td>0.8370</td>
</tr>
<tr>
<td>2458110</td>
<td>621</td>
<td>39</td>
<td>0.0248</td>
<td>0.0177</td>
<td>2.6354</td>
<td>11.0135</td>
<td>0.4348</td>
</tr>
<tr>
<td>2458113</td>
<td>160</td>
<td>1</td>
<td>0.4649</td>
<td>0.1554</td>
<td>2.2268</td>
<td>30.4810</td>
<td>6.8484</td>
</tr>
<tr>
<td>2458113</td>
<td>169</td>
<td>0</td>
<td>0.2826</td>
<td>0.1459</td>
<td>2.5933</td>
<td>13.9014</td>
<td>6.6344</td>
</tr>
<tr>
<td>2458115</td>
<td>160</td>
<td>1</td>
<td>0.4299</td>
<td>0.1554</td>
<td>1.9252</td>
<td>27.0328</td>
<td>6.8484</td>
</tr>
<tr>
<td>2458115</td>
<td>802</td>
<td>30</td>
<td>0.0052</td>
<td>0.0062</td>
<td>-2.0759</td>
<td>-11.2726</td>
<td>0.0608</td>
</tr>
<tr>
<td>2458116</td>
<td>160</td>
<td>1</td>
<td>0.4509</td>
<td>0.1554</td>
<td>2.1057</td>
<td>29.0968</td>
<td>6.8484</td>
</tr>
</tbody>
</table>
of the interferometer. Unlike self-calibration, however, the visibilities when doing redundant calibration are only estimated in the relatively few sampled cells in the $u$-$v$ plane where the redundant baselines happen to lie. In self-calibration, since the model of the sky is estimated in the image plane, the technique effectively estimates the visibility values across the whole $u$-$v$ plane. Consequently, self-calibration benefits greatly from combining multiple times and frequencies into a single estimate of the sky, as earth-rotation and sampling across the radio frequency bandwidth fill out the coverage of the $u$-$v$ plane. Since redundant calibration does not gain anything by filling out the $u$-$v$ plane, the incentive to combine multiple times and frequencies of observations is less apparent. There is, however, great merit in combining the solving so that a single all-encompassing redundant estimate is made based on as much data as possible. Some reasons for this are:

1) Independently solving for the visibility values introduces more degrees of freedom than needed. For example, the sky visibilities must be exactly equal at the same LST on different days, so we could unify the calibration by only solving for a single set of visibility solutions: if we combine the calibration across days, then the total number of complex parameters to be solved is $N_{\text{days}} \times N_{\text{ants}} + N_{\text{bl\_grps}}$ compared to $N_{\text{days}} \times N_{\text{ants}} + N_{\text{days}} \times N_{\text{bl\_grps}}$ if solving individually for each JD. Moreover, as frequency channels are made narrower, the visibilities of neighbouring redundant calibration solutions must converge too; the same goes for visibilities for adjacent time integrations. Having more degrees of freedom typically means a higher possibility of over-fitting and some loss on S/N.

2) Redundant visibility solutions from different initializations (e.g. different days but same LST) are not degenerately consistent. As discussed in Section 4.4, if we want to compare redundantly calibrated visibilities solved separately for each JD and to run meaningful statistics on them, the degenerate parameter offsets between the visibility solutions must be solved. This is an additional data processing step that we could avoid. A unified solver would also facilitate the identification and flagging of corrupted data, since all raw visibilities can be compared directly to the solved redundant visibility solutions (after multiplication with the solved gains).

3) The statistics of non-Gaussian distributions can be meaningless (e.g. the median of the median $\neq$ the median of the whole dataset). By considering the entire dataset, we obtain the best visibility estimates for the data. This is especially relevant when dealing with robust distributions, such as the Cauchy distribution.

4) RFI introduces a strongly non-Gaussian error into radio observations. Robust estimators of visibilities and antenna gains are far easier to construct as more data are considered.

---

*In this combined calibration, we need to solve for all the antenna gains (which must be found separately for each day) and the redundant visibility solutions for the redundant baseline groups.*
simultaneously, since rare events are more prominent with increased data volumes.

For these reasons, I build on the redundant calibration described in Section 4.1 so that it takes in visibilities from multiple days (LST alignment is required) and outputs a single set of visibility solutions and the gains for each day, as motivated and described in 1) above.

In such a unified solve, it is even easier to compare redundant calibration estimates of visibilities across days to test the stability of the data and to find outliers, since all solutions will be in the same degenerate space, as indicated in 2). There may be a drift in the degenerate parameters for different times and frequencies; constraints can be implemented to fix these.

4.5.1 Unified MLE

I solve for all visibility data across JDs simultaneously to find a single set of redundant visibility solutions at any LST in a procedure that we shall call across days relative redundant calibration:

**Gaussian assumed noise**

\[-\ln(L_{xd,rel}^G)(\nu) = \frac{1}{2} \sum_d \sum_\alpha \sum_{\{i,j\}_\alpha} \ln(2\pi \sigma_{ij,d}^2(\nu)) + \frac{(V_{ij,d}^{obs}(\nu) - g_{i,d}(\nu)g_{j,d}^*(\nu)U_\alpha(\nu))^2}{\sigma_{ij,d}^2(\nu)}\]

\[(4.28)\]

**Cauchy assumed noise**

\[-\ln(L_{xd,rel}^C)(\nu) = \sum_d \sum_\alpha \sum_{\{i,j\}_\alpha} \ln(\pi \gamma_{ij,d}(\nu)) + \ln \left(1 + \frac{(V_{ij,d}^{obs}(\nu) - g_{i,d}(\nu)g_{j,d}^*(\nu)U_\alpha(\nu))^2}{\gamma_{ij,d}(\nu)}\right)^2\]

\[(4.29)\]

where we now also sum across days, indexed by \(d\). The LST dependence is implicit, and this optimization must be done independently for each frequency and time slice (I do not implement any convergence criteria for adjacent frequencies and times at this moment). As with the previous MLEs in this chapter, those with Cauchy-distributed noise grant location and scale parameters that are not distorted by outliers and hence provide the best median estimator of the data.

4.5.2 Results

I show example results for the redundant calibration across all H1C_IDR2.2 JDs (except for JD 2458109, which is known to be mostly faulty) for a single frequency and time slice (channel 600 and LST 5.44 h), with Gaussian assumed noise and with variance input. The gain amplitudes
Chapter 4. Robust redundant calibration

and phases solved from this calibration are plotted in Figure 4.21. The per-baseline normalized residuals from this solve are plotted in Figure 4.22. The code to obtain these results is written in Appendix B.2.3.

In addition, I plot the raw and calibrated visibilities for each baseline group for the same sample frequency and time slice in Figures 4.23 and 4.24, respectively, with visibilities coloured by JD.

Interestingly, in Figure 4.23, visibilities for certain baseline groups are clumped into separate clusters in the complex plane (more noticeable for e.g. baseline groups 54 and 75). These clumps correspond to different baselines within each group, and these differences are due to phase offsets that can be calibrated out. The calibrated visibilities in Figure 4.24 are well reconciled and no obvious outlier can be detected.

I next consider visibilities for the LST range 5.28 – 5.46 h (within Field 2) and frequency channels 500 – 700 (contains Band 2) that I redundantly calibrate across the same JDs, again with a Gaussian assumed noise distribution with variance input. I plot \(-\ln(L^G_{\delta d, rel})\) and \(R_{\text{man}}\) for these solves in Figure 4.25 to show the quality of the calibration for this dataset. Darker regions in these heatmaps indicate some disagreement between the considered visibilities, possibly due to anomalous JDs; this does not mean that the entire multi-JD data slice should be thrown away, but it instead confirms the presence of outliers for at least one of the considered JDs. I further plot the redundant visibility solutions for each baseline group in Figure 4.26 for a sample time bin.

I showed illustrative results as to how across days redundant calibration works. To reiterate, the advantage of this method is that we obtain a single redundant visibility estimate \(U_\alpha\) for each frequency and LST slice. As we saw in Figure 4.11, the visibility estimates from redundant calibration do not necessarily match with standard location estimators such as the mean or the geometric median. The solved gains and \(U_\alpha\) are the best possible estimates for the whole multi-JD dataset; the latter can then also be used as the central locations of outlier rejection algorithms (much like the robust Mahalanobis distance contours in Figure 4.24, but where the location is set instead of being calculated from MCD considerations).

4.5.3 Performance

In Figure 4.27, I plot the compute time for redundant calibration with Gaussian assumed noise distribution for the same frequency/LST slice across an increasing number of days. The results from the unified MLE solve are shown (xd_rel_cal), as well as those from performing redundant calibration individually for each JD (rel_cal) followed by transforming to the same degenerate parameter space (deg_cal). Very similar results are obtained for a Cauchy assumed distribution.
4.5. *Unified generalized redundant calibration*

---

(a) Gain amplitude solutions for all JDS. 

(b) Gain phase solutions for all JDS. 

**Figure 4.21:** Gain solutions coloured by JD for the unified redundant calibration across days example at frequency channel 600 and LST 5.44 h, with Gaussian assumed noise and variance input. The standard deviations across JDS for these results are shown in the bottom subplots. The jumps in phases for antenna 25 and antennas 124 onwards in Figure 4.21b are due to phase wrapping. The gains for this data slice appear to be stable across days.

---

(a) Re component of the normalized residual. 

(b) Im component of the normalized residual. 

**Figure 4.22:** Normalized residuals coloured by JD for the unified redundant calibration across days example at frequency channel 600 and LST 5.44 h with Gaussian assumed noise and with variance input (each baseline index has datapoints for all considered JDS). The residual in this case is $V_{ij}^{\text{obs}}(v) - g_i(v)g_j^*(v)U_{\alpha}(v)$. The normalization is done by dividing by $\left| V_{ij}^{\text{obs}}g_i g_j^* U_{\alpha} \right|^{1/2}$. The mean $\mu$ and $\mu \pm 2\sigma$ for all the normalized residuals are also shown.
Figure 4.23: Raw visibilities by redundant baseline group for all H1C_IDR2.2 JDs at frequency channel 600 and LST 5.44 h. The redundant baseline group index and the first baseline in each group are shown at the bottom of each plot. 5σ MCD Mahalanobis distance contours are drawn for each group to get an indication of scatter and to see the proportion of outliers.

The compute time grows exponentially with the number of days for the xd_re1_ca1 procedure, as expected from the $O(N^2)$ complexity of the BFGS algorithm. The compute time for the rel_ca1 + deg_ca1 combination grows linearly with the number of days. With this latter method, however, we do not find the best fit calibration solutions for the entire multi-JD dataset. We also have to solve for degenerate offsets, which, while not computationally expensive, introduces additional room for error and also relies on the anchor JD to be of good quality.

Should significantly more JDs be considered, as well as more antennas and baselines (as is the case for subsequent HERA observing campaigns), then the unified redundant calibration across days would become expensive. There is scope to speed-up this computation (with e.g. GPUs); however, with the current setup, we are ultimately limited by the $O(N^2)$ computational
4.6. Tolerance for termination in redundant calibration

Figure 4.24: Redundantly calibrated visibilities by redundant baseline group for all H1C_IDR2.2 JDs at frequency channel 600 and LST 5.44h. As in Figure 4.23, the redundant baseline group index and the first baseline in each group are also shown at the bottom of each plot. The 5σ MCD Mahalanobis distance contours are also drawn.

complexity of the BFGS algorithm. Nevertheless, this unified solve is the best redundant calibrator of the whole data; the added cost is worthwhile if we are to strive for calibration solutions that are accurate enough to push for a 21 cm PS detection.

4.6 Tolerance for termination in redundant calibration

The redundant calibration routine developed uses the SciPy minimizer with the BFGS method. An argument of this function is the gradient tolerance for termination `gtol`, which has default value $10^{-5}$. The algorithm measures the size (Euclidean norm) of the gradient of the function $f$ to be minimized with respect to its parameters at each iteration step $k$ to measure convergence:

$$v_k = \| \nabla f_k \|$$  \hspace{1cm} (4.30)
Chapter 4. Robust redundant calibration

Figure 4.25: \(-\ln(L_{\text{sd, rel}})\) and \(R_{\text{man}}\) for the across days redundant calibration for channels 500–700 and LSTs 5.28–5.46 h.

