Radio Recombination Line Forecasts with the Canadian Hydrogen Observatory and Radio-transient Detector (CHORD)

by

Duncan Cameron-Steinke

B.Sc., Queen's University, 2020

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE COLLEGE OF GRADUATE STUDIES

(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

September 2026

© Duncan Cameron-Steinke, 2026

The following individuals certify that they have read, and recommend to the College of Graduate Studies for acceptance, a thesis entitled:

Radio Recombination Line Forecasts with the Canadian Hydrogen Observatory and Radio-transient Detector (CHORD)

submitted by Duncan Cameron-Steinke in partial fulfilment of the requirements of the degree of Master of Science

Alex S. Hill, Irving K. Barber Faculty of Science
Supervisor

John Braun, Irving K. Barber Faculty of Science
Supervisory Committee Member

Timothy Robishaw, Irving K. Barber Faculty of Science
Supervisory Committee Member

Jonathan Holzman, Faculty of Applied Science
University Examiner

Abstract

The Canadian Hydrogen Observatory and Radio-transient Detector (CHORD) is a 512-dish radio interferometer telescope under construction at the Dominion Radio Astrophysical Observatory (DRAO) near Penticton, British Columbia. This work estimates the capabilities of CHORD for detecting Radio Recombination Lines (RRLs) in the frequency range of 300-1500 MHz. This work analyzes the Radio Frequency Interference (RFI) present at DRAO and finds that 79 hydrogen Alpha RRL transitions are likely to be detectable. This work derives the theoretical sensitivity of CHORD to diffuse sources and explores the expected signal-to-noise gains by stacking multiple RRLs. We propose site locations to add additional dishes as part of Extension CHORD, and find that adding 32 dishes across two locations can improve angular resolution by 44% while decreasing sensitivity by 25%. Finally, we take published results of the Survey of Ionized Gas in the Galaxy Made with the Arecibo Telescope (SIGGMA), and estimate CHORD's detection capabilities given those results. We find that CHORD should have the sensitivity to match each of SIGGMA's detections using four observing days and can improve upon each SIGGMA detection in both sensitivity and angular resolution after 9 observing days.

Lay Summary

The Canadian Hydrogen Observatory and Radio-transient Detector (CHORD) is a radio telescope under construction at the Dominion Radio Astrophysical Observatory (DRAO) near Penticton, British Columbia. A scientific goal for the telescope is to map the extremely faint radio waves originating from distant gas clouds in our Milky Way Galaxy. To accomplish this, the telescope must be extremely sensitive at specific radio frequencies. In this thesis, we analyze which frequencies are usable by CHORD and which are corrupted by interference from terrestrial radio emitters. We calculate the theoretical sensitivity which we expect CHORD to achieve and provide data processing strategies that can improve sensitivity. We study the impact of adding additional antennas to CHORD and propose a design which optimizes scientific outcomes. Finally, we compare the expected performance of CHORD to previously published surveys and find that CHORD is likely to be more precise, likely allowing for new scientific discoveries.

Preface

This thesis was wholly written by Duncan Cameron-Steinke, who also performed all of the analysis presented in this thesis. This work was supervised by Dr. Alex Hill, with additional guidance from Dr. John Braun and Dr. Tim Robishaw. Dr. Nicholas Bruce assisted in the technical description of the RFI monitor and provided guidance for data processing of the RFI monitor data. Dr. Théo Michelot previously served as a committee member and provided continued guidance and support.

Generative AI in the form of Claude Code Opus 4.8 was used to accomplish specific, tightly scoped, coding tasks, such as generating plots and debugging. I carefully reviewed all generated code to ensure correctness and accuracy. Generative AI in the form of Claude Code Opus 4.8 was used to edit versions of this thesis for grammar, clarity and to review accuracy of stated concepts and mathematical derivations. Generative AI was used as a resource while researching this thesis to provide background and answer questions.

Generative AI was not used to write any sections of this thesis and was not used to accomplish any of the analysis presented in the thesis.

Table of Contents

Abstract . . . . . **iii**

Lay Summary . . . . . **iv**

Preface . . . . . **v**

Table of Contents . . . . . **vi**

List of Tables . . . . . **ix**

List of Figures . . . . . **x**

Acknowledgements . . . . . **xiii**

Dedication . . . . . **xiv**

Chapter 1: Introduction **1**
1.1 Interstellar Medium 1
1.2 Radio Recombination Line Emission 2
1.3 CHORD Telescope 4

Chapter 2: RFI Analysis **6**
2.1 RFI Impact 6
2.1.1 Definition of RFI 6
2.1.2 RFI Impact on RRL Science 6
2.2 RFI Detection Methods 7
2.2.1 Excess Kurtosis for Non-Gaussianity 7
2.2.2 Spectral Kurtosis for Power Spectra 8
2.3 DRAO RFI Monitor 10
2.3.1 RFI Monitor Specifications 10
2.3.2 RFI Monitor Data Statistics 10

2.4 Application of RFI Detection 12
2.4.1 Excess Kurtosis RFI Detection 12
2.4.2 Spectral Kurtosis RFI Detection 15
2.4.3 Radio Recombination Line Visibility 18
2.4.4 Analysis Limitations 21

Chapter 3: CHORD Sensitivity Analysis 23
3.1 Point Source Sensitivity 23
3.2 Angular Resolution 27
3.3 Brightness Sensitivity 29
3.4 Sensitivity with Radio Background 31

Chapter 4: Radio Line Stacking 33
4.1 Inverse-Variance-Weighted Average 33
4.2 Signal-to-Noise Ratio by Sky Region 34
4.2.1 Signal-to-Noise Ratio after Smoothing 36

Chapter 5: Extension CHORD 39
5.1 Candidate Site Selection 39
5.2 (u, v) Coverage and Synthesized Beam 41
5.3 Weighting Methods 41
5.4 Evaluating Beam Patterns 43
5.5 Optimal Extension CHORD Configurations 43

Chapter 6: Simulations 47
6.1 Simulations of SIGGMA Observations 47

Chapter 7: Conclusion 53
7.1 Conclusions 53

Bibliography 55

Appendix 58
Appendix A: RFI 59
A.1 Log-Gamma vs Gaussian Distributions 59
A.1.1 Gamma Distribution 59
A.1.2 Log-Gamma Distribution 60
A.2 Spectral Kurtosis Pair Plot 62
A.3 Alternative RFI Thresholds 62
Appendix B: Extension CHORD 65
B.1 Additional Extension CHORD Plots 65

B.1.1 Pairwise Comparisons 65
B.1.2 (u, v) Coverage of B + D 65
B.1.3 Beam Plots for Each Individual Location 65
B.1.4 Beam Plots for Each Location Pair 65
B.1.5 (u, v) Coverage and Beam Pattern for the Complete Array 66
Appendix C: Appendix C: SIGGMA Results 79
C.1 Complete SIGGMA Tables 79
C.1.1 SIGGMA H II Regions 79
C.1.2 SIGGMA Stacked SNR Predictions 80
C.1.3 SIGGMA Stacked SNR Predictions with Smoothing . 81

List of Tables

Table 2.1 Detectable lines by frequency range 20

Table 5.1 Extension CHORD candidate locations 40

Table 6.1 A subset of H II region results within CHORD coverage 48
Table 6.2 SNR after stacking the best 1, 15 and 60 lines for CHORD observations of SIGGMA H II regions 51
Table 6.3 SNR for CHORD observations of SIGGMA H II regions with smoothed angular resolution 51

Table A.1 Detectable RRLs by aggregation and SK threshold . . 63

Table C.1 Set of SIGGMA H II regions and candidate H II regions visible to CHORD. 80
Table C.2 Predicted SNR for CHORD observations of SIGGMA H II regions using one observing day, showing the SNR from stacking the highest 1, 15 and 60 stacked lines. . 81
Table C.3 SNR for CHORD observations of SIGGMA H II regions with a smoothed angular resolution of 5, 6, and 8 arcminutes. 82

List of Figures

Figure 1.1 CHORD survey range 5

Figure 2.1 Sample of RFI moniter data 10
Figure 2.2 RFI monitor gain factor effects on mean power at 1420 MHz 13
Figure 2.3 Excess kurtosis histogram for simulated and observed RFI-free spectrum 14
Figure 2.4 Power and excess kurtosis from 1400 MHz to 1430 MHz 15
Figure 2.5 Power and excess kurtosis from 590 MHz to 610 MHz 16
Figure 2.6 Spectral kurtosis shape parameter calibration 17
Figure 2.7 Spectral kurtosis histograms for simulated and observed RFI-free spectrum 17
Figure 2.8 Power and spectral kurtosis from 590 MHz to 610 MHz 18
Figure 2.9 Spectral kurtosis histogram across the full CHORD band 19
Figure 2.10 RFI labels for full CHORD spectrum 20
Figure 2.11 RRL detectability scores for each RRL in the CHORD range 21

Figure 3.1 Comparison of Airy disk and Gaussian beam 25
Figure 3.2 CHORD primary beam attenuation 26
Figure 3.3 CHORD point source sensitivity 27
Figure 3.4 CHORD and Pathfinder angular resolution as a function of observing frequency. 28
Figure 3.5 CHORD and Pathfinder surface brightness sensitivity 29
Figure 3.6 CHORD and Pathfinder brightness temperature sensitivity 30
Figure 3.7 CHORD brightness temperature sensitivity maps assuming empty sky 30
Figure 3.8 Radio background maps 31

Figure 3.9 Brightness temperature sensitivity maps with radio background 32

Figure 4.1 Sample SNR locations map 35
Figure 4.2 SNR per-line by-sky location with native velocity and angular resolutions 35
Figure 4.3 SNR of system only effects and background only effects 36
Figure 4.4 SNR after smoothing to a common velocity and angular resolutions 38

Figure 5.1 Site map for Extension CHORD 40
Figure 5.2 Core array (u, v) coverage and synthesized beam . . . 42
Figure 5.3 Beam patterns with striping for location A 44
Figure 5.4 Uniform-weighted resolution for extension CHORD location pairs 45
Figure 5.5 Locations B, D and B+D beam patterns 46

Figure 6.1 SIGGMA data with CHORD coverage 48
Figure 6.2 H II regions over CHORD sensitivity 48
Figure 6.3 Single-day per-line SNR of SIGGMA H II regions . . 49
Figure 6.4 Single-day per-line SNR of SIGGMA H II regions, smoothed to common resolution 50

Figure A.1 Pair plot of RFI label agreement 62

Figure B.1 Comparison of each location pairing 67
Figure B.2 (u, v) coverage of locations B, D, and B + D 68
Figure B.3 Beam plots for locations A, A2, B, and C 69
Figure B.4 Beam plots for locations D, E, F, and G 70
Figure B.5 Beam plots for location pairs A+A2, A+B, A+C, and A+D 71
Figure B.6 Beam plots for location pairs A+E, A+F, A+G, and A2+B 72
Figure B.7 Beam plots for location pairs A2+C, A2+D, A2+E, and A2+F 73
Figure B.8 Beam plots for location pairs A2+G, B+C, B+D, and B+E 74
Figure B.9 Beam plots for location pairs B+F, B+G, C+D, and C+E 75
Figure B.10 Beam plots for location pairs C+F, C+G, D+E, and D+F 76

Figure B.11 Beam plots for location pairs D+G, E+F, E+G, and F+G 77
Figure B.12 (u, v) coverage and uniform-weighted beam pattern for the complete array including all candidate locations 78

Acknowledgements

Thank you to my supervisor Dr. Alex Hill, for being consistently available throughout my studies to answer questions and provide feedback. Also thank you Alex, for allowing me the flexibility to pursue my ever-changing research interests and always helping me along the way.

Thank you to my committee members for their guidance and support throughout my research. To Dr. John Braun thank you for the statistical feedback, to Dr. Tim Robishaw thanks for the detailed review which greatly improved this work, and to Dr. Théo Michelot thanks for your mentorship and encouragement which allowed me to expand my research scope.

Thank you to all the members of the UBCO/UofC ISM group, Dr. Anna Ordog, Dr. Tom Landecker, Dr. Jo-Anne Brown, Dr. Jennifer West, Dr. Rebecca Booth, Ciara Chisholm, Artem Davydov, Dr. Josh MacEachern, and Veyanna Hassani, for the many reviews and discussions.

Thank you to the CHORD spectral lines group, Dr. Kristine Spekkens, Akanksha Bij, Hans Hopkinson, and Josh Goodeve, for your contributions.

Dedication

To Rebeka, thank you for your unwavering support.

Chapter 1

Introduction

Radio recombination lines (RRLs) trace ionized gas in the interstellar medium (ISM), providing information on the structure and evolution of our Galaxy. This thesis forecasts the ability of the Canadian Hydrogen Observatory and Radio-transient Detector (CHORD) to detect RRLs. These forecasts allow us to predict the scientific outcomes of an RRL survey and to optimize survey strategies toward specific science goals. The remainder of this chapter provides background on the ISM, RRL emission, and the CHORD telescope. In later chapters, we analyze the impact of radio frequency interference (RFI) on RRL detection, derive CHORD's expected sensitivity, explore line stacking to improve detection, and evaluate the performance benefits of extending the array with additional dishes. Finally, we compare CHORD's expected performance to the Survey of Ionized Gas in the Galaxy, Made with the Arecibo Telescope (SIGGMA), and show that CHORD can improve on SIGGMA in both sensitivity and angular resolution.

1.1 Interstellar Medium

Stars are born in the coalescing gas and dust that permeate our Milky Way Galaxy, known as the ISM [5]. As stars evolve, they recycle material to the ISM through stellar winds and supernovae, enriching the ISM with new elements. To understand the lifecycle of stars, it is important to study the nature of the ISM. Beyond its importance in star formation, the ISM traces the structure and dynamics of our Milky Way Galaxy and other galaxies [5], allowing for the study of galactic evolution. Finally, the ISM impacts astronomical observations of everything from nearby stars to distant galaxies, making its study a crucial component of many other areas of astrophysics.

Part of the ISM is interstellar dust, which absorbs or scatters visible and ultraviolet light in a process called interstellar extinction. Dust extinction acts like a veil, limiting the distance that visible light can travel, which limits the area we can study through visible light. Radio waves do not suffer

from dust extinction, so radio observations can reveal regions of the ISM that are inaccessible to visible light.

The dominant component of the ISM is hydrogen gas, comprising approximately 70% of the ISM by mass, with the remainder consisting of helium and trace amounts of heavier elements [5]. Most hydrogen in the diffuse interstellar medium is neutral hydrogen (H I), which contains one electron in the ground state. Neutral hydrogen is very stable and is generally incapable of producing emission or absorption lines by the transition of the electron between orbitals. The 21-cm line, which arises from electron spin reversals in neutral hydrogen, is the main method used to detect neutral hydrogen.

When neutral hydrogen absorbs a photon in excess of 13.6 eV, the electron is ejected from the hydrogen atom, producing ionized hydrogen (H II). When the electron recombines, it can cascade downwards through the energy levels, producing a number of low-energy photons. Some very hot stars, or regions of very hot stars, can produce sufficient radiation to fully ionize the surrounding hydrogen gas in what are called H II regions. For these regions to maintain local thermodynamic equilibrium (LTE), the rate of ionization must equal the rate of recombination, which produces continuous emission of low-energy photons. We discuss emission in the radio range in more technical detail in section 1.2.

1.2 Radio Recombination Line Emission

RRLs are spectral lines emitted in an ionized medium when an electron recombines with an ion and then cascades through successively lower energy levels. As the negatively charged electron moves closer to the positive nucleus, it decreases its electric potential energy, and the excess energy is emitted as a photon. Since the energy levels of an atom are quantized, the emitted photons are also quantized and appear only at specific frequencies. Photon energy and frequency are interchangeable quantities related by $E = h\nu$, where E is the photon energy, h is Planck's constant, and ν is the frequency. RRLs specifically refer to photons in the radio frequency range, which for a hydrogen atom are emitted when electrons transition between high energy levels (e.g., $n = 100$ to $n = 99$). The exact frequency of RRLs can be calculated using the Rydberg formula, which for hydrogen is given

by

$$\nu = R_H c \left(\frac{1}{n^2} - \frac{1}{(n+\Delta n)^2} \right), \tag{1.1}$$

where R_H is the Rydberg constant for hydrogen, c is the speed of light, n is the final quantum number, and Δn is the change in quantum numbers. Transitions where $\Delta n = 1$ are called alpha transitions; those where $\Delta n = 2$ and $\Delta n = 3$ are called beta and gamma transitions, respectively.