Figure 4.26: Visibilities solved from the across days redundant calibration for channels 500–700 at LST 5.44 h. The amplitude of the visibility estimates are smoother than those in Figure 4.13a. The phases in Figure 4.26b occasionally jump around because the minimizer sometimes finds itself in a different region of the degenerate parameter space, which leads to different but equivalent solutions.

Successful termination occurs once this gradient norm \(\nu_k\) is less than \(\text{gtol}\).

Results in Section 4.3.1 leave this value unchanged, as computations for a single day/frequency/time slice without noise input worked well out of the box, and no adjustment was needed. No modification to \(\text{gtol}\) was needed for the degenerate comparison in Section 4.4 either.

The role and importance of \(\text{gtol}\) became more apparent when incorporating the noise parameter in the MLE (Section 4.3.3) and when calibrating across multiple JDs (Section 4.5); these enhancements required \(\text{gtol}\) to be adjusted for these different minimizations to terminate successfully.

To get a better understanding of the sensitivity of the minimization to \(\text{gtol}\), I perform redundant calibration for a typical data slice for various values of \(\text{gtol}\). I consider data on JD 2458098, EE polarization, frequency channel 600, LST 5.44 h, and redundantly calibrate
4.6. **Tolerance for termination in redundant calibration**

(a) Compute time for redundant calibration across an increasing number of JDs, without noise input.

(b) Compute time for redundant calibration across an increasing number of JDs, with noise input (see Section 4.3.3).

**Figure 4.27:** In blue: compute times for redundant calibration with Gaussian assumed noise distribution, with different scalings shown, for frequency channel 600 and LST 5.44 h (found to be the least flagged slice across all H1C_IDR2.2 days), across an increasing number of days, such that 1 day is the redundant calibration of JD 2458098, 2 days is the redundant calibration of JDs 2458098 and 2458099 and so on, for all the days in H1C_IDR2.2 minus JD 2458109. In orange: compute times for redundant calibration + degenerate transformation, with previous solutions being used as initial parameters (for the \( \text{rel\_cal} \) part; this significantly speeds up computation). Noise is neglected and used for the MLE calculations in Figure 4.27a and Figure 4.27b, respectively. We note that previous solutions can also be reused as initial parameters for \( \text{xd\_rel\_cal} \), which decreases computation time by \( \sim 55\% \) (both with and without noise input), thus, making it as fast as the \( \text{rel\_cal} + \text{deg\_cal} \) approach. For a batch \( \text{xd\_rel\_cal} \) run, the biggest time sinks come from problematic frequency/time slices; if these are already flagged (using e.g. final calibration files from the analysis pipeline), then the batch run could be done in a reasonable time. I also note that for these computations, the tolerance for termination \( \text{gtol} \) (see Section 4.6) was fixed to be \( \text{gtol} = 10^{-5} \) and \( \text{gtol} = 5 \times 10^{3} \) without and with noise input, respectively; since the compute time is stable by a couple of orders of magnitude about these tolerance values, it is not necessary to increase the tolerance proportionally to the number of days under consideration. I do, however, recommend that this is done in general; my redundant calibration implementation defaults to multiplying \( \text{gtol} \) by the number of days.
assuming a Gaussian noise distribution. I plot these results in Figure 4.28; from these, the default \( gtol = 10^{-5} \) value used up until now is justified, and, for single-day redundant calibrations without noise input, it does fine.

In Section 4.3.3, I included noise input in the MLE calculations for redundant calibration. The noise for visibilities is small and has the effect of blowing up the \(-\ln(L_{rel})\). For this reason, \( gtol \) must be adjusted. I repeat the tolerance analysis in Figure 4.29, and find that a suitable tolerance value is \( gtol \sim 10^{4} \).

Performing this analysis with a Cauchy assumed noise distribution leads to suitable tolerances of \( 10^{-5} \lesssim gtol \lesssim 10^{-3} \) if noise is omitted from the MLE.

For redundant calibration across multiple JDs (Section 4.5), we expect \(-\ln(L_{rel}) \propto N_{\text{days}}\). Hence, we adjust \( gtol \) for such computations by multiplying it by \( N_{\text{days}} \). For the number of days in H1C_IDR2.2 (i.e. 18 days), this adjustment will not result in any material change in \( gtol \); this can be seen in Figures 4.28 and 4.29, as for the default values of \( gtol \) chosen, \(-\ln(L_{rel})\) remains stable for a couple of orders of magnitude on either side. However, as the number of days considered is further increased, this adjustment becomes necessary.

### 4.7 Discussion

I presented a generalized approach to redundant calibration that uses MLE to obtain best-fit location parameters for both the antenna gains and redundant visibilities for each frequency and time slice, which is accelerated with the JAX ML library. With this increase in performance, through the use of AD and JIT compilation, we can generalize beyond Gaussian models and assume different noise distributions for the data, with little added runtime. I chose JAX for this research, as it has source-level compatibility, delivers both forward and reverse-mode AD (efficient for Jacobian and Hessian computation), and has an attractive, functional framework and architecture that leverages from the Tensorflow ecosystem. With its simple and clean API (that is the same for CPU/GPU/TPU), JAX can be easily implemented into existing and new code to provide impressive out-of-the-box performance.

Redundant calibration was performed on H1C_IDR2.2 raw visibilities, using both Gaussian and Cauchy assumed noise distributions for the visibilities, with the latter distribution defined by robust location and scale parameters (the median and the HWHM). I find that the gain and visibility solutions using these two distributions are very similar for most situations, although the calibration results with Cauchy assumed noise may prove to be favourable as larger datasets are considered, since this robust MLE has improved resilience to RFI or instrumental artefacts over and above established flagging methods. Such robust statistics also require less ad hoc intervention.
4.7. Discussion

Figure 4.28: Gradient tolerance analysis for redundant calibration minimization with Gaussian assumed noise distribution and uniform variance across baselines (setting the variance $\sigma_{ij}$ to unity), for $10^{-10} \leq \text{gtol} \leq 1$, evenly spaced on a log scale. In the first graph from the top, I plot the solved minimum $-\ln(L_{\text{rel}})$ for each $\text{gtol}$. In the middle graph, the number of iterations taken to reach termination (set by $\text{gtol}$) is plotted. In the bottom graph, the compute time is plotted. The data is also coloured according to the success of the minimizer; at the lower end of the graphs, coloured in crimson, are data when the minimizer is limited by machine precision. The minimum $-\ln(L_{\text{rel}})$ approaches a value $\approx 1.930 \times 10^{-2}$; hence, we expect that for a good solve, $\text{gtol} < 10^{-2}$, which is what we observe. For $\text{gtol} < 10^{-5}$, the decrease in $-\ln(L_{\text{rel}})$ is negligible: around $\text{gtol} \sim 10^{-5}$, it is of $O(10^{-6})\text{k}$, while around $\text{gtol} \sim 10^{-6}$, it is of $O(10^{-11})\text{k}$. Choosing a $\text{gtol}$ between $10^{-5} - 10^{-6}$ is, therefore, recommended: we stick with $\text{gtol} \sim 10^{-5}$. We notice that the compute time does not suffer significantly as the tolerance decreases; it is best, however, to stick with a tolerance that does not linger too close to the edge of the minimizer being unsuccessful due to precision loss.
Figure 4.29: Gradient tolerance analysis for the redundant calibration minimization with Gaussian assumed noise distribution and input noise variance (predicted from the autocorrelations), for $10^{-1} \leq \text{gtol} \leq 10^9$, evenly spaced on a log scale. The different plots are as in Figure 4.28. The minimum $-\ln(L_{\text{rel}})$ approaches a value $\approx 1.820 \times 10^7$. 
I showed how pairs of redundant calibration visibility solutions can be compared by translating between their degenerate parameters. Through this comparison, the stability of redundant calibration visibility solutions can be probed, and the negative log-likelihoods for this comparison can be used as a metric for further flagging: certain effects that may appear as correct calibration solutions may not be present day-to-day and should be flagged. This pairwise comparison can also be used to compare a set of JDs to a single anchor JD.

The generalized redundant calibration was extended in Section 4.5 to take in data across multiple JDs to find a single set of visibility solutions for any LST. While there is only a single set of resulting visibilities, gains are calculated for each JD. This unified method is optimal and provides the best estimates for multi-JD interferometric observations. The gains can be used to reconstruct the visibility solutions for each JD, which will be consistent (with respect to the degeneracies of Section 4.1.3). This way, statistics on all these visibility solutions can be performed at once. While there is a clear advantage in this unified solver, it is computationally expensive as it has increased time complexity compared to running redundant calibration separately for each JD. There is scope for further acceleration with ML software and hardware.

While I focused on redundant calibration in this work, the generalized MLE framework can also be applied to other calibration strategies, such as absolute sky-based calibration where we solve $V_{ij}^{\text{obs}}(\nu) = g_i(\nu)g_j^*(\nu)M_{ij}(\nu) + n_{ij}(\nu)$ for model visibilities $M_{ij}$, where the model is taken to be the sky itself (see e.g. the last paragraph of Section 2.2.2.4, which details how model visibilities are constructed in the HERA pipeline).
The FT (and hence PS) works very well in transforming a signal from its time domain to its frequency domain when the frequency spectrum is stationary and does not evolve in time. However, in the case of a dynamic signal (whose spectral features are not constant with time), the performance of the FT will be degraded; this is the case for most of the signals seen in nature. A non-stationary alternative to the FT is, therefore, required for the spectral decomposition of dynamic signals. In particular, for 21 cm cosmology, we compute PS over considerable frequency bandwidths: the Universe can change over such periods; this is known as the light-cone effect (Barkana & Loeb 2006), so a multiscale analysis may better constrain the statistics of the observed era. Wavelet transforms (WTs) offer just this capability by being able to localize the 21 cm signal in both frequency and delay, and the corresponding power provides an unbiased estimator that can fully characterize the 21 cm in all spatial dimensions (since we now include the LoS axis). I introduce wavelets and their transforms in Section 5.2.

Wavelets can also be used as both an additional diagnostic and outlier detection tool. If erroneous modes exist in a given signal, it is beneficial to simultaneously locate the fault to a point in the signal and Fourier domain, which is not possible with standard spectral decomposition. I present an error detection technique based on wavelets in Section 5.3 and identify contaminated times and baselines that were used in the final H1C_IDR2.2 PS results (see Section 2.2.4, which discusses the results in The HERA Collaboration et al. 2022c).

Power spectral estimates with wavelets are also computed for the 14 m and 29 m EW baselines in Section 5.4 as a new potential method to evaluate spectral power that is not yet established in 21 cm cosmology.
Throughout this chapter and Appendix A, wavelet and Fourier theory are introduced by considering a time signal whose Fourier dual is in the frequency domain. These methods are, however, applied to radio interferometric visibilities that are functions of frequency and that are transformed into the delay domain.

## 5.1 The short-time Fourier transform

A natural and commonly used method in time-frequency analysis is to divide the full time signal into shorter segments of equal length and to apply the FT locally on those segments. This operation is called the short-time Fourier transform (STFT), and is done by multiplying the signal $x(t)$ with a window function $w$ that is nonzero for the concerned period before transforming to frequency space. The continuous formulation is given by

$$X(\tau, f) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-2\pi if t} \, dt$$  \hspace{1cm} \text{(5.1)}$$

with the discrete case by

$$X_{m,k} = \sum_{n=0}^{N-1} x_ne^{-2\pi i \frac{k}{N} n}$$  \hspace{1cm} \text{(5.2)}$$

We look at the squared magnitude of the STFT $|X|^2$ to obtain the localized PS of the signal, which we can plot for different time localizations to get a spectrogram.

A caveat of the STFT is that it has fixed resolution, as determined by the width of the windowing function, with narrow windows giving good temporal resolution but poor spectral resolution and vice versa. In general, we wish to have good temporal resolution at high frequencies while also having good spectral resolution at low frequencies; this is where the WT comes in.

## 5.2 The wavelet transform

The wavelet transform (WT) builds on the STFT by having increased temporal resolution at higher frequencies at the sacrifice of spectral information, with the opposite occurring at lower frequencies. We, thus, sample the dual time-frequency space at different resolutions to get a more complete and often more helpful representation of the signal. This principle is illustrated in Figure 5.1, where the time and frequency resolutions for the different transforms introduced in this thesis are shown.

Wavelets are a powerful tool with many applications and have been used with great success in other fields of astrophysics. Their usage in 21 cm cosmology for EoR detection has, however, been limited. Trott (2016) studied wavelets and their potential in detecting the 21 cm signal, and
Figure 5.1: Time-frequency tiling illustration for the FT, STFT and WT. In signal (time) space, the tiling would be purely vertical with full temporal resolution and no spectral information.

showed that this novel approach has improved estimation performance compared to the standard Fourier PS that is commonplace in the literature. As EoR experiments advance by growing and improving their sensitivity and resolution, wavelets and their transforms will become increasingly important since they are able to retain location (frequency/redshift) information. Wavelets will likely become a key tool to statistically probe the $21\,\text{cm}$ signal.

The basic principles of wavelet analysis are described in the rest of this section. I refer to Mallat (2009); Debnath & Shah (2015) for comprehensive overviews of wavelets and their properties.

We use the continuous wavelet transform (CWT) as the tool to represent the signal; it is defined as

$$\gamma(\tau, a) = \langle x(t)\psi_{\tau,a}(t) \rangle = \int_{-\infty}^{\infty} x(t)\psi_{\tau,a}^*(t) \, dt$$

where $\psi_{\tau,a}$ are a set of basis functions called wavelets, which are generated from a single wavelet $\psi$ called the *mother* wavelet:

$$\psi_{\tau,a}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right)$$

where $a$ is the scale factor and $\tau$ is the translation factor.

A commonly used wavelet is the complex Morlet (also called Gabor) wavelet, which is given by

$$\psi_{cM}(t; B, C) = \frac{1}{\sqrt{\pi B}} e^{-\pi^2 t^2} e^{i2\pi C t}$$

where $B$ is the bandwidth and $C$ is the centre frequency. The former, $B$, is the inverse of the variance of the Morlet wavelet in the frequency domain; therefore, increasing $B$ reduces the variance in frequency, resulting in a higher energy concentration around the central frequency. The Morlet wavelet is composed of the product of a complex carrier (i.e. the standard Fourier
mode) with a Gaussian envelope (with an energy normalization factor in front). See Figure 5.2 for plots of the Morlet wavelet function and its corresponding FT for $B = 1.5$ and $C = 1$. The CWT with $\psi_{cM}(t; B = 1, C = 1)$ is a minor modification of the STFT with Gaussian window, and is also known as a Gabor transform (Gabor 1946).

In Figure 5.3, I also depict how the scaling and translation of wavelets affect their spreads in frequency and time.

The CWT for a finite sequence of equally spaced samples $x_n$ (not to be confused with the discrete wavelet transform, see Appendix A.6) is given by

$$\gamma_{m,a} = \sum_{n=0}^{N-1} x_n \psi^{*}_{(m,a), n}$$

with Equations 5.4 and 5.5 similarly discretized.

There are a plethora of wavelets to choose from, each with its own characteristics and benefits. In general, it is helpful to select a wavelet that is similar to the signal under investigation. The Morlet wavelet $\psi_{cM}(t; B = 1.5, C = 1)$ is a good all-purpose wavelet with similar variance in both time and frequency (slightly better frequency resolution); I use this wavelet for the work in the rest of this chapter. See Appendix A.5 for further information on how Morlet wavelets are constructed. A wavelet with similar characteristics to the signal under investigation will improve results. The choice of wavelet to study the 21 cm cosmological signal could be explored in future work.

The continuous wavelets used for CWT analysis do not form an orthogonal basis, and the wavelet coefficients $\gamma_{m,a}$ are highly redundant. Thus, this representation of the signal is
5.2. The wavelet transform

Heisenberg Boxes

Figure 5.3: Heisenberg boxes of two Morlet wavelets with different translation and scaling factors, showing the time-frequency spread of the given wavelets. We can see that a scaling \( a \) in time corresponds to a dilation in frequency by \( 1/a \). The wavelet functions and their FTs are normalized to unity amplitude, and the lengths and widths of the boxes are not to scale with the wavelet functions and their FTs; this is done for pictorial purposes. As per the uncertainty principle, the minimum area of such a Heisenberg box is \( 1/4\pi \), which is achieved with Morlet wavelets (see Appendices A.4 and A.5).

inefficient; it is, however, more than adequate if we wish to analyse and better understand the spectral structure of a signal, especially through visual means. For optimal signal decomposition and eventual reconstruction, discrete wavelets are used.

We call the square of the modulus of the CWT the wavelet power spectrum (WPS):

\[
\text{WPS}_{m,a} = |\gamma_{m,a}|^2
\]

which is analogous to the standard PS, but has the additional dimension indexed by \( m \) (or \( \tau \) for the continuous case) that locates the spectral power in the signal domain. We predominantly use this measure to spectrally describe the signal using wavelets.

Let us not forget that in obtaining a time-frequency wavelet representation of a signal, the temporal and spectral information gained are constrained by the uncertainty principle (see Appendix A.4).
When dealing with interferometric visibilities, the signal lives in the frequency domain and its Fourier dual in the delay domain. The CWT and WPS graphed in this chapter, therefore, have units [Jy Hz] and [Jy^2 Hz^2] (like the FT and PS), unless stated otherwise. I use the PyWavelets\textsuperscript{a} (Lee et al. 2019) package for WT computations.

In Figure 5.4, I show the WPS for a sample HERA visibility on a frequency-delay heatmap, which we call a scaleogram. The visibility signal and its corresponding FT are also plotted on the sides of the scaleogram. I refer to Table 2.2 for conversion between delay and $k_{\parallel}$.

5.3 Error detection

As a first pass, scaleogram cubes with dimensions (scales, frequencies, times, baselines) were created. The napari\textsuperscript{b} (Sofroniew et al. 2022) multidimensional image viewing software was used to browse through the scaleograms for each time and baseline combination, for both Bands 1 and 2. While this is an intensive exercise, it provides a good feel for the data. Visual inspection of large amounts of data is laborious, meaning that anomalous modes can be missed. Furthermore, logarithmic scales can be deceiving; a more rigorous approach is needed.

I develop detection process whereby the modified-Z score (see Equation 2.37) is computed for each scaleogram, where the location and scales in the computation of $Z_{\text{mod}}$ are the median and the MAD across times and baselines, both of which are shown in Figure 5.5 for Band 2. CWT coefficients that have support outside of the cone of influence (CoI) are also masked to minimize false detections; practically, I mask coefficients with $\sigma_{\text{B1}} > 2 \times 10^1$ and $\sigma_{\text{B2}} > 7 \times 10^{-2}$ for Bands 1 and 2 (determined by essay), respectively, as the edge-effect artefacts leak further into the data than the estimated $\sqrt{2} s_{\text{CoI}}$.

The effect of flagging channels 570 – 573 (within Band 2) can be clearly seen in Figure 5.5: before the preprocessing pipelines, certain channel ranges are flagged for exhibiting narrowband features in the inspection of residual visibilities after wideband delay clean. These flagged channels are then inpainted with a clean model, which will be smooth. It is, therefore, expected to see lower power at high delays (low scales) for those channels. Interestingly, there is slightly elevated power for those channels from scales $\sim 7$, which corresponds to the 2 $\mu$s delay threshold used for the clean based inpainting. As we saw, though, in Figure 3.20, the final spherically averaged PS results are largely unchanged by the flagging of these channels.

Histograms of the number of power coefficients in the scaleograms as a function of $|Z_{\text{mod}}|$ following the outlier detection procedure are shown in Figure 5.6 for both bands. A fair number of outliers are detected from this method, with a non-negligible amount attaining extremely

\textsuperscript{a}https://github.com/PyWavelets/pywt
\textsuperscript{b}https://napari.org/
5.3. Error detection

Figure 5.4: CWT scaleogram, with complex visibility signal shown in the upper plot and its FT, using a Blackman-Harris window, shown on the right plot. Wavelet scales are from $2^{-59}$, corresponding to delays of $5.120 - 0.174 \mu s$; the starting wavelet scale of 2 is chosen to match the Nyquist frequency. The cone of influence (CoI) is clearly delineated in the scaleogram, showing the regions of frequency-time space that are corrupted by edge-effect artefacts. In the FT plot, I show $|\tilde{V}|$ for both positive and negative delays, for completeness. This particular example is for the final visibility of the 14 m (12, 13, pI) EW baseline for a 214 s time bin at LST 2.12 h (in Field 1) and for Band 1, after all the reduction steps from the analysis and preprocessing pipelines. This particular data slice is well-behaved, with no anomalies appearing across the three plots.
As an example of a particularly bad data slice, I look at baseline (66, 85, pl) at LST 8.52 h that has an isolated region of higher power around frequency 154 MHz and delay \(\approx 1\ \mu s\). I show its scaleogram with signal and FT appended to the relevant axes in Figure 5.7, and further examine it in comparison to its redundant group in Figure 5.8. The region of high power is a significant outlier with peak |\(Z_{\text{mod}}\)| = 103. It seems to indicate a transient event that is not present in adjacent time integrations or other baselines. The anomaly is not as appreciable when solely looking at the signal and Fourier domains, but clearly stands out by looking at the scaleograms and |\(Z_{\text{mod}}\)| heatmaps.

I further reduce the outlier detection data to make it more manageable and to potentially identify patterns. The maximum |\(Z_{\text{mod}}\)| is taken for each scaleogram and is plotted as a function of time and baseline in Figure 5.9 for all bands and fields. From this visualization, it is clear that certain time bins and baseline groups suffer more from these transient effects in their WPS: e.g. the bottom time bin of Field 1 (more prominent in Band 1) and top bin of Field 3 have noticeably more anomalies than the other bins. The short baseline groups, particularly those with length \(\approx 25\ m\) (corresponding to baseline indices 81 – 158), also seem to have more outliers, as seen from the brighter rectangular regions at the lower end of the baseline indices.

I repeat this analysis but look at the maximum |\(Z_{\text{mod}}\)| per antenna, with results plotted in Figure 5.10. The findings are similar to Figure 5.9, with the last time bin of Band 1 Field 1 and the first time bin of Band 2 Field 3 showing the worst outliers. There is no antenna that is
5.4 The 21 cm wavelet power spectrum

To demonstrate the potential of wavelets and their WPS as combined spatial and spectral estimator of the 21 cm signal, I form cross wavelet power spectra (CWPS) in a similar fashion evidently faulty, although it seems that antennas closer to the centre of the array configuration (see Figure 2.6a for array layout) seem to have a higher prevalence of strong outlier events.

As a further illustration of the affected baselines, I plot baselines that have outliers in their scaleograms $|Z_{\text{mod}}| > 5$ for Band 1 in Figure 5.11, with the methodology for drawing the baselines explained in the caption. From such plots, baselines consisting of antennas closer to the array’s centre can be seen to have a higher proportion of severe outliers.

From all these considerations, I infer that Band 1 as a whole is of worse quality than Band 2, with many more extreme outliers seen in the baseline and antenna heatmaps in Figures 5.9 and 5.10. These could be the source of the elevated power at lower $k$ modes seen in Figure 2.14 that does not agree with the theoretical noise floor. Moreover, the transient anomalous effects should be considered as per-baseline issues, and any additional flagging should be as such; there is no clear evidence to discard any antennas. Shorter baselines also seem to be more affected, and some time bins at the edges of the windows of observations for the fields also appear more problematic.