Line intensity estimates in the radio range assuming optically thin emission and local thermodynamic equilibrium are [1]

$$\left(\frac{T_L}{\mathrm{K}}\right) \backsimeq 3076 \left(\frac{T_e}{\mathrm{K}}\right)^{-3/2} \left(\frac{\mathrm{EM}}{\mathrm{pc\,cm^{-6}}}\right) \left(\frac{\nu_0}{\mathrm{GHz}}\right)^{-1} \left(\frac{\Delta V}{\mathrm{km\,s^{-1}}}\right)^{-1} \Delta n M_{\Delta n}, \tag{1.2}$$

where T_e is the electron temperature, EM is the emission measure, ν_0 is the line frequency, ΔV is the FWHM line width, and $M_{\Delta n}$ is a quantum mechanical correction factor of $M_{\Delta n} = 0.19, 0.026$ and 0.0081 for $\Delta n = 1, 2$ and 3, respectively.

Emission measure is defined as the integral of the electron density squared along the line of sight, given by

$$\mathrm{EM} = \int n_e^2(l) dl \tag{1.3}$$

where n_e is the electron density and l is the line of sight distance, meaning EM is a measure of the total quantity of electrons producing photons within the line of sight. Electron temperature is a measure of the average kinetic energy of electrons, and is assumed by L. D. Anderson et al. [1] to be $\approx 8000\,\mathrm{K}$. Taken together, temperature and emission measure are the two physical parameters that determine the strength of RRL emission.

The two important features to note from equation (1.2) are, first, that alpha transitions are expected to be an order of magnitude stronger than beta transitions and two orders of magnitude stronger than gamma transitions, and, second, that line intensity is inversely proportional to frequency, meaning that lower-frequency lines are expected to be strongest.

While line emission from a source at rest should classically occur only at the specific frequency characterized by the Rydberg formula of equation (1.1), when measuring an emission line produced by an astronomical source,

the line will be broadened in its frequency profile to a shape approximately Gaussian characterized by its FWHM line width ΔV. Line broadening is caused by a combination of thermal broadening (the random motion of electrons), turbulence, the bulk motion of the gas, and pressure broadening (the perturbations of energy levels between closely packed neighboring particles).

The integrated line intensity—the line intensity integrated over the line width, which eliminates the dependence on ΔV—is further shown by L. D. Anderson et al. [1] to be

$$\frac{W_{\mathrm{RRL}}}{\mathrm{K\,km\,s^{-1}}} \backsimeq 3274 \left(\frac{T_e}{\mathrm{K}}\right)^{-3/2} \left(\frac{\mathrm{EM}}{\mathrm{pc\,cm^{-6}}}\right) \left(\frac{\nu_0}{\mathrm{GHz}}\right)^{-1} \Delta n M_{\Delta n}. \tag{1.4}$$

1.3 CHORD Telescope

CHORD is a radio interferometer being constructed at the Dominion Radio Astrophysical Observatory (DRAO) in British Columbia, Canada. The telescope consists of a 512-dish core array of 6-meter dishes operating from 300–1500 MHz. A 64-dish subset of the full array, called CHORD Pathfinder, is expected to provide early science and testing data by late 2026. The full array is scheduled to be commissioned in 2028.

CHORD is an interferometer, meaning it does not generate conventional images directly; instead, each pair of dishes measures a single component of the sky's 2-dimensional Fourier transform. Each Fourier component measures the intensity of a unique 2-dimensional sinusoid, with alternating sensitive and non-sensitive regions. Combining many Fourier components reconstructs a conventional image. CHORD consists of a highly redundant array, meaning multiple pairs of dishes measure the same Fourier components. This design reduces the total unique Fourier components, however, each component receives multiple independent samples, reducing measurement error and assisting with antenna calibration. As a result, CHORD is expected to achieve higher sensitivity compared to a non-redundant design.

As a drift-scan telescope, CHORD maintains a fixed declination pointing along the local meridian and measures continuously along the strip of right ascension (RA) that passes through its field of view as the Earth rotates. The full survey is then built up by re-pointing the dishes to different declinations. The full declination range accessible to CHORD is $+20\,^\circ < \delta < +80\,^\circ$,

which is planned to be fully covered over a 5-year survey [19]. The full possible CHORD survey extent is plotted in Figure 1.1.

The instrument will have a native spectral resolution of 195.3 kHz, with the option of upchannelizing via a two-stage polyphase filter bank (PFB) to higher resolution [2]. We henceforth assume a spectral resolution of 5 kHz, corresponding to a velocity resolution of $5\,\mathrm{km\,s^{-1}}$ at 300 MHz and $1\,\mathrm{km\,s^{-1}}$ at 1500 MHz. Beam characteristics of CHORD and Pathfinder are discussed in detail in section 3.2.

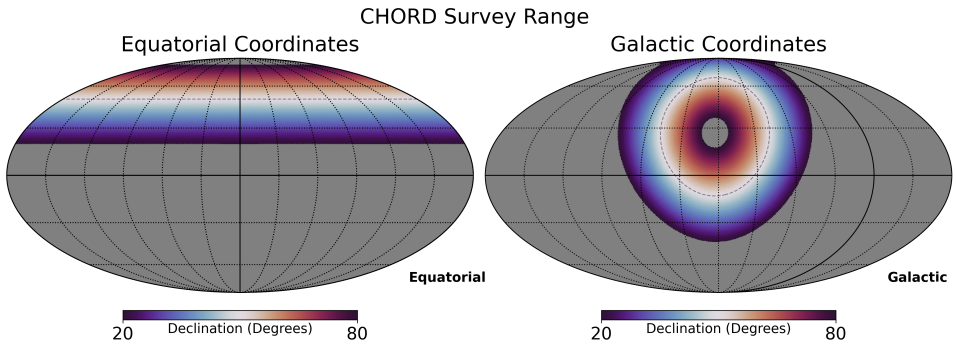


Figure 1.1: The full declination range accessible to CHORD, $+20\,^\circ < \delta < +80\,^\circ$. The left plot shows equatorial coordinates centered on the equator, the right plot shows Galactic coordinates centered on the Galactic plane.

Chapter 2

RFI Analysis

In section 2.1 we define radio frequency interference (RFI) and its impact on RRL science. In section 2.2 we describe the methods used to detect RFI, including excess kurtosis and spectral kurtosis. In section 2.3 we describe the DRAO RFI monitor and its data products. In section 2.4 we apply the RFI detection methods to the RFI monitor data and characterize the radio spectrum surrounding each target frequency.

2.1 RFI Impact

2.1.1 Definition of RFI

RFI consists of radio waves emitted from any Earthly source that can overcome or be confused with the target signal. Examples of RFI can include satellites, electronic equipment, and local broadband or narrowband communications. Unlike statistical noise, RFI cannot be removed through smoothing or time averaging because RFI is a distinct signal and not a random variable. Instead, frequency bins or time intervals corrupted by RFI must be identified and removed.

2.1.2 RFI Impact on RRL Science

RRL science is particularly sensitive to RFI for multiple reasons:

- RRLs are weak signals operating at or below the instrument's sensitivity level, so even low-level RFI is impactful.
- RRLs occur only at specific frequencies, so there is no option to shift observations to avoid RFI.
- RRL detection requires sensitive measurements across a wide frequency profile, so any RFI within the profile could corrupt the measurement.

With these considerations in mind, it is vital to understand the RFI environment surrounding each target frequency to inform RRL survey performance.

2.2 RFI Detection Methods

2.2.1 Excess Kurtosis for Non-Gaussianity

Kurtosis-based RFI detection originates from the principle that true astronomical signals, along with thermal noise, should appear Gaussian, while RFI should not. Intermittent RFI, produced from lightning strikes, engine ignition, and radar pulses, is leptokurtic, having more outliers than expected in Gaussian distributions. Continuous RFI, produced from communication signals with defined power envelopes, is platykurtic, having fewer outliers. Kurtosis is highly sensitive to the outliers of a distribution, which tests Gaussianity specifically in the area where we expect to find RFI. Alternative statistical tests, such as the Shapiro-Wilk test, measure Gaussianity across the full distribution. In an empirical study of eighteen normality tests, A. Jo et al. [8] found that the kurtosis test had the greatest statistical power[1] for symmetric data and the Shapiro–Wilk for asymmetric data; both ranked highly among the remaining sixteen tests overall. We expect leptokurtic RFI to be asymmetric, having a long tail of outliers, while platykurtic RFI should be symmetric. While we expect Shaprio-Wilk to have good statistical power on leptokurtic RFI, it is computationally expensive or impossible to implement without completely redesigning a radio receiver's streaming data pipeline. The Shapiro-Wilk test requires storing and sorting all accumulated samples per frequency channel; in comparison, the kurtosis test can be computed in a single streaming pass, which explains why kurtosis-based methods are preferred in astronomy [17]. While either test could be performed in this analysis, we use kurtosis rather than Shapiro–Wilk to simulate the compute constraints expected of the CHORD streaming pipeline.