5.4 The 21 cm wavelet power spectrum

Figure 5.6: Number of points in scaleograms binned by $|Z_{\text{mod}}|$ score. 1.39% of Band 1 and 1.56% of Band 2 power coefficients exceed $|Z_{\text{mod}}| = 5$. Band 1 also has more extreme outliers than Band 2. Note that each count corresponds to a point on a scale-frequency scaleogram; therefore, if there is a region of higher power (consisting of many points) for a given scaleogram, that scaleogram will be counted multiple times.
Figure 5.7: Visibility, FT and scaleogram for visibility with baseline (66, 85, pI) for a 214 s time bin at LST 8.52 h (in Field 3) and for Band 2. Wavelet scales are from 2 – 59, corresponding to delays of 5.120 – 0.174 µs, same as Figure 5.4. We see a clear triangular shaped region of higher power at frequencies ≈ 154 MHz and delays ≈ 1 µs, which is not evident in the visibility and FT plots, despite corresponding to a peak $|\tilde{Z}_{\text{mod}}| = 103$ (as shown in Figure 5.8).
5.4. The 21 cm wavelet power spectrum

\[ |Z^{\text{mod}}| \]

\[ \text{WPS} \]

\[ \text{PS} [\text{Jy}^2 \text{Hz}^2] \]

\[ \text{Delay} [\mu s] \]

\[ \text{Scale} \]

\[ |Z^{\text{mod}}| \]

\[ \text{WPS} \]

\[ \text{PS} [\text{Jy}^2 \text{Hz}^2] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Delay} [\mu s] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Scale} \]

\[ |Z^{\text{mod}}| \]

\[ \text{WPS} \]

\[ \text{PS} [\text{Jy}^2 \text{Hz}^2] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Delay} [\mu s] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Scale} \]

\[ |Z^{\text{mod}}| \]

\[ \text{WPS} \]

\[ \text{PS} [\text{Jy}^2 \text{Hz}^2] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Delay} [\mu s] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Scale} \]

\[ |Z^{\text{mod}}| \]

\[ \text{WPS} \]

\[ \text{PS} [\text{Jy}^2 \text{Hz}^2] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Delay} [\mu s] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Scale} \]

\[ |Z^{\text{mod}}| \]

\[ \text{WPS} \]

\[ \text{PS} [\text{Jy}^2 \text{Hz}^2] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Delay} [\mu s] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Scale} \]

\[ |Z^{\text{mod}}| \]

\[ \text{WPS} \]

\[ \text{PS} [\text{Jy}^2 \text{Hz}^2] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Delay} [\mu s] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Scale} \]

\[ |Z^{\text{mod}}| \]

\[ \text{WPS} \]

\[ \text{PS} [\text{Jy}^2 \text{Hz}^2] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Delay} [\mu s] \]

\[ \text{Frequency} [\text{MHz}] \]

\[ \text{Scale} \]

| Figure 5.8: Top row: $|Z^{\text{mod}}|$ for the scaleogram shown in Figure 5.7 (left) and the mean WPS for the redundant group (right). Bottom row: plot of the complex visibilities (left) and their PS (right) for the baselines in the redundant group. The visibility and PS for the faulty baseline are highlighted. Areas of interest in frequency and delay space are also shaded in red; it is not entirely evident that these are areas of concern if purely looking at the visibility and PS plots. |

\[ \text{CWT scaleograms are formed for all baselines and times. Each scaleogram slice, containing the CWT coefficients for different scales and frequencies, is cross-multiplied with the one adjacent in time (separated by 214 s), with the cross-multiplication being done with all baselines in its redundant set (that share the same length and orientation). Considering all baseline permutations in the Fourier domain rather than redundant averaging in the visibility domain reduces the impact of baseline-dependent systematics. Cross-multiplying adjacent time bins to the HERA PS analysis:} \]

\[ \text{CWPS}_{m,a} = \gamma^i_{m,a} \gamma^j_{m,a} \] (5.8)
Figure 5.9: Maximum $|\mathcal{Z}^{\text{mod}}|$ for WPS scaleograms by baseline. Baseline length increases with baseline index. Only baselines taken forward for PS computation in the standard HERA pipeline are shown. The bottom time bin of Band 1 Field 1 and the top time bin of Field 3 (both bands) contain many extreme outliers. Some of the shorter baselines are also afflicted by strong outliers, indicated by the brighter regions in both bands.
Figure 5.10: Maximum $|\mathcal{Z}_{\text{mod}}|$ for WPS scaleograms by antenna. As with Figure 5.9, the biggest outliers are contained in the bottom time bin of Band 1 Field 1 and the top time bin of Field 3 (both bands). No antenna is clearly faulty from this metric.
Figure 5.11: Array layout of unflagged antennas with baselines drawn according to their $|Z^\text{mod}|$ score for Band 1. First, all baseline/time slices with (at least) a $|Z^\text{mod}| > 5$ data point in their scaleograms are found (there are 39,966), and the maximum $|Z^\text{mod}|$ is taken for each scaleogram. I then average the $|Z^\text{mod}|$ score across the time dimension, subtract the minimum $|Z^\text{mod}|$ score from all points and then divide by the maximum $|Z^\text{mod}|$ score to get a weight for each baseline. This weight is then used for the thickness and transparency parameters in drawing these baselines. The thicker and more opaque baselines are those with, on average, more serious outliers in their scaleograms. Baselines with central antennas appear to be more affected. This plot is merely illustrative, as there are many ways to fold the data and adjust the plotting parameters. Even so, from looking at various metrics, antennas 85, 86, 52, 40, 25 and 66 recur as being involved in baselines that suffer from the worst outliers, although no statistic is significant enough to justify discarding them.

further reduces any noise bias. Next, all such cross-power scaleograms are incoherently averaged across redundant baseline sets and across the time axis.

I show the final CWPS scaleograms for 14 m and 29 m EW baselines in Figures 5.12 and 5.13, respectively, for all bands and fields. These final scaleograms all show a horizontal feature at delay $\approx 0.4$ µs (or, in $k_\parallel$ units, $0.20 \, h \, \text{Mpc}^{-1}$ at $z = 10.4$, and $0.22 \, h \, \text{Mpc}^{-1}$ at $z = 7.9$), which is close to the HERA H1C_IDR2.2 upper limits set in Figure 2.14 and could be cause for concern. The computation time for the calculations to obtain the shown CWPS for all bands and fields from the fully reduced and calibrated pI visibilities is $\sim O(10 \, \text{min})$. 
5.4. *The 21 cm wavelet power spectrum*

![14 m EW CWPS](image)

**Figure 5.12:** CWPS for the 14 m EW baselines for all bands and fields of the H1C_IDR2.2 analysis.
Chapter 5. Wavelets and multiresolution analysis

29 m EW CWPS

<table>
<thead>
<tr>
<th>Band 1</th>
<th>Band 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redshift</td>
<td>Redshift</td>
</tr>
<tr>
<td>11.0</td>
<td>8.4</td>
</tr>
<tr>
<td>10.5</td>
<td>8.2</td>
</tr>
<tr>
<td>10.0</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>7.6</td>
</tr>
</tbody>
</table>

Field 1

Delay [µs]

Redshift

Band 1

10

10

Field 2

Delay [µs]

10

10

Field 3

Delay [µs]

10

10

Frequency [MHz]

10^{-5} 10^{-4} 10^{-3} 10^{-2} 10^{-1}

|CWPS|

10^{-1} 10^{-3} 10^{-5}

Figure 5.13: CWPS for the 29 m EW baselines for all bands and fields of the H1C_IDR2.2 analysis.
5.5. Discussion

A standardized way of presenting WPS results in 21 cm cosmology is needed. Converting the scaleograms in Figures 5.12 and 5.13 to cosmological units is a two-step process that involves first converting visibilities from Jy to mK and then multiplying the WPS by a cosmological scalar quantity to finally get units mK$^2$ h$^{-3}$ Mpc$^3$ (much like the conversion of the PS in Equation 2.21). Such a process, however, leads to added complexity, as $k_{\parallel} = k_{\parallel}(\eta, z)$ (where $\eta$ is delay, cf. Equation 2.19), so some regridding would be required since each vertical slice in the scaleograms of Figures 5.12 and 5.13 would be slightly shifted. In addition, WPS from different baselines must be combined appropriately. Moreover, the scaleograms plotted in this chapter are created using the CWT, which takes the product of the signal and wavelet at each frequency and for all specified wavelet scales. These scaled functions that are continuously shifted over the signal do not form an orthogonal basis; therefore, the resulting coefficients form a highly redundant representation of the signal. Wavelets can be discretized such that they are scaled and translated in discrete steps, with the modified transform operation called the discrete wavelet transform (DWT), which provides an optimal representation of the signal. The CWT remains useful, though, as this oversampling provides a better way of picturing the signal. See Appendix A.6 for a brief rundown of the DWT.

5.5 Discussion

Wavelets and their transform provide an alternative method of spectral decomposition of a signal while also retaining spatial locality. Such a multiresolution analysis is ideal when dealing with a dynamic signal and has been used extensively in other branches of physics. The 21 cm signal is expected to evolve over redshift (due to the light-cone effect), so any PS estimate over a considerable bandwidth will be biased, which in turn reduces the detectability of the signal. Wavelets have been an untapped resource for 21 cm cosmology, with few mentions of their application in the literature. They are, however, capable of providing an unbiased estimator of spectral power that can fully characterize the 3D nature of the EoR. Wavelets may eventually become the primary tool for 21 cm spectral analysis as data volumes grow (increased number of antennas and evenings of observation) and frequency resolution increases.

This chapter outlined the fundamentals of WT theory, and as a proof of concept I computed the CWPS for the 14 m and 29 m EW baselines. A standardized way of presenting WPS scaleograms is needed for consistency between different experiments, which may be challenging due to the increased dimensionality of the result. Simulations with e.g. 21cmFAST (Mesinger et al. 2011; Murray et al. 2020) are also needed to compute fiducial scaleograms for comparison.

The use of the WPS can also be extended to advanced error detection of fully calibrated and reduced visibilities. By inspecting frequency-delay scaleograms, anomalous regions of
higher than average power can be detected. RFI or other transient events would then manifest themselves as pyramids (see e.g. Figure 5.7), as the effect would first be noticed at high delays (low scales) and would then be subsequently detected by bigger wavelets (higher scales) at lower delays. The analysis of the data in this dual space also reveals structure that can be missed if looking at one of the dimensions alone.

From visual and $Z_{\text{mod}}$ considerations, scaleograms with regions of high power were identified and problematic baseline/time slices catalogued for each band separately. It is clear that some time bins at the edges of the fields of observation should have been discarded for PS analysis, as they contain a considerable number of extreme outliers that would corrupt any estimates after time averaging. The shorter 14 – 28 m baselines experience outlier events over particular time windows, with baselines within a redundant group appearing to be affected together. Further looking at outliers on a per-antenna basis does not reveal that any one antenna is prominently faulty, although there are signs that antennas closer to the centre of the array (and that are, thus, surrounded by more neighbouring antennas) are involved with more severe outliers.

Anomaly detection with WPS scaleograms presents itself as a great flagging tool that can spot erroneous visibility slices. With further automation, this technique can be used in production and can provide per-baseline flags for PS estimation.
This thesis has presented several robust statistical methods for radio interferometry that provide uncontaminated visibility estimates that are not unduly affected by RFI or other spurious effects, which can be used with confidence for 21 cm cosmology. These estimation techniques are statistically more rigorous than the de facto methods in the literature, and benefit from not requiring the same scrutiny vis-à-vis flagging. Commonly used flagging algorithms are ad hoc, can miss low-level effects, and often require considerable human oversight; it is, therefore, advantageous to supersede them by robust estimators.

Advances in radio astronomy are limited by non-thermal errors; the robust statistics outlined in this thesis suitably deal such non-Gaussianities. Robust statistics can be expensive, but many tools exist to speed up such computations. In this regard, this thesis also examined the balance between speed and efficiency, i.e. feasibility and accuracy, when implementing robust statistics.

In the introductory part of this thesis, Chapter 1 describes the lay of the land for our physical understanding of the Epoch of Reionization. The theory behind 21 cm cosmology is also reviewed, which presents itself as an auspicious probe to explore the early Universe. Chapter 2 runs through the mathematical formalism of radio interferometric measurement for the power spectrum computation of the 21 cm line, and introduces the Hydrogen Epoch of Reionization Array experiment and its data reduction pipeline.