Kurtosis is the standardized fourth central moment of a distribution and is a measure of how strong the tails of a distribution are. The k-th central moment of a distribution is defined by the expectation value $\mu_k \equiv \mathbb{E}[(X-\mu)^k]$. The first four central moments are given by:

$$\text{Mean:} \quad \mu = \mathbb{E}[X] = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2.1}$$

$$\text{Variance:} \quad \mu_2 = \mathbb{E}[(X-\mu)^2] = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 \equiv \sigma^2 \tag{2.2}$$

[1] Here, statistical power denotes the probability of correctly rejecting the null hypothesis of Gaussianity when the data are in fact non-Gaussian.

$$\text{Skewness:} \quad \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mathbb{E}[(X-\mu)^3]}{\mathbb{E}[(X-\mu)^2]^{3/2}} \qquad \text{(Standardized)} \tag{2.3}$$

$$\text{Kurtosis:} \quad \frac{\mu_4}{\mu_2^2} = \frac{\mathbb{E}[(X-\mu)^4]}{\mathbb{E}[(X-\mu)^2]^2}. \qquad \text{(Standardized)} \tag{2.4}$$

Skewness and kurtosis are standardized by dividing by the variance raised to the power of $\frac{3}{2}$ and 2, respectively, making them independent of the scale of the data.
Defining the residuals $r \equiv X - \mu$, the standardized kurtosis of a Gaussian is

$$\mathbb{E}[r^4] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} r^4 e^{-\frac{r^2}{2\sigma^2}} dr = 3\sigma^4 \tag{2.5}$$

$$\mathbb{E}[r^2] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} r^2 e^{-\frac{r^2}{2\sigma^2}} dr = \sigma^2 \tag{2.6}$$

$$\text{Kurtosis} = \frac{\mathbb{E}[r^4]}{\mathbb{E}[r^2]^2} = \frac{3\sigma^4}{\sigma^4} = 3. \tag{2.7}$$

From equation 2.7 we can see that a Gaussian distribution is expected to have a kurtosis of 3, regardless of the scale of the underlying data. Excess kurtosis is defined as the kurtosis minus 3, representing the amount of kurtosis above or below the expected value for a Gaussian distribution:

$$\text{EK} = \frac{\mu_4}{\mu_2^2} - 3. \tag{2.8}$$

A positive excess kurtosis indicates more samples in the tails, while a negative excess kurtosis indicates fewer samples in the tails as compared to a Gaussian distribution. Excess kurtosis can therefore be used in RFI detection to identify a spectrum that deviates from Gaussian.

Equation 2.8 is the population excess kurtosis, for a finite sample size M the unbiased sample excess kurtosis is given by M. G. M. G. Kendall [9] as

$$\widehat{EK}_s = \frac{M-1}{(M-2)(M-3)} \left((M+1)\frac{\mu_4}{\mu_2^2} - 3(M-1) \right). \tag{2.9}$$

2.2.2 Spectral Kurtosis for Power Spectra

Spectral kurtosis was developed by G. M. Nita & D. E. Gary [13] for RFI detection in single-look power spectra. The theory of spectral kurtosis

is that a power spectral density (PSD) per unit frequency, hereafter referred to as power (P_k) for frequency bin k, is proportional to voltage squared ($P_k \propto |V_k|^2$) and follows an exponential distribution. Spectral kurtosis acts as a test on the power spectrum to determine whether the underlying voltage distribution is Gaussian or not. Spectral kurtosis is defined as:

$$\widehat{SK} = \frac{\widehat{\sigma}_k^2}{\widehat{\mu}_k^2}, \tag{2.10}$$

where $\hat{\sigma}_k^2$ is the sample variance and $\hat{\mu}_k$ is the sample mean of power in frequency bin k over M independent samples, given by

$$\hat{\mu}_k = \frac{1}{M}\sum_{i=1}^{M} P_{k,i} \tag{2.11}$$

$$\hat{\sigma}_k^2 = \frac{1}{M-1}\sum_{i=1}^{M}(P_{k,i} - \hat{\mu}_k)^2. \tag{2.12}$$

Power is the squared magnitude of voltage, so the spectral kurtosis applied to a power measurement is conceptually similar to measuring kurtosis of the voltage, but not identical to it. G. M. Nita & D. E. Gary [13] show that an unbiased estimate of spectral kurtosis can be computed using the simplified formula

$$\text{let:} \quad S_1 = \sum_{i=1}^{M} P_{k,i}, \quad S_2 = \sum_{i=1}^{M} P_{k,i}^2 \tag{2.13}$$

$$\widehat{SK} = \frac{M+1}{M-1}\left(\frac{MS_2}{S_1^2} - 1\right). \tag{2.14}$$

For voltages drawn from a Gaussian distribution, the spectral kurtosis should be exactly 1.

Most systems do not record single-look power and instead record integrated power. To accommodate this G. M. Nita & D. E. Gary [14] developed a generalized spectral kurtosis estimator which acts on integrated power. The generalized spectral kurtosis estimator is given by

$$\text{let:} \quad S_1 = \sum_{i=1}^{M}\sum_{j=1}^{N} P_{k,i,j}, \quad S_2 = \sum_{i=1}^{M}\left(\sum_{j=1}^{N} P_{k,i,j}\right)^2 \tag{2.15}$$

$$\widehat{SK} = \frac{NMd+1}{M-1}\left(\frac{MS_2}{S_1^2} - 1\right), \tag{2.16}$$

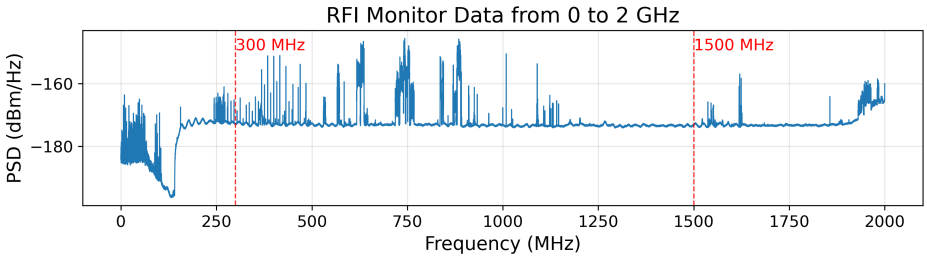


Figure 2.1: RFI monitor data from 0 to 2 GHz for a single time sample recorded on May 5, 2026 at 6:00 PM PDT.

where N is the number of samples per integration, and d is the shape parameter of the underlying distribution, which for an exponential distribution is 1.

2.3 DRAO RFI Monitor

2.3.1 RFI Monitor Specifications

A roof-mounted receiver located at the DRAO site, called the "RFI monitor," records radio signals from all directions, providing a means of estimating the RFI that CHORD is expected to receive [4]. The monitor records complex voltages at a rate of 4 gigasamples per second (GSPS). The stream is processed by an oversampled polyphase filterbank (OSPFB) with a 16/15 oversampling ratio yielding 16 frequency bands each with 125 MHz bandwidth and 133 MHz sampling rate. Each band is sent through a critically sampled filterbank, which performs a Fourier transform, and converted from voltage to power. Power is integrated over 2500 time samples to obtain integrations lasting 0.75 seconds. The final data product contains 600,000 channels at 3.33 kHz resolution from 0 to 2 GHz. The data are stored in PSD units of dBm/Hz. This analysis uses one hour of RFI monitor data collected at 6:00 PM PDT on May 5, 2026 comprising 4800 time samples at 0.75 second resolution, a sample of which is plotted in figure 2.1.

2.3.2 RFI Monitor Data Statistics

To understand the statistics of kurtosis and spectral kurtosis we trace how statistical noise propagates through the RFI monitor. An analog voltage measures incoming signals, which are digitized to real and imaginary

components V_R and V_I, respectively, given by

$$V = V_R + jV_I. \tag{2.17}$$

In both statistical noise and astronomical signals, the real and imaginary components are independent and identically distributed Gaussian random variables with identical variance σ^2. The mean of each component is unknown and non-zero. However, the mean does not impact the distribution type or its kurtosis and is therefore ignored. Next, power is computed as the magnitude of voltage squared, given by

$$P = |V|^2 = V_R^2 + V_I^2. \tag{2.18}$$

The sum of the squares of two independent Gaussian random variables follows a chi-squared distribution with two degrees of freedom, which is equivalent to an exponential distribution with mean $2\sigma^2$, meaning that power at this stage is exponentially distributed.

$$P \sim \sigma^2 \chi_2^2 \tag{2.19}$$

$$P \sim \exp(2\sigma^2) \tag{2.20}$$

The stream is then integrated over N samples, which forms a gamma distribution with shape parameter N and scale parameter $2\sigma^2$.

$$P_N = \sum_{i=1}^{N} P_i \tag{2.21}$$

$$P_N \sim \text{Gamma}(N, 2\sigma^2) \tag{2.22}$$

Finally the integrated power is converted to power in decibel units of dBm/Hz by scaling power with a gain calibration factor (G_{cc}) and taking the log,

$$P_{\text{dB}} = 10 \log_{10}(G_{\text{cc}} \cdot P_N). \tag{2.23}$$

To perform analysis we need to recover the integrated power P_N from the data product P_{dB} which is done by exponentiating

$$P_N = \frac{10^{P_{\text{dB}}/10}}{G_{\text{cc}}}. \tag{2.24}$$

The gain factor G_{cc} is unknown and changes between each calibration cycle, which occurs every 5 minutes as shown in the top panel of figure 2.2.

An analysis of power at 1420 MHz, which is an RFI protected frequency, shows a statistically significant change in power between each calibration cycle (bottom panel figure 2.2), which can only be attributed to the gain factor as there is not expected to be any time-dependent signal change at 1420 MHz. Therefore, computing kurtosis of P_N between calibration cycles will be overestimated as some of the variance would be caused by the changing gain factor and not the intrinsic variance of P_N. We split the time range into 11 distinct calibration segments of 382 samples each, equating to 4.7 minutes per segment. We discard the first and last segments shown in figure 2.2 as they do not contain a complete set of 382 samples.

Within each calibration segment the gain factor remains constant and can be safely ignored, we compute spectral kurtosis of integrated power P_N using the generalized spectral kurtosis estimator, knowing that Gaussian noise should have a spectral kurtosis of 1. Additionally, under the central limit theorem, the sum of N independent, identically distributed, exponential random variables, converges to a Gaussian. In Appendix A.1 we show that a gamma distribution with shape N and scale $2\sigma^2$ converges to a Gaussian distribution with mean $N \cdot 2\sigma^2$ and variance $N \cdot (2\sigma^2)^2$. Since the number of integrated samples ($N = 2500$) is large, it is reasonable to apply the Gaussian approximation given by

$$P_N \approx \text{Gaussian}(N \cdot 2\sigma^2, N \cdot (2\sigma^2)^2). \tag{2.25}$$

By making the Gaussian approximation, we can compute excess kurtosis of the integrated power P_N and expect it to be 0 for a Gaussian power distribution.

2.4 Application of RFI Detection

In this section we apply excess kurtosis to test if the power distribution is Gaussian, and spectral kurtosis to test if the underlying voltage distribution is Gaussian.

2.4.1 Excess Kurtosis RFI Detection

To obtain the expected distribution of excess kurtosis under the null hypothesis of Gaussian power, we performed a Monte Carlo simulation. We drew 382 samples from a Gaussian distribution and computed excess kurtosis. We repeated this for 100,000 trials, giving the expected distribution of

[2] A violin plot is a box-plot with a rotated kernel density plot on each side.

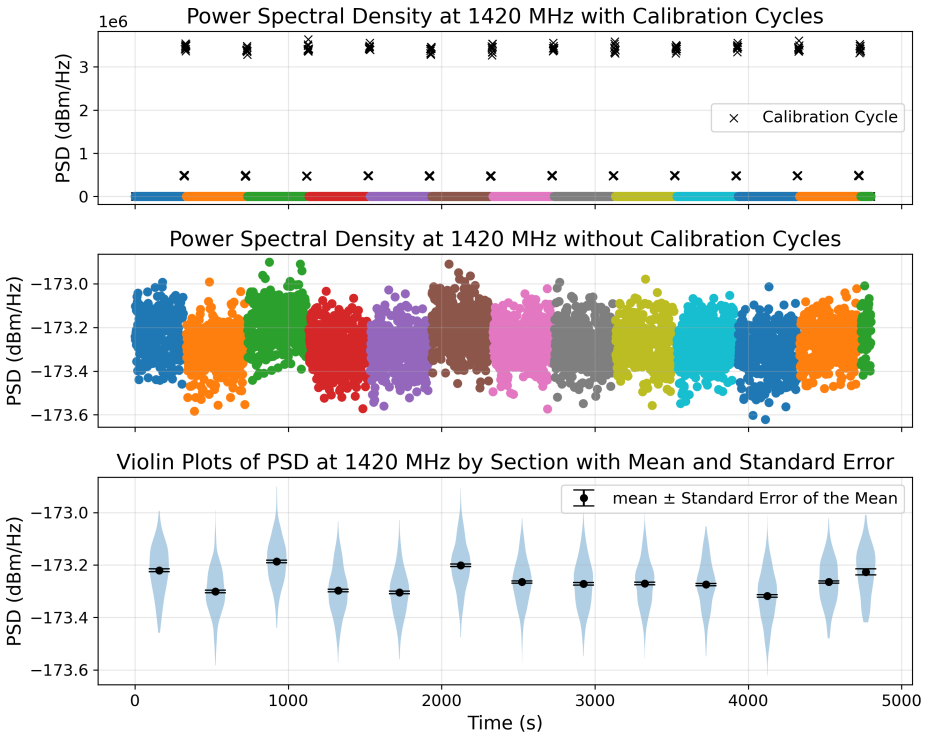


Figure 2.2: RFI monitor data at 1420 MHz. Top panel shows periodic calibration cycle, middle panel shows PSD between cycles, and bottom panel shows a violin plot[2] with mean and standard error of PSD within each cycle. The results show a statistically significant change in mean power between cycles caused by changes in the gain factor G_{cc}.

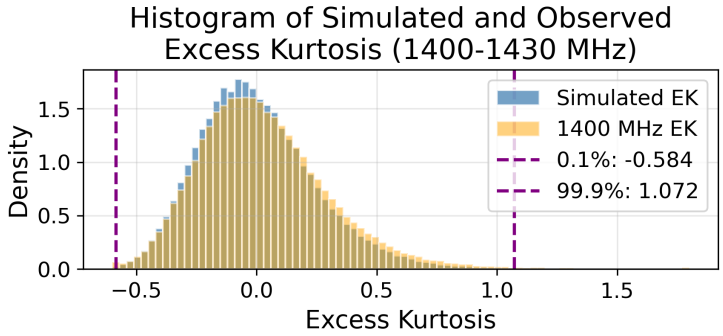


Figure 2.3: Histogram of excess kurtosis for 100,000 Monte Carlo simulations of Gaussian noise in blue, and histogram of observed excess kurtosis within the RFI-free 1400 MHz to 1430 MHz band in yellow. Showing observed excess kurtosis is slightly higher than predicted by theory. The 0.1% and 99.9% simulated thresholds are shown as vertical dashed lines.

excess kurtosis shown in Figure 2.3. We take the RFI detection thresholds as the 0.1st and 99.9th percentiles of the resulting distribution, giving -0.584 and 1.072 threshold bounds, respectively. The thresholds are asymmetric, which arises from the positive skew of excess kurtosis.

To validate the theoretical bounds, we selected a clean portion of the spectrum from 1400 MHz to 1430 MHz because it sits within an RFI-protected band. Figure 2.4 shows the PSD spectrum of the clean band alongside its excess kurtosis, confirming that there is no significant RFI. We aggregated the observed excess kurtosis of the clean band across each calibration cycle and plotted it alongside the simulated results in figure 2.3. The observed excess kurtosis closely matches theory which is expected for an RFI-free zone. The observations are slightly shifted right compared to simulated results which suggests more outliers in the true data than predicted by theory. A likely cause is low-level, intermittent, broad-spectrum RFI, which nevertheless impacts protected spectrum. We selected intentionally wide bounds of 0.1% and 99.9% to limit false-positives; these bounds classified 99.5% of the observed data between 1400 MHz and 1430 MHz as RFI-free, whereas theory predicts that 99.8% should be RFI-free.