In Chapter 3, I test for the normality of HERA visibilities aggregated over evenings of observation and over redundant baselines. With evidence for deviation from Gaussianity, the use of robust statistics is justified; I proceed to review their principles and I introduce several robust location estimators, including the geometric median. I use this latter estimator for LST-
averaging in the re-analysis of the results presented in The HERA Collaboration et al. (2022c), and find that the geometric median performs just as well, if not better, than the previously published method. I thus demonstrate that these robust location estimators can be relied upon for good estimates and that their application is far-reaching to all the different averaging axes encountered in radio interferometry. This chapter also includes research that uses multivariate outlier detection with robust Mahalanobis distances to better identify outliers in the complex plane than methods that do not take the covariance of the data distribution into account. This is a superior flagging algorithm to basic algorithms such as MAD-clipping that is ubiquitous in the literature. This chapter also emphasizes that complex data in an interferometric context should be wholly considered for any location or scatter computation, rather than treating the $\Re e$ and $\Im m$ components separately.

I broach the topic of calibration in Chapter 4, and generalize the maximum likelihood estimation formalism of redundant calibration to other distributions than a Gaussian. I work with both Gaussian and Cauchy assumed noise distributions, and find that in this particular instance, results for both distributions are very similar in most cases, likely due to the constriction of the redundant calibration problem. Still, calibration requirements need to be excellent for the purposes of a 21 cm detection, so assuming a non-Gaussian distribution for the noise is a conservative approach that avoids, insofar as possible, the absorption of glitches into calibration solutions. Degeneracies inherent to redundant calibration introduce an additional layer of complexity that need to be dealt with if comparing redundant visibility solutions. Research is, thus, conducted to compare redundant calibration solutions by solving for the degenerate parameters that translate between sets of solutions. I further show that this comparison can be used for outlier detection purposes. In a unified step, I build the framework for redundant calibration across multiple days that solves for a single set of visibilities and for per-antenna and per-day gains. This is an optimal calibration routine, as it wholly considers visibilities from a multi-JD dataset. Ultimately, this chapter demonstrates that we can extend beyond Gaussian assumed noise distributions in the calibration of radio data with the assistance of machine learning libraries to provide the required acceleration to make the computations tractable.

Lastly, in Chapter 5, I highlight the application of wavelets to 21 cm cosmology. Wavelet transforms enable us to get a dual representation of visibilities in both frequency and delay space. The computation of 21 cm wavelet power spectra grant power estimates as a function of redshift, instead of getting an average value (over a frequency band) that would otherwise be obtained from standard power spectrum estimation; this has the potential to paint a fuller picture of reionization at finer detail, and will likely surpass normal power spectrum methods as the favourite tool for spectral analysis of the 21 cm line. In this chapter, I additionally show that wavelet power spectra, visualized as scaleograms, can be used as an advanced flagging
mechanism; regions of high power in the dual frequency-delay space are indicative of outlier events. I show that, even after the full reduction of HERA visibilities that are used for power spectrum computation in The HERA Collaboration et al. (2022c), final pseudo-Stokes I visibilities in the selected bands and fields still show some level of contamination that inevitably end up corrupting power spectrum estimates. The multiresolution analysis detailed in this chapter offers an alternative way to examine the data; it is able to identify anomalous data slices from a dual-domain view, which may otherwise be missed if only looking at the data from one of the domains.

Future experiments, such as the SKA, will perform much of their data reduction in real time; the data volumes will be so large that the raw data simply cannot all be stored. This means that any calibration and averaging step need to be done right the first time: there is little margin for error. While many of the flagging procedures used in radio astronomy are increasingly sophisticated, much of it still is not automated; simple yet robust statistics could be hugely beneficial in providing uncorrupted estimates out of the box. The computation of robust statistics is expensive, which is part of the reason as to why they are not part of the mainstream astronomical data reduction procedures. It is only with the recent exponential increase in compute power that employing robust statistics has become feasible. There is now also a surfeit of ML libraries to choose from that accelerate high-level programs, requiring just slight and intuitive modification. For the optimizations performed in Chapters 3 and 4, packages such as jaxopt, which performs batched optimization on GPUs as part of the JAX ecosystem, show promise for great yet attainable speedup. The agenda of this thesis is to push forward the case of robust statistics to the field of radio interferometry; it is hoped that the effectiveness, ease of use and accelerability of the presented methods are attractive enough to warrant incorporation into existing and future reduction pipelines.

In the now James Webb Space Telescope (JWST) (Gardner et al. 2006) era of astronomy, we are able to better research the cosmic dawn of the Universe. Through its exceptional observational capabilities in the infrared (its primary imager, the Near Infrared Camera, is sensitive to wavelengths between 0.6 – 5 µm), the JWST will vastly improve our understanding of the first primordial luminous sources and the subsequent reionization of the IGM. We are obtaining pictures and spectrographs of the high-redshift Universe, and are even imaging typical galaxies from the cosmic dawn that reionized the Universe. At the time of submission, the JWST has already identified candidate galaxies up to \( z \sim 16 \) (Naidu et al. 2022; Finkelstein et al. 2022; Donnan et al. 2022), just weeks into its science mission (spectroscopic confirmation is still pending). These finding are not just serendipitous, they also indicate that there is likely a significant population of such UV luminous sources at \( z > 11 \), considering that these

---

[https://github.com/google/jaxopt](https://github.com/google/jaxopt)
findings are solely from an early data release that covers mere arcmins$^2$ on the sky. These early discoveries give a taste of things to come, and highlight the immense potential of deep JWST observations that will push the frontiers of astrophysics and cosmology. Observations from the JWST are complementary to those from 21 cm cosmology; such deep-field surveys at high redshifts with spectroscopic backing will teach us much about the cosmic dawn through the observation of primordial galaxies and quasars, but 21 cm observations (PS, global signal and tomography) will further shed light on the large-scale structure and evolution of the Universe, and will be able to see even further back, all the way to the edge of the dark ages. Radio interferometry, however, needs to continue to make progress if it is to stay at the forefront of astronomy. Next-generation experiments, such as the SKA, have been hindered by delays, and a 21 cm detection still remains elusive. Once constructed, much of the work will revolve around combatting RFI; the robust techniques presented in this thesis should help with this endeavour.
A.1 The Fourier transform

Throughout this thesis, I employ the ordinary unitary form of the Fourier transform (FT). The FT of a function $\xi(t)$ is given by $\mathcal{F}(\xi(t)) = \hat{\xi}(f)$, with the forward and inverse transforms defined as

$$\hat{\xi}(f) = \int_{-\infty}^{\infty} \xi(t) e^{-2\pi i f t} \, dt \quad (A.1)$$
$$\xi(t) = \int_{-\infty}^{\infty} \hat{\xi}(f) e^{2\pi i f t} \, df \quad (A.2)$$

thus forming the FT pair

$$\xi(t) \overset{\mathcal{F}}{\rightarrow} \hat{\xi}(f) \quad (A.3)$$

Equivalently, for finite equally-spaced sequences $x_n$, we compute the discrete Fourier transform (DFT) and its inverse, which are given by

$$X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i k n / N} \quad (A.4)$$
$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{2\pi i k n / N} \quad (A.5)$$

I use the fast Fourier transform (FFT) algorithm for all DFT computations for its reduced complexity from $O(N^2) \rightarrow O(N \log N)$, which was popularized by Cooley & Tukey (1965) but first discovered by Gauss in 1805 and published posthumously in Gauss (1866).
A.2 The power spectrum

The distribution of power of a time signal by frequency component is given by its power spectrum (PS). The energy of a signal is equal in both the time and frequency domain by the Plancherel identity

\[ ||\xi||^2 = \int_{-\infty}^{\infty} |\xi(t)|^2 \, dt = \int_{-\infty}^{\infty} |\hat{\xi}(f)|^2 \, df \]  

(A.6)

We, thus, define the PS as

\[ S(f) = |\hat{\xi}(f)|^2 = \hat{\xi}^*(f)\hat{\xi}(f) \]  

(A.7)

By the Wiener–Khinchin theorem, the PS is also equal to the FT of the two-point autocorrelation function, such that

\[ S(f) = \int_{-\infty}^{\infty} R_{\xi\xi}(t) e^{-2\pi i ft} \, dt = \hat{R}_{\xi\xi}(f) \]  

(A.8)

where

\[ R_{\xi\xi}(t) = \int_{-\infty}^{\infty} \xi(t')\xi^*(t-t) \, dt' = \int_{-\infty}^{\infty} \xi^*(t')\xi(t+t') \, dt' \]  

(A.9)

which is also commonly used to define the PS.

In the discrete case, these definitions become

\[ S_k = X_k^* X_k \]  

(A.10)

\[ = \sum_{n=0}^{N-1} R_{xx,n} e^{-2\pi i \frac{kn}{N}} \]  

(A.11)

with

\[ R_{xx,n} = \sum_{n'=0}^{N-1} \xi (n') \xi^*(n'-n) \]  

(A.12)

The approach in Equation A.10 is the most direct and can be efficiently implemented owing to the speed-up from the FFT algorithm.

A.2.1 The cross power spectrum

For two given signals \( \xi(t) \), \( \zeta(t) \), we can define the cross power spectrum (CPS) as

\[ C_{\xi\zeta}(f) = \int_{-\infty}^{\infty} R_{\xi\zeta}(t)e^{-2\pi i ft} \, dt = \hat{R}_{\xi\zeta}(f) \]  

(A.13)

where \( R_{\xi\zeta}(t) \) is the cross-correlation of \( \xi(t) \) with \( \zeta(t) \), as defined by

\[ R_{\xi\zeta}(t) = \int_{-\infty}^{\infty} \xi^*(t')\zeta(t+t') \, dt' \]  

(A.14)

\( C_{\xi\zeta}(f) \) can be found in a similar fashion by swapping the order of the signals. From these definitions, we find that \( C_{\xi\zeta}(f) = C_{\zeta\xi}(f) \), and the PS is a special case when \( \xi(t) = \zeta(t) \). In Equations A.7 and A.8, the (auto) PS \( S(f) \) is implied to mean \( S_{\xi\xi}(f) \).
Discrete analogies can also be made for the CPS. In practice, we compute the product of the complex conjugate of the FT of one signal with the FT of the other:

$$C_k = X_k^*Y_k$$ (A.15)

In CPS computation, we obtain a complex spectrum; for two coherent signals that aim to measure/detect the same phenomenon (e.g. redundant baselines in radio interferometry seeking to detect the 21 cm signal), we are usually only interested in the real part of the CPS, since that is where the phenomenon of interest lives, with the imaginary part being a realization of the noise.

In this thesis, I refer to the PS and CPS somewhat interchangeably, unless specified otherwise.

### A.3 Windowing

When performing an FFT, the two endpoints of the signal are interpreted as being connected to form a circularly continuous (i.e. periodic) signal. Modes with a non-integer number of cycles across the sample interval will then be discontinuous at the signal boundaries, which will induce structure in the frequency domain, therefore, distorting the spectrum. The finite length of the sampling, thus, causes spectral leakage. This can also be seen as multiplication of an infinite signal with a rectangular $\text{rect}(t)$, which has FT $\text{sinc}(\pi f)$.

A common way to mitigate spectral leakage is to apply a window function $W(t)$ to the data, which smoothly approaches zero at the edges of the specified interval. By the convolution theorem

$$\mathcal{F}(\hat{\xi}(t)W(t)) = \hat{\xi}(f) * \hat{W}(f)$$ (A.16)

Popular window functions are bell-like; they are symmetric and are maximal at the centre of the interval. The Blackman-Harris window (Harris 1978), which is used throughout this thesis in FT and PS computations, is shown in Figure A.1, and is given by

$$W_{BH}^n = 0.35875 - 0.48829 \cos \left( \frac{2\pi n}{N} \right) + 0.14128 \cos \left( \frac{4\pi n}{N} \right) - 0.01168 \cos \left( \frac{6\pi n}{N} \right)$$ (A.17)

Other popular windows include the Hann window (Blackman & Tukey 1958), $W_n^H = \sin^2(\pi n/N)$, or modified Gaussians.

### A.4 The uncertainty principle

Let us adjust the temporal spread of a function $\xi(t)$. We can contract/dilate the function by a factor $s > 0$ such that

$$\xi_s(t) = \frac{1}{\sqrt{s}}\xi\left(\frac{t}{s}\right)$$ (A.18)
Appendix A. The Fourier domain

where energy normalization is ensured, such that $\|\xi_s\|^2 = \|\xi\|^2$, where $\|\xi\|^2 = \int_{-\infty}^{\infty} |\xi(t)|^2 \, dt$ is the norm of $\xi(t)$. The spread of $\xi(t)$ is increased for $s > 1$ and decreased for $s < 1$. Taking the FT of $\xi_s(t)$, we obtain

$$\hat{\xi}_s(f) = \sqrt{s} \hat{\xi}(sf)$$  \hfill (A.19)

We see that the scaling has opposite effects in Equations A.18 and A.19. For a temporal dilation with $s > 1$, we observe a decrease in spread in the frequency domain. Thus, there is a trade-off in how localized the function is in time and frequency.

Taking the limiting case of a Dirac $\delta(t)$, the function is precisely localized in time. The FT gives $\hat{\delta}(f) = 1$; thus, the FT of the function is uniformly spread across all frequencies.
Conversely, a function $\varepsilon(t) = e^{iat}$ has FT $\hat{\varepsilon}(f) = \delta(f - \frac{a}{2\pi})$, so is completely localized in frequency but not in time.

This negotiation is at the foundation of time-frequency analysis and is quantified by the uncertainty principle.

**Theorem 1** (The uncertainty principle). *The temporal and frequency standard deviations of a function $\xi \in L^2(\mathbb{R})$ must satisfy*

$$\sigma_t \sigma_f \geq \frac{1}{4\pi} \quad (A.20)$$

**Proof.** Following Mallat (2009), we prove the uncertainty principle (with the original proof presented in Weyl 1928) for Schwartz class functions $\xi \in \mathcal{S}(\mathbb{R})$, which are infinitely differentiable functions $\xi : \mathbb{R} \to \mathbb{C}$ such that $\xi^{(n)}(t)$ rapidly decreases $\forall n \geq 0$; formally, we write this as

$$\sup_{t \in \mathbb{R}} |t^m \xi^{(n)}(t)| < \infty, \quad \forall m, n \geq 0 \quad (A.21)$$

The FT is well defined for $\xi \in \mathcal{S}(\mathbb{R})$, with $\mathcal{F} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$. For this proof, we only need Equation A.21 to hold for $m = 1/2, n = 1$. This proof can be extended to any square-integrable function $\xi(t) \in L^2(\mathbb{R})$, such that $\int_{-\infty}^{\infty} |\xi(t)|^2 \, dt < \infty$.

The temporal variance of $\xi(t)$ is given by

$$\sigma_t^2 = \frac{\int_{-\infty}^{\infty} (t - \bar{t})^2 \xi(t) \xi^*(t) \, dt}{\int_{-\infty}^{\infty} \xi(t) \xi^*(t) \, dt} = \frac{1}{\|\xi\|^2} \int_{-\infty}^{\infty} (t - \bar{t})^2 |\xi(t)|^2 \, dt \quad (A.22)$$

where $\bar{t}$ is the mean value of $\xi(t)$, which is given by

$$\bar{t} = \frac{1}{\|\xi\|^2} \int_{-\infty}^{\infty} t|\xi(t)|^2 \, dt \quad (A.23)$$

In the frequency domain, the variance is similarly given by

$$\sigma_f^2 = \frac{1}{\|\xi\|^2} \int_{-\infty}^{\infty} (f - \bar{f})^2 |\hat{\xi}(f)|^2 \, df \quad (A.24)$$

with

$$\bar{f} = \frac{1}{\|\xi\|^2} \int_{-\infty}^{\infty} f|\hat{\xi}(f)|^2 \, df \quad (A.25)$$

where we have used the Plancherel identity to get $1/\|\hat{\xi}\|^2 = 1/\|\xi\|^2$ in front of the integrals in Equations A.24 and A.25.

For mean localizations in time $\bar{t}$ and frequency $\bar{f}$ of $\xi(t)$, the mean average time and frequency location of $e^{-2\pi it\beta} \xi(t + \bar{t})$ are null, where we have simply applied a translation in time $\xi$ by $\bar{t}$ and a modulation in frequency by $\bar{f}$. Hence, we find it sufficient to prove the theorem for $\bar{t} = 0$ and $\bar{f} = 0$.
Let us consider \( \sigma_t^2 \sigma_f^2 \):

\[
\sigma_t^2 \sigma_f^2 = \frac{1}{\| \xi \|^4} \int_{-\infty}^{\infty} |t \xi(t)|^2 \, dt \int_{-\infty}^{\infty} |f \hat{\xi}(f)|^2 \, df
\]  
(A.26)

Now, \( \mathcal{F}(\xi'(t)) = 2\pi f \hat{\xi}(f) \), therefore, by the Plancherel identity we get

\[
\sigma_t^2 \sigma_f^2 = \frac{1}{4\pi^2 \| \xi \|^4} \int_{-\infty}^{\infty} |t \xi(t)|^2 \, dt \int_{-\infty}^{\infty} |\xi'(t)|^2 \, dt
\]  
(A.27)

We now then use the Cauchy-Schwarz inequality:

\[
\int |\xi(t)\xi^*(t)|^2 \, dt \leq \int |\xi(t)|^2 \, dt \int |\xi(t)|^2 \, dt
\]  
(A.28)

valid for \( \xi, \xi^* \in L^2(\mathbb{R}) \), to obtain

\[
\sigma_t^2 \sigma_f^2 \geq \frac{1}{4\pi^2 \| \xi \|^4} \left[ \int_{-\infty}^{\infty} |t \xi'(t)\xi^*(t)| \, dt \right]^2
\]  
(A.29)

\[
\geq \frac{1}{4\pi^2 \| \xi \|^4} \left[ \int_{-\infty}^{\infty} \frac{t}{2} (\xi''(t)\xi^*(t) + \xi^''(t)\xi(t)) \, dt \right]^2
\]  
(A.30)

\[
\geq \frac{1}{16\pi^2 \| \xi \|^4} \left[ \int_{-\infty}^{\infty} t \left( |\xi(t)|^2 \right)' \, dt \right]^2
\]  
(A.31)

Integrating by parts:

\[
\sigma_t^2 \sigma_f^2 \geq \frac{1}{16\pi^2 \| \xi \|^4} \left[ t|\xi(t)|^2 |_{-\infty}^{\infty} - \int_{-\infty}^{\infty} |\xi(t)|^2 \, dt \right]^2
\]  
(A.32)

The first term in brackets disappears for \( \xi \in S(\mathbb{R}) \) as \( \lim_{|t| \to \infty} t^{1/2} \xi^{(1)}(t) = 0 \). The bracket then becomes \( \| \xi \|^4 \), which is cancelled by the same factor in the denominator of the fraction out in front. Taking the square root on both sides, we obtain Equation A.20.

\[\Box\]

### A.5 Morlet wavelets

Morlet wavelets originate from the motivation to find a wavelet function \( \psi(t) \) that simultaneously minimizes its standard deviation in time and frequency.

From the uncertainty principle in Theorem 1, we know that there is a fundamental lower bound on the spreads in time and frequency, regardless of the original wavelet function \( \psi(t) \). To force equality on this lower bound, we must force equality in the Cauchy-Schwarz inequality, which is satisfied iff \( \xi(t) \propto \xi(t) \) in Equation A.28. The Cauchy-Schwarz inequality applied at Equation A.27, therefore, requires \( t\psi(t) \propto \psi'(t) \). We solve the differential equation

\[
\frac{d\psi(t)}{dt} = -2pt\psi(t)
\]  
(A.33)
A.6. The discrete wavelet transform

for some \( p \in \mathbb{C} \).

Rearranging and separating variables, we obtain

\[
\int \frac{d\psi(t)}{\psi(t)} = -2p \int t \, dt
\]

which gives us

\[
\ln \psi(t) = -pt^2 + c
\]

We, therefore, obtain Gaussian solutions of the form

\[
\psi(t) = q e^{-p(t-a)^2} e^{ibx}
\]

Therefore, we have shown that the unique family of signals that achieve the lower bound \( \sigma_t \sigma_f = \frac{1}{4\pi} \) are complex exponentials multiplied by a Gaussian, thus, giving the Morlet wavelet functions in Equation 5.5.

A.6 The discrete wavelet transform

The continuous wavelet transform (CWT) is a redundant measure as it uses a non-orthogonal set of wavelets, which means that the resulting coefficients are highly correlated. In addition, the translation \( \tau \) and scale \( a \) factors for the mother wavelets in Equation 5.4 are continuous, so we effectively have an infinite number of wavelets. For these reasons, we are motivated to reduce the parameterization of the WT in order to spectrally describe the signal with a minimal number of components.

In the discrete wavelet transform (DWT), we discretize the mother wavelet of Equation 5.4:

\[
\psi_{j,k}(t) = \frac{1}{\sqrt{a_0^j}} \phi \left( \frac{t - k\tau_0 a_0^j}{a_0^j} \right)
\]

where \( j, k \in \mathbb{Z} \) scale and translate the mother wavelet, such that \( j \) sets the wavelet's width and \( k \) sets its position. \( a_0 > 1 \) is a fixed dilation step and \( \tau_0 \) is the translation factor. The wavelets in Equation A.37 are usually continuous functions; it is the sampling in the time-scale space that is now discrete. The convention \( a_0 = 2, \tau_0 = 1 \) is usually followed, which corresponds to dyadic grid sampling; this is a natural choice and is also computationally efficient.

Any signal \( \xi(t) \) can, hence, be reconstructed by and expressed as the superposition

\[
\xi(t) = \sum_j \sum_k \zeta_{j,k} \psi_{j,k}(t)
\]

for wavelet coefficients \( \zeta_{j,k} \)

\[
\zeta_{j,k} = \int_{-\infty}^{\infty} \xi(t) \psi_{j,k}(t) \, dt
\]
as long as the energy of the wavelet coefficients are bound by constants $A, B$ (Daubechies 1990), such that

$$A\|\xi(t)\| \leq \sum_j \sum_k \left| \int_\infty^\infty \zeta_{j,k} \psi_{j,k}(t) \, dt \right|^2 \leq B\|\xi(t)\| \quad (A.40)$$

with $0 < A < B < \infty$. When $A = B$, the discrete wavelets form an orthonormal basis:

$$\int_\infty^\infty \psi_{j,k}(t) \psi_{j',k'}(t) \, dt = \delta_{j,j'} \delta_{k,k'} \quad (A.41)$$

For a complete signal representation, the spectrum needs to be covered all the way to zero; stretching the wavelet in time by a factor of 2 contracts the frequency response by a factor of 2; thus, as per Zeno’s dichotomy paradox, an infinite number of wavelets should still be needed. Practically, we choose a suitable cutoff (i.e. set a maximum $j$) and then fill this gap with a low-pass spectrum that we call the scaling function (or father wavelet). Intuitively, the scaling function extracts the global features that are missed by the wavelet coefficients. See Mallat (2009); Debnath & Shah (2015) for further information on the scaling function and wavelet decomposition/reconstruction.
B.1 robstat

I introduce basic functionality of the robstat\textsuperscript{a} Python package, which contains various robust statistical functions to estimate the locations of directional and multivariate data. The package requires an installation of R, the \texttt{rpy2}\textsuperscript{b} Python-R bridge, as well as other standard packages common in data science. R is called upon as it has packages available for the evaluation of depth measures and their associated medians; these are not yet found in Python.

The geometric and Tukey medians (see Sections 3.4.1.2 and 3.4.1.3) can be computed by calling \texttt{geometric\_median} and \texttt{tukey\_median} from robstat, respectively. The former is a SciPy minimization of the $L_1$ cost function (see Equation 3.31) using the BFGS algorithm, and the latter calls from the TukeyRegion R package.