We evaluated the excess-kurtosis RFI detection method by examining known RFI regions and observing the results. A plot of power and excess

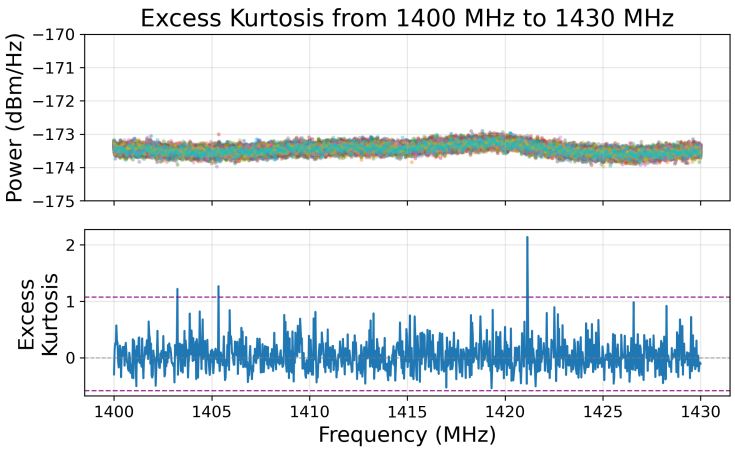


Figure 2.4: Top: PSD in dBm/Hz across the 1400 MHz to 1430 MHz band. Bottom: Excess kurtosis of P_N for each frequency bin with the 0.1% and 99.9% simulated thresholds shown as horizontal dashed lines, indicating that the band is largely free of RFI.

kurtosis from 590 MHz to 610 MHz is shown in Figure 2.5, demonstrating both clean regions (592.5 MHz to 595 MHz) and RFI-contaminated regions (597 MHz to 602 MHz). Excess kurtosis fails to capture these areas of known RFI, suggesting it is not a sensitive enough statistic to be used further.

2.4.2 Spectral Kurtosis RFI Detection

To apply the generalized spectral kurtosis estimator, we must specify the shape parameter d and integration count N. The theoretical value d of an exponential distribution is 1 [14], and the total integration count is $N = 2500$ [4]. However, in practice, not every integrated sample is fully independent; oversampling introduces duplicate samples, and windowing and filtering within the filterbank may induce correlation between adjacent samples. A standard practice employed by E. T. Smith et al. [16] and G. Nita et al. [12] is to empirically estimate the shape parameter d from the data to account for these instrumental effects.

We reduce the integrated sample count N from 2500 to $2500 \times 15/16$ to account for the 16/15 oversampling ratio. Next, we compute spectral kurtosis on integrated power P_N using the RFI-free band of 1400 MHz to

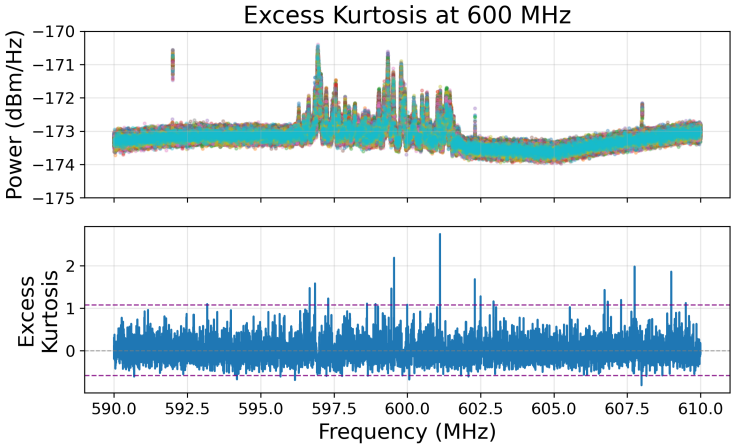


Figure 2.5: Power and excess kurtosis from 590 MHz to 610 MHz, showing that excess kurtosis is not able to clearly distinguish between clean and RFI-contaminated spectrum.

1430 MHz shown in figure 2.4. We compute the mode of spectral kurtosis across the clean spectrum to empirically estimate the calibrated shape parameter $d = 0.96$, such that the calibrated mode sits at exactly 1, as shown in figure 2.6.

We computed the RFI thresholds following the same approach as for excess kurtosis. We computed spectral kurtosis using 382 samples drawn from a gamma distribution with the calibrated shape parameter, and repeated the process for 100,000 Monte-Carlo simulations to obtain the 0.1% and 99.9% percentiles. We plotted the spectral kurtosis of the clean band from 1400 MHz to 1430 MHz as a histogram alongside the simulated results in figure 2.7. The observed spectral kurtosis matches the simulated data nearly perfectly, with 99.7% of observed data falling within the simulated 99.8% bounds.

To assess the quality of the spectral kurtosis RFI method, we examined spectrum containing known RFI, an example of which is shown in figure 2.8. Spectral kurtosis clearly distinguishes the clean spectrum (592.5 MHz to 595 MHz) from the RFI-contaminated spectrum (597 MHz to 602 MHz). From this and other comparisons we concluded that our approach to RFI detection using spectral kurtosis is effective and proceeded to use it for the

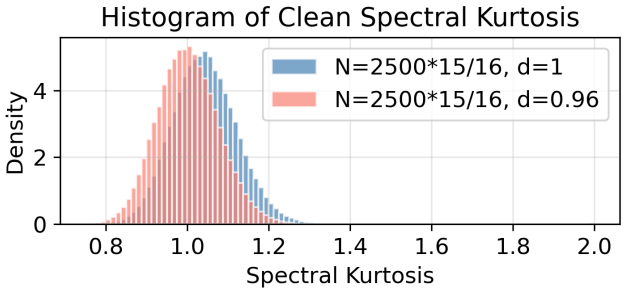


Figure 2.6: The blue histogram shows spectral kurtosis on RFI-free spectrum using the theoretical shape parameter $d = 1$ with a mode of 1.04. The salmon histogram shows the same spectrum using an empirically estimated shape parameter $d = 0.96$, which places the mode at 1.

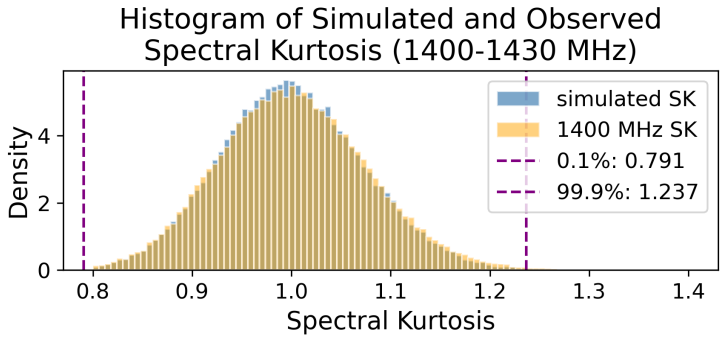


Figure 2.7: Histogram of spectral kurtosis for 100,000 Monte Carlo simulations of Gaussian noise in blue, and histogram of observed spectral kurtosis using the RFI-free spectrum from 1400 MHz to 1430 MHz across all time segments in yellow, showing near perfect agreement between the distributions. The 0.1% and 99.9% thresholds are shown as horizontal dashed lines.

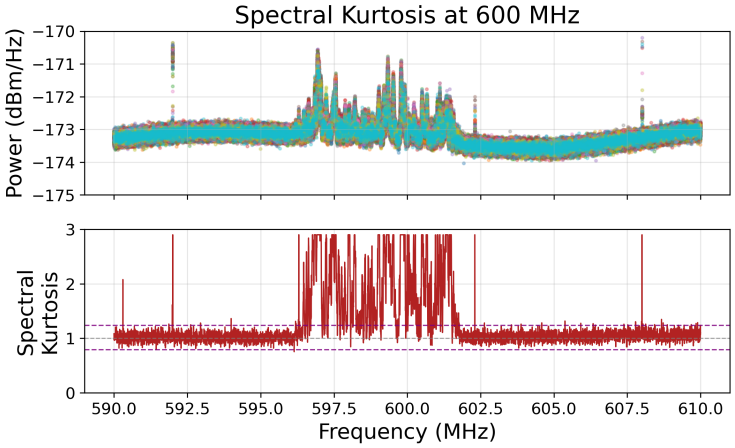


Figure 2.8: Power and spectral kurtosis from 590 MHz to 610 MHz, showing a clear distinction in spectral kurtosis between clean and RFI-contaminated regions.

remainder of the analysis.

When applied to the entire CHORD band, spectral kurtosis finds that 75.6% of the spectrum is RFI-free, while 24.4% is contaminated with RFI. A histogram of spectral kurtosis across the entire CHORD band is shown in figure 2.9.

2.4.3 Radio Recombination Line Visibility

Because each calibration segment must be treated independently, the spectral kurtosis results produce 11 distinct RFI labels for each frequency bin, one from each segment. Typically there is strong agreement between segments. On average any two segments provide the same RFI labels for 93.8% of frequency bins, with the lowest segment agreement being 91.7% and the highest being 95.9%. A full pairwise comparison is presented in Appendix A.2. Disagreement between segments is likely driven in part by random noise fluctuations and in part by intermittent RFI.