Both of these median estimates are computed in the example below:

```python
import numpy as np
from robstat.robstat import geometric_median, tukey_median

# generate noisy data with some outliers
np.random.seed(1)
points = np.random.random(500).reshape(-1, 2)*2
points = np.concatenate((points+2, np.random.random(50).reshape(-1, 2)+5))
```

\textsuperscript{a}\url{https://github.com/matyasmolnar/robstat}
\textsuperscript{b}\url{https://github.com/rpy2/rpy2}
## Appendix B. Code implementations

### Figure B.1: Location estimates for noisy data with outliers.

The geometric and Tukey medians are not skewed by outliers, unlike the mean, and are at roughly the same location.

```python
# compute location estimates
sample_mean = np.mean(points, axis=0)
sample_gmed = geometric_median(points, weights=None)
sample_tmed = tukey_median(points)['barycenter']

med_ests = list(zip(['Mean', 'Geometric median', 'Tukey Median'],
                    [sample_mean, sample_gmed, sample_tmed]))

for med_est in med_ests:
    print('{}: {}'.format(med_est[0], med_est[1]))

# Mean : [3.21828555 3.26175702]
# Geometric median: [3.06956999 3.15071265]
# Tukey Median : [3.09257657 3.16488129]

These results are also plotted in Figure B.1.
```
B.2  
simpleredcal

The below sections show example code for redundant calibration, as described in Chapter 4, with the simpleredcal\textsuperscript{\textcopyright} \textsc{python} package

B.2.1  Relative redundant calibration with \textsc{JAX}

Below is example code for a relative redundant calibration run with Cauchy assumed noise distribution for a single frequency channel and single time bin. This minimization exercise is accelerated with \textsc{JAX}.

We start by loading a file from the HERA H1C_IDR2.2 dataset.

```python
from jax import numpy as jnp
from simpleredcal.red_likelihood import doRelCal, group_data, relabelAnts
from simpleredcal.red_utils import find_flag_file, find_zen_file, get_bad_ants

# Select dataset to calibrate
JD = 2458098.43869
pol = 'ee'  # polarization of data
chan = 605  # frequency channel
tint = 0    # time integration

zen_fn = find_zen_file(JD)  # find path of dataset
bad_ants = get_bad_ants(zen_fn)  # get bad antennas from commissioning
flags_fn = find_flag_file(JD, 'first')  # import flags from firstcal

# Load dataset from uvh5 file to numpy array, with flagging applied
hdr, RedG, cMData = group_data(zen_fn, pol=pol, chans=chan, tints=tint, 
    bad_ants=bad_ants, flag_path=flags_fn)
# 0 out of 741 data points flagged for visibility dataset
# zen.2458098.43869.HH.uvh5

cData = jnp.squeeze(cMData.filled())  # filled with nans for flags
no_ants = jnp.unique(RedG[:, 1:]).size  # number of antennas
```

\textsuperscript{\textcopyright}https://github.com/bnikolic/simpleredcal
no_unq_bls = jnp.unique(RedG[:, 0]).size  # number of redundant baselines

cRedG = relabelAnts(RedG)  # relabel antennas with consecutive numbering

We then perform relative redundant calibration.

res_rel = doRelCal(cRedG, cData, no_unq_bls, no_ants, distribution='cauchy', coords='cartesian', bounded=False, norm_gains=True)

# Optimization terminated successfully.

All the magic happens in the doRelCal function. The negative log-likelihood is calculated in relative_logLkl, which is partially filled and JIT’d, and appears in doRelCal as

from jax import jit

ff = jit(functools.partial(relative_logLkl, credg, distribution, obsvis, no_unq_bls, coords))

We then seek to minimize the negative log-likelihood (Equation 4.5), which is done through the SciPy minimization:

from jax import jacfwd, jacrev

jac = jacrev(ff)  # Jacobian; rev-mode faster for fewer outputs than # inputs
hess = jacfwd(jacrev(ff))  # Hessian; fwd-over-rev is more efficient

res = minimize(ff, initp, bounds=bounds, method=method, jac=jac, hess=hess, options={'maxiter':max_nit})

with the Jacobian being used for algorithms such as BFGS, L-BFGS-B and SLSQP, and the Hessian for the trust-region method (these are the tried and tested SciPy minimization methods that I used). Recently, a JAX version of SciPy’s minimize has been released, which can be found under jax.scipy.optimize.minimize. This solver, however, can only use the BFGS method for the time being.

B.2.1.1 Adding noise

The code for redundant calibration with noise included in the calculations requires little modification:

from jax import numpy as jnp
from red_likelihood import doRelCalD, group_data, relabelAnts
from red_utils import find_flag_file, find_zen_file, get_bad_ants

# Select dataset to calibrate
JD = 2458098.43869
pol = 'ee'  # polarization of data
chan = 605  # frequency channel
tint = 0    # time integration
noise_dist = 'gaussian'  # assumed noise distribution

zen_fn = find_zen_file(JD)  # find path of dataset
bad_ants = get_bad_ants(zen_fn)  # get bad antennas from commissioning
flags_fn = find_flag_file(JD, 'first')  # import flags from firstcal

# Load dataset from uvh5 file to numpy array, with flagging applied
hd_raw, RedG, cMData, cNoise = group_data(zen_fn, pol=pol, chans=chan, tints=tint, 
                                       bad_ants=bad_ants, flag_path=flags_fn, 
                                       noise=True)

# 0 out of 741 data points flagged for visibility dataset
# zen.2458098.43869.HH.uvh5

cData = jnp.squeeze(cMData.filled())  # filled with nans for flags
cNoise = jnp.squeeze(cNoise)
no_ants = jnp.unique(RedG[:, 1:]).size  # number of antennas
no_unq_bls = jnp.unique(RedG[:, 0]).size  # number of redundant baselines

cRedG = relabelAnts(RedG)  # relabel antennas with consecutive numbering

res_rel = doRelCalD(cRedG, cData, no_unq_bls, no_ants, distribution=noise_dist, 
                    noise=cNoise)
res_rel_vis, res_rel_gains = split_rel_results(res_rel['x'], no_unq_bls)

The only subtlety when changing to a minimization with noise input is that the gradient 
tolerance for termination must be changed. This is because the negative log-likelihood being 
minimized becomes rather big since the noise is small ($O(10^{-5})$). Example $-\ln(L_{rel})$ for 
minimizations with and without noise are shown in Figure B.2.
Appendix B. Code implementations

Figure B.2: \(-\ln(L_{rel})\) for redundant calibration minimizations with and without noise input, for the EE polarization of visibilities on JD 2458098 at LST 5.28 h (i.e. the 0th time integration of the dataset labelled by 2458098.43869). The shapes for in Figures B.2a and B.2b are inverted, and the ranges of values for successful iterations are: \(10^{-3} \leq -\ln(L_{rel}) \leq 10^{-1}\) and \(10^7 \leq -\ln(L_{rel}) \leq 10^9\).

B.2.2 Degenerate comparison of redundant visibility solutions

The code for comparing a pair of relatively calibrated visibility solutions for a particular frequency and time slice on separate JDs, and finding the degenerate parameters that translate between, is shown below.

Here, we use HERA data observed on JDs 2458098.43869 and 2458099.43124, taken from H1C_IDR2.2. We select frequency channel 605 and the zeroth time integration of the first dataset as an example. We only look at the EE polarization, assume that the noise follows a Gaussian distribution, and neglect the noise contribution (first term in Equation 4.23) in MLE calculations (effectively assuming it is uniform across baselines).

We start by loading the first HERA H1C_IDR2.2 dataset.

```python
from hera_cal.io import HERAData
from jax import numpy as jnp
from simpleredcal.red_likelihood import doDegVisVis, doRelCal, group_data, \
    red_ant_sep, relabelAnts, split_rel_results
from simpleredcal.red_utils import find_flag_file, find_nearest, find_zen_file, \
    get_bad_ants, match_lst

# Select 1st dataset to relatively calibrate
```
B.2. simpliedcal

JD1 = 2458098.43869
chan = 605 # frequency channel
time_int1 = 0 # time integration of 1st dataset
noise_dist = 'gaussian' # assumed noise distribution
coords = 'cartesian' # parameter coordinate system

zen_fn1 = find_zen_file(JD1) # find path of dataset
bad_ants1 = get_bad_ants(zen_fn1) # get bad antennas from commissioning
flags_fn1 = find_flag_file(JD1, 'first') # import flags from firstcal
print('Bad antennas for JD {} are: 
{}'.format(JD1, bad_ants1))
# Bad antennas for JD 2458098.43869 are:
# [0 2 11 24 50 53 54 67 69 98 122 136 139]

# Load dataset from uvh5 file to numpy array, with flagging applied
hdraw1, RedG1, cMData1 = group_data(zen_fn1, pol='ee', chans=chan, 
    tints=time_int1, bad_ants=bad_ants1, flag_path=flags_fn1)
# 0 out of 741 data points flagged for visibility dataset
# zen.2458098.43869.HH.uvh5
cData1 = jnp.squeeze(cMData1.filled()) # filled with nans for flags
ants = jnp.unique(RedG1[:, 1:])
no_ants = ants.size # number of antennas
no_unq_bls = jnp.unique(RedG1[:, 0]).size # number of redundant baselines
cRedG1 = relabelAnts(RedG1) # relabel antennas with consecutive numbering

We then load the second dataset on another JD that matches the first dataset in LST:

# Select 2nd dataset to relatively calibrate that matches the LST of the 1st
JD2 = match_lst(JD1, 2458099, tint=time_int1) # finding the JD_time of the dataset that matches the LST of the dataset used in 1
zen_fn2 = find_zen_file(JD2)
bad_ants2 = get_bad_ants(zen_fn2)
flags_fn2 = find_flag_file(JD2, 'first')

# Find the time integration in dataset 2 that corresponds to the closest LST to that of dataset 1
194

Appendix B. Code implementations

```python
hdraw2 = HERAData(zen_fn2)

# Load dataset from uvh5 file to numpy array, with flagging applied
_, RedG2, cMData2 = group_data(zen_fn2, pol='ee', chans=chan, 
    tints=time_int2, bad_ants=bad_ants2, flag_path=flags_fn2)
# 0 out of 741 data points flagged for visibility dataset
# zen.2458098.43869.HH.uvh5

cData2 = jnp.squeeze(cMData2.filled())  # filled with nans for flags

# Do the visibilities for JDs {} and {} have:
# the same bad antennas? {}
# the same redundant grouping? {}.

# Do the visibilities for JDs 2458098.43869 and 2458099.43124 have:
# the same bad antennas? True
# the same redundant grouping? True

Relative redundant calibration is now performed on both datasets, with Gaussian assumed
noise:

# Relative redundant calibration of the 1st dataset
res_rel1, initp = doRelCal(cRedG1, cData1, no_unq_bls, no_ants, 
    distribution=noise_dist, coords=coords, norm_gains=True, 
    return_initp=True)
# Optimization terminated successfully.

# Relative redundant calibration of the 2nd dataset
res_rel2 = doRelCal(cRedG1, cData2, no_unq_bls, no_ants, 
    distribution=noise_dist, coords=coords, norm_gains=True, 
    initp=initp, phase_reg_initp=True)
# Optimization terminated successfully.

# Get the relatively calibrated gain and visibility solutions
res_rel_vis1, res_rel_gains2 = split_rel_results(res_rel1['x'], no_unq_bls, 
    coords=coords)
res_rel_vis2, res_rel_gains2 = split_rel_results(res_rel2['x'], no_unq_bls, 
    coords=coords)
```

```
We now compare the redundant visibility solutions from both of these datasets, again, using Gaussian assumed noise:

```python
# Translating between relatively calibrated visibility sets
ant_sep = red_ant_sep(RedG1, hdraw1.antpos)
res_deg = doDegVisVis(ant_sep, res_rel_vis1, res_rel_vis2, 
                      distribution=noise_dist)
# Optimization terminated successfully.
print('Degenerate parameters are:
      Amplitude = {}
      Phase gradient in x = {:e}
      Phase gradient in y = {:e}
'.format(*res_deg['x']))
```

The \( \Re \) and \( \Im \) residuals for the above are plotted in Figure B.3. The results if we had assumed Cauchy noise (done with \( \text{noise\_dist} = \text{cauchy} \)) are very similar.

### B.2.2.1 Batch processing

The full relative redundant calibration of HERA datasets can be done with the `rel_cal.py` script:

```bash
python rel_cal.py '2458099.43124' --pol 'ee' --flag_type 'first' \ --dist 'gaussian'
```

```bash
python rel_cal.py '2458099.43869' --pol 'ee' --flag_type 'first' \ --dist 'gaussian'
```

```bash
python rel_cal.py '2458098.43869' --pol 'ee' --flag_type 'first' \ --dist 'gaussian' --initp_jd 2458099
```

In the last command, we impose the tilt constraint from Equation 4.26 to ensure that the solutions found have similar tilts as for JD 2458099. Note that in order to reuse the solutions for JD 2458099, the first two executions must have ended.
Appendix B. Code implementations

Figure B.3: Normalized residuals, by visibility amplitude, between visibility solutions on JDs 2458098 and 2458099 at LST 5.28 h for frequency channel 605. The baseline length increases and the number of baselines of each type decreases with increasing redundant baseline type ID. The 14 m and 29 m EW baselines are represented by baseline types 2 and 6, respectively. Note that the normalization is done by dividing the residuals by the geometric mean of the redundant visibility amplitude solutions for the two days being compared.

Comparison of two datasets by fitting for the degenerate parameters that translate between their relative redundant calibration visibility solutions can be made with deg_cal.py:

```python
deg_cal.py '2458098.43869' --deg_dim 'jd' --pol 'ee' \
--dist 'gaussian' --tgt_jd 2458099
```

Note that adjacent times and frequencies can be compared by specifying 'tint' and 'freq' after the deg_dim argument, respectively.

B.2.2.2 Multi-day comparison

Once the redundant calibration solutions for each JD and their degenerate offsets compared to the anchor day have been found, the degenerate comparison of redundant visibility solutions for a given LST range (see Section 4.4.4) can be done with the align_deg.py script, which
B.2. **simpleredcal**

degenerately transforms all the redundantly calibrated solutions to the same degenerate space as that of the anchor day, aligns the solutions in LST and then merges them in the same dataframe:

```python
python align_deg.py 2458098.43869 --jd_comp 'idr2_jdsx' --jd_anchor 2458099 \ --pol 'ee' --dist 'gaussian'
```

The first argument '2458098.43869' specifies the LST range and alignment of the resultant dataframe (such that time integrations are numbered according to those from HERA dataset zen.2458098.43869.HH.uvh5). The second argument 'idr2_jdsx' specifies that the JDs to use are those from H1C_IDR2; the x at the end of that argument indicates that JD 2458109 has been removed since that day is mostly flagged and antenna 137 is additionally flagged, which complicates the analysis for little reward; JD 2458109 is, therefore, usually omitted from analyses.

This script returns a dataframe for the complex visibility solutions for each redundant baseline group, indexed by frequency channel, time integration and JD.

### B.2.3 Redundant calibration across JDs

In the below example, we perform redundant calibration across all the days in H1C_IDR2.2 (except for JD 2458109), assuming Gaussian distributed noise.

```python
from jax import numpy as jnp
from simpleredcal.align_utils import idr2_jdsx
from simpleredcal.red_likelihood import doRelCalD, relabelAnts, split_rel_results
from simpleredcal.xd_utils import XDgroup_data

jdt = 2458098.43869  # dataset to use for LST alignment
JDs = idr2_jdsx  # days to consider in the redundant calibration
pol = 'ee'  # polarization of data
chan = 600  # frequency channel
time_int1 = 53  # time integration of 1st dataset
noise_dist = 'gaussian'  # assumed noise distribution

# Loading the data
hd, redg, cdata, cndata = XDgroup_data(jdt, JDs, pol, chans=chan, \
    tints=time_int1, bad_ants=True, \
    use_flags='first', noise=True)

cdata = jnp.squeeze(cdata.data)
```
Appendix B. Code implementations

cndata = jnp.squeeze(cndata)
no_unq_bls = jnp.unique(redg[:, 0]).size  # number of redundant baselines
no_ants = jnp.unique(redg[:, 1:]).size    # number of antennas

res_rel = doRelCalD(relabelAnts(redg), cdata, no_unq_bls, no_ants, 
    distribution='cauchy', noise=cndata, xd=True)
# Optimization terminated successfully.

Note that we used the default implementation doRelCalD instead of the development version doRelCalD. The former has been simplified and stripped of trial features found in doRelCal for improved legibility and (slight) speed-up.

B.2.3.1 Batch Processing

The full redundant calibration across JDs can be done with the xd_rel_cal.py script:

```
python xd_rel_cal.py '2458098.43869' --jds 'idr2_jd' --pol 'ee' --flag_type 'first' --dist 'gaussian'
```

with arguments similar to the multi-day degenerate comparison script in Appendix B.2.2.2.
<table>
<thead>
<tr>
<th><strong>ACRONYMS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AD</strong></td>
</tr>
<tr>
<td><strong>BFGS</strong></td>
</tr>
<tr>
<td><strong>BH</strong></td>
</tr>
<tr>
<td><strong>BIGHORNS</strong></td>
</tr>
<tr>
<td><strong>CASA</strong></td>
</tr>
<tr>
<td><strong>CDF</strong></td>
</tr>
<tr>
<td><strong>CDM</strong></td>
</tr>
<tr>
<td><strong>CHIME</strong></td>
</tr>
<tr>
<td><strong>CMB</strong></td>
</tr>
<tr>
<td><strong>CoI</strong></td>
</tr>
<tr>
<td><strong>CPS</strong></td>
</tr>
<tr>
<td><strong>CPU</strong></td>
</tr>
<tr>
<td><strong>CWPS</strong></td>
</tr>
<tr>
<td><strong>CWT</strong></td>
</tr>
<tr>
<td><strong>DFT</strong></td>
</tr>
<tr>
<td><strong>DLA</strong></td>
</tr>
<tr>
<td><strong>DM</strong></td>
</tr>
<tr>
<td><strong>DWT</strong></td>
</tr>
<tr>
<td><strong>EDGES</strong></td>
</tr>
<tr>
<td><strong>EE</strong></td>
</tr>
<tr>
<td><strong>EoR</strong></td>
</tr>
<tr>
<td><strong>EW</strong></td>
</tr>
<tr>
<td>Acronym</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>FD</td>
</tr>
<tr>
<td>FFT</td>
</tr>
<tr>
<td>FoV</td>
</tr>
<tr>
<td>FT</td>
</tr>
<tr>
<td>FWHM</td>
</tr>
<tr>
<td>GC</td>
</tr>
<tr>
<td>GLEAM</td>
</tr>
<tr>
<td>GMRT</td>
</tr>
<tr>
<td>GP</td>
</tr>
<tr>
<td>GPU</td>
</tr>
<tr>
<td>HDF</td>
</tr>
<tr>
<td>HERA</td>
</tr>
<tr>
<td>HLO</td>
</tr>
<tr>
<td>HPF</td>
</tr>
<tr>
<td>HWHM</td>
</tr>
<tr>
<td>IGM</td>
</tr>
<tr>
<td>IMF</td>
</tr>
<tr>
<td>IR</td>
</tr>
<tr>
<td>ISM</td>
</tr>
<tr>
<td>JD</td>
</tr>
<tr>
<td>JIT</td>
</tr>
<tr>
<td>JWST</td>
</tr>
<tr>
<td>KDE</td>
</tr>
<tr>
<td>kSZ</td>
</tr>
<tr>
<td>LEDA</td>
</tr>
<tr>
<td>LHS</td>
</tr>
<tr>
<td>LOFAR</td>
</tr>
<tr>
<td>LoS</td>
</tr>
<tr>
<td>Acronym</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>LST</td>
</tr>
<tr>
<td>MAD</td>
</tr>
<tr>
<td>MCD</td>
</tr>
<tr>
<td>ML</td>
</tr>
<tr>
<td>MLE</td>
</tr>
<tr>
<td>MWA</td>
</tr>
<tr>
<td>NN</td>
</tr>
<tr>
<td>NRAO</td>
</tr>
<tr>
<td>NS</td>
</tr>
<tr>
<td>PAPER</td>
</tr>
<tr>
<td>PS</td>
</tr>
<tr>
<td>PSF</td>
</tr>
<tr>
<td>QSO</td>
</tr>
<tr>
<td>REACH</td>
</tr>
<tr>
<td>RFI</td>
</tr>
<tr>
<td>RHS</td>
</tr>
<tr>
<td>RMD</td>
</tr>
<tr>
<td>RMS</td>
</tr>
<tr>
<td>S/N</td>
</tr>
<tr>
<td>SARAS</td>
</tr>
<tr>
<td>SCI-HI</td>
</tr>
<tr>
<td>SKA</td>
</tr>
<tr>
<td>SLSQP</td>
</tr>
<tr>
<td>SMBH</td>
</tr>
<tr>
<td>STFT</td>
</tr>
<tr>
<td>TPU</td>
</tr>
<tr>
<td>Acronym</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>UV</td>
</tr>
<tr>
<td>WIMPs</td>
</tr>
<tr>
<td>WPS</td>
</tr>
<tr>
<td>WT</td>
</tr>
<tr>
<td>XLA</td>
</tr>
<tr>
<td>ZEBRA</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY

Barkana R., Loeb A., 2001, Physics Reports, 349, 125
Bernardi G., et al., 2009, Astronomy and Astrophysics, 500, 965
Blackman R. B., Tukey J. W., 1958, Bell System Technical Journal, 37, 185
Bromm V., 2013, Reports on Progress in Physics, 76, 112901

203
Byrne R., 2019, *HERA Memorandum #63: A Comparison of Two Absolute Calibration Methods*
Caprini C., Figueroa D. G., 2018, *Classical and Quantum Gravity*, 35, 163001
Dillon J. S., 2019, *HERA Memorandum #69: H1C IDR 2.2*
Field G. B., 1958, Proceedings of the IRE, 46, 240
Fritz H., Filzmoser P., Croux C., 2012, Computational Statistics, 27, 393
Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Physics Reports, 433, 181
Gardner J. P., et al., 2006, Space Science Reviews, 123, 485
Loeb A., 2006, Scientific American, 295, 46
Mahalanobis P. C., 1936, Proceedings of the National Institute of Science of India, 2, 49
Margossian C. C., 2019, WIREs Data Mining and Knowledge Discovery, 9, e1305
Mellema G., et al., 2013, Experimental Astronomy, 36, 235
Molnar M., Nikolic B., 2020, HERA Memorandum #84: A Generalized Approach to Redundant Calibration with JAX
Molnar M., Nikolic B., 2021a, HERA Memorandum #106: Non-Gaussian Effects and Robust Location Estimates of Aggregated Calibrated Visibilities
Molnar M., Nikolic B., 2021b, HERA Memorandum #94: Comparing Visibility Solutions from Relative Redundant Calibration by Degenerate Translation
Molnar M., Nikolic B., 2022, HERA Memorandum #110: Multivariate Outlier Detection Using Robust Mahalanobis Distances
Bibliography

Nambissan T. J., et al., 2021, Experimental Astronomy, 51, 193
Oja H., 1983, Statistics & Probability Letters, 1, 327
Patra N., Subrahmanyan R., Raghunathan A., Udaya Shankar N., 2013, Experimental Astronomy, 36, 319
Planck Collaboration et al., 2020, Astronomy and Astrophysics, 641, A6
doi:10.7551/mitpress/3206.001.0001
Rousseeuw P. J., Driessen K. V., 1999, Technometrics, 41, 212
Savitzky A., Golay M. J. E., 1964, Analytical Chemistry, 36, 1627
Schaeper D., 2003, Astronomy and Astrophysics, 397, 527
Shapiro S. S., Wilk M. B., 1965, Biometrika, 52, 591
Singh S., et al., 2022, Nature Astronomy, 6, 607
Bibliography

The HERA Collaboration et al., 2022a, preprint (arXiv:2210.04912)
The HERA Collaboration et al., 2022b, The Astrophysical Journal, 924, 51
The HERA Collaboration et al., 2022c, The Astrophysical Journal, 925, 221
Vardi Y., Zhang C.-H., 2000, Proceedings of the National Academy of Science, 97, 1423
Verma A., 2000, Current Science, 78, 804
Weiszfeld E., 1937, Tohoku Mathematical Journal, First Series, 43, 355
Weyl H., 1928, Gruppentheorie und Quantenmechanik. S. Hirzel, Leipzig
Wieringa M. H., 1992, Experimental Astronomy, 2, 203
Woosley S. E., Heger A., Weaver T. A., 2002, Reviews of Modern Physics, 74, 1015
Bibliography

Zel’dovich Y. B., 1970, Astronomy and Astrophysics, 5, 84
Zernike F., 1938, Physica, 5, 785