A given frequency bin is observable when it is free of RFI at least 50% of the time. CHORD will remain at the same pointing for long periods (weeks to months), which allows for sufficient observation time in the periods when the RFI is not present. We therefore chose to label a frequency bin as clean

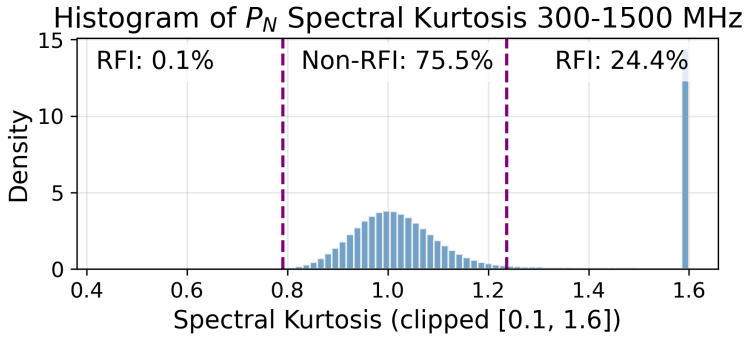


Figure 2.9: Histogram of spectral kurtosis across the full 300 MHz to 1500 MHz CHORD band for all calibration segments, which finds that 75.6% of the sampled frequency bins are RFI-free. Spectral kurtosis above 1.6 is aggregated at 1.6 for visibility.

if it is free of RFI in at least 50% of the calibration segments. A sensitivity analysis of the 50% threshold is presented in Appendix A.3 which shows the number of detectable RRL lines using different threshold assumptions. With the 50% threshold, we find that 76.3% of the spectrum is RFI-free, which is consistent with the overall estimate in figure 2.9 of 75.6%. The full CHORD spectrum, with RFI labels and the locations of hydrogen RRLs, is shown in figure 2.10.

For an RRL to be detectable, it should be measurable across a width of at least 100 km/s, and ideally up to 200 km/s. To characterize this full width rather than the line centre alone, we compute a Gaussian-weighted average of the binary RFI labels, centred on each line's rest frequency f_0 with a standard deviation of 100 km/s. We convert velocity widths to frequency using $\Delta f = f_0 \Delta v/c$, resulting in a wider Gaussian at higher frequencies (100 km/s equates to 30 bins at 300 MHz and 148 bins at 1500 MHz). This approach places 68% of the weighted importance within ±100 km/s of the centre and 95% within ±200 km/s. The resulting detectability score is

$$s(f_0) = \frac{\sum_i w_i \,\mathrm{RFI}_i}{\sum_i w_i}, \qquad w_i = \exp\left(-\frac{(f_i - f_0)^2}{2\sigma_f^2}\right), \tag{2.26}$$

where f_i is the frequency of bin i and $\mathrm{RFI}_i \in \{0, 1\}$ is its RFI label. The score is bounded by $[0, 1]$, where 1 indicates spectrum that is completely clean and

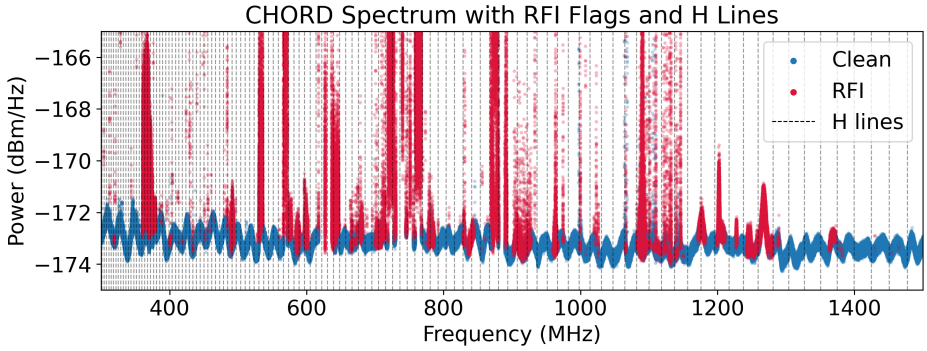


Figure 2.10: Sample RFI monitor power spectrum coloured by RFI label, the locations of hydrogen RRLs are shown as vertical dashed lines. Some spectrum with intermittent RFI is still considered clean if it is RFI-free in at least 50% of calibration segments.

Table 2.1: Table of detectable hydrogen RRLs versus RFI-contaminated RRLs by frequency range.

Frequency Range (MHz)	**Detectable Transitions**	**Contaminated Transitions**
300–470	29	10
470–800	23	16
800–1500	27	11
Total	**79**	**37**

0 that it is completely RFI. A plot of detectability scores versus frequency for all hydrogen RRLs is shown in figure 2.11. Using a threshold that a line must have a detectability score of at least 99%, we find that 79 of the 116 hydrogen RRLs in the CHORD band are expected to be detectable, and 37 are not. The number of detectable transitions per frequency range is presented in table 2.1. If we instead consider a frequency bin clean when a minimum of one segment is RFI-free, the total number of detectable lines becomes 91. A sensitivity analysis of spectral kurtosis thresholds and combination methods is presented in Appendix A.3.

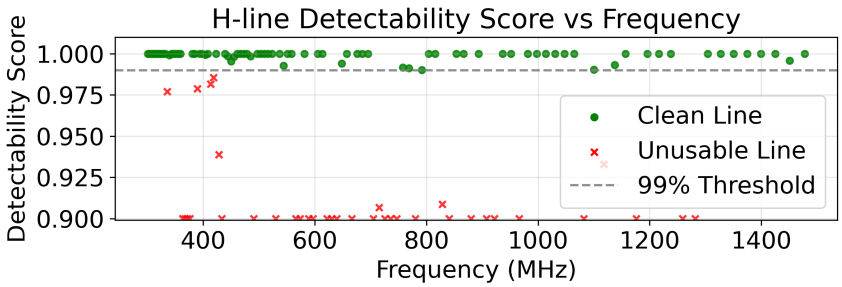


Figure 2.11: Detectability score of each hydrogen RRL in the CHORD range using a 99% detectability score cutoff to classify each line as detectable or contaminated.

2.4.4 Analysis Limitations

There are several important limitations to consider about conclusions from the RFI monitor:

- **Sensitivity mismatch:** CHORD is orders of magnitude more sensitive than the RFI monitor; signals not detectable by the RFI monitor may nonetheless impede CHORD.
- **Beam geometry:** CHORD's beam is directional and points upwards, whereas the RFI monitor's beam is omnidirectional and points horizontally and downward. As a result, the RFI monitor may not detect RFI signals coming from above that will impact CHORD. In particular, the GNSS satellite communication band from 1164–1215 MHz is not detected by the RFI monitor.
- **Limited sampling window:** This analysis is based on a single hour of RFI monitor data, which may not be representative of the long-term RFI environment.
- **Temporal evolution:** RFI is not static: as new RFI sources emerge, spectrum that is currently free of RFI might become corrupted.
- **Gaussian RFI:** Some RFI, such as spread-spectrum RFI, appears Gaussian and could be missed by this analysis.
- **Scope:** This analysis only considers the impact of RFI on detectability; it does not address how line broadening or the superposition of

hydrogen β, hydrogen γ, helium, and carbon recombination lines might also impact detectability.

Chapter 3

CHORD Sensitivity Analysis

We derive the instrumental point source sensitivity of CHORD in section 3.1, the angular resolution in section 3.2, and the surface brightness sensitivity in section 3.3. We then incorporate radio background to obtain the overall sensitivity across the full CHORD field of view in section 3.4.

3.1 Point Source Sensitivity

The point source sensitivity of a radio interferometer is the minimum flux density of a point source that can be detected at a given signal-to-noise ratio (SNR). The standard approach is to choose an SNR of 1, in which case the sensitivity is equal to the root-mean-square (RMS) noise. We will use the terms sensitivity and RMS noise interchangeably. The RMS noise can be theoretically computed using the radiometer equation [6]

$$\sigma_{\text{RMS, single pol}} = \frac{2kT_{\text{sys}}}{A_e\sqrt{N(N-1)\,\Delta\nu\,\tau_{\text{eff}}}}, \tag{3.1}$$

where $k = 1.38 \times 10^{-23}\,\text{J}\,\text{K}^{-1}$ is the Boltzmann constant, $T_{\text{sys}} = 30\,\text{K}$ is the expected system temperature of each antenna, and $A_e = \pi(D/2)^2 \cdot 0.5 = 14.14\,\text{m}^2$ is the effective collecting area assuming a 50% illumination efficiency [18]. The remaining quantities are the bandwidth $\Delta\nu$ in Hz, the number of antennas N and the effective integration time τ_{eff} in seconds, defined below.

The physical antenna parameters of T_s and A_e are typically combined into a single quantity called the system equivalent flux density (SEFD), defined as

$$\text{SEFD} = \frac{2kT_{\text{sys}}}{A_e} \tag{3.2}$$

$$= 5868\,\text{Jy}. \tag{3.3}$$

The SEFD, in units of janskys, represents the flux density of a point source that would produce the same measured power as the system noise.

While equation 3.1 provides RMS noise for a single polarization, CHORD will measure dual polarizations. The total intensity, denoted as a Stokes I measurement, combines both orthogonal polarizations. Since each polarization is independent, the RMS noise is reduced by a factor of $\sqrt{2}$ due to the averaging of two independent noise contributions. Thus, the one-sigma sensitivity for Stokes I can be expressed as

$$\sigma_{\mathrm{RMS}} = \frac{\mathrm{SEFD}}{\sqrt{2\,\Delta\nu\,\tau_{\mathrm{eff}}\,N(N-1)}}. \tag{3.4}$$

Equation 3.1 assumes that the source is directly in the centre of the primary beam, which is the power response pattern produced by the dish geometry. However, CHORD is a drift scan telescope, meaning that instead of tracking a source of interest, it maintains a fixed pointing and sources move through the beam as the sky rotates overhead. We define the effective integration time τ_{eff} as the equivalent time that a source would need to spend in the centre of the primary beam to accumulate the same total power as it does given its drift path.

The CHORD primary beam is an Airy disk, whose response pattern is given by

$$A(\rho) = \left[\frac{2J_1\left(\frac{\pi D\nu}{c}\sin\rho\right)}{\frac{\pi D\nu}{c}\sin\rho}\right]^2, \tag{3.5}$$

where J_1 is the first-order Bessel function of the first kind, D is the dish diameter, ν is the observing frequency, and ρ is the angle from the beam centre. The full width at half maximum (FWHM) of the Airy disk is given by

$$A\left(\frac{\rho_{\mathrm{FWHM}}}{2}\right) = \frac{1}{2} \tag{3.6}$$

$$\Rightarrow \rho_{\mathrm{FWHM}} \approx 1.029\frac{\lambda}{D}. \tag{3.7}$$

For simplicity we will model the primary beam as a Gaussian with a matched FWHM and ignore the sidelobes. The Gaussian beam response is given by

$$A(\rho) = \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \tag{3.8}$$

$$= \exp\left[-4\ln 2\ \left(\frac{\rho}{\rho_{\,\mathrm{FWHM}}}\right)^2\right] \tag{3.9}$$

where $\rho_{\mathrm{FWHM}} = 2\sqrt{2\ln 2}\,\sigma$ is the primary beam FWHM. A plot of an Airy disk overlaid with a Gaussian beam is shown in figure 3.1 and shows that the Gaussian is a good approximation to the Airy disk when considering only the main lobe.

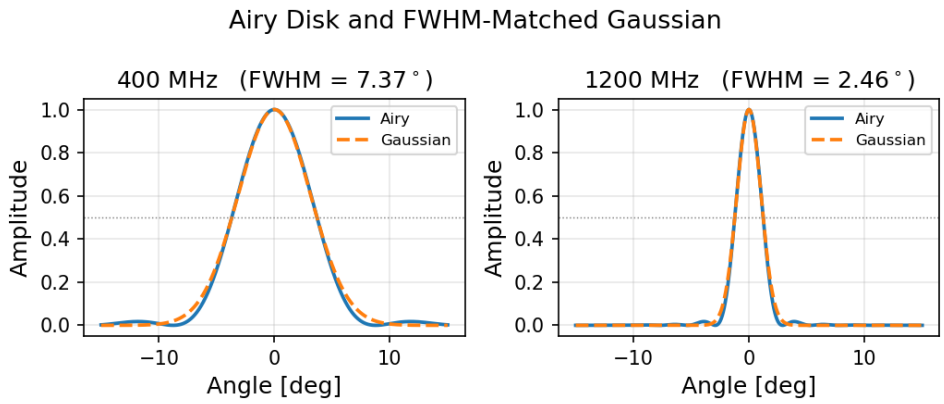


Figure 3.1: Comparison of the Airy disk and FWHM-matched Gaussian beam for observing frequencies of 400 and 1200 MHz, demonstrating that a Gaussian is a good approximation to the Airy disk.

The radial distance ρ can be expressed in terms of the azimuth and zenith angles as

$$\rho(t) = \sqrt{\theta(t)^2 + \phi^2}, \tag{3.10}$$

where $\theta(t) = \theta^*(t) - \theta_0$ and $\phi = \phi^* - \phi_0$ are the offset angles in azimuth (EW) and zenith (NS) between the beam centre (θ_0, ϕ_0) and a given sky location $(\theta^*(t), \phi^*)$. We can further express the azimuth angle in terms of Earth's declination-specific apparent angular velocity (ω) as $\theta^*(t) = t \cdot \omega \cos\phi_0$, and assume for simplicity that $\theta_0 = 0\,^\circ$ (that the azimuth angle is zero at time $t = 0$). The radial distance from the beam centre in terms of time and zenith offset is then

$$\rho(t) = \sqrt{t^2\omega^2 \cos^2\phi_0 + \phi^2}. \tag{3.11}$$

Making the flat-sky approximation, we will ignore curvature over the small angle of the primary beam, and treat θ, ϕ as Cartesian coordinates. The primary beam power response for sources with a zenith angle offset ϕ as a function of time is obtained by combining equations 3.9 and 3.11 to yield

$$A(t, \phi) = \exp\left[-4\ln 2\,\frac{t^2\omega^2\cos^2\phi_0 + \phi^2}{\rho^2_{\mathrm{FWHM}}}\right]. \tag{3.12}$$

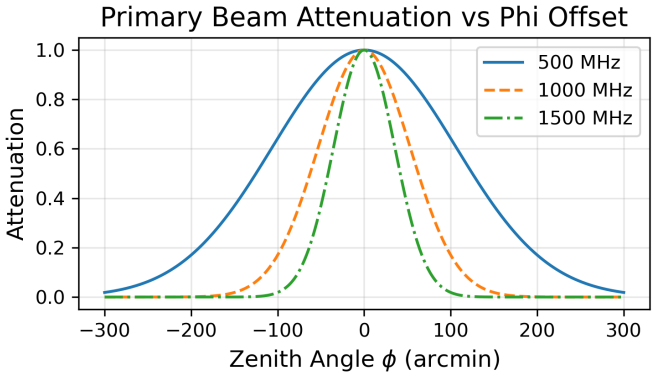


Figure 3.2: CHORD primary beam attenuation factor $D(\phi)$ versus zenith angle ϕ for different frequencies, showing that lower frequencies have wider beams and thus less attenuation for sources outside the beam path.

The total power accumulated from a point source per day of drift scan is obtained by integrating equation 3.12. During mapmaking, a matched filter reconstructs a sky image by reweighting visibilities relative to the image centre [7]. The reweighting is done using the primary beam response. As a result, the effective power accumulation is the integral of the square of the power response and not the total power accumulation. Taking $t = 0$ to be when the source is directly over the beam with $\theta(0) = 0$, the effective integration time, τ_{eff}, which is the equivalent time that a source would spend in the centre of the beam to accumulate the same effective power, is given by:

$$\tau_{\text{eff}} = \int_{-\infty}^{\infty} A^2(t)dt \tag{3.13}$$

$$= \int_{-\infty}^{\infty} \exp\left[-8\ln(2)\,\frac{t^2\omega^2\cos^2\phi_0 + \phi^2}{\rho^2_{\text{FWHM}}}\right] dt \tag{3.14}$$

$$= \exp\left(-8\ln(2)\frac{\phi^2}{\rho^2_{\text{FWHM}}}\right)\int_{-\infty}^{\infty} \exp\left[-8\ln(2)\,\frac{t^2\omega^2\cos^2\phi_0}{\rho^2_{\text{FWHM}}}\right] dt \tag{3.15}$$

$$= D(\phi)\sqrt{\frac{\pi}{8\ln(2)}}\frac{\rho_{\text{FWHM}}}{\omega\cos\phi_0}. \tag{3.16}$$

The factor $D(\phi) = \exp\left(-8\ln(2)\frac{\phi^2}{\rho^2_{\text{FWHM}}}\right)$ is a term in the range $]0, 1]$ accounting for the attenuation of sources outside the beam path.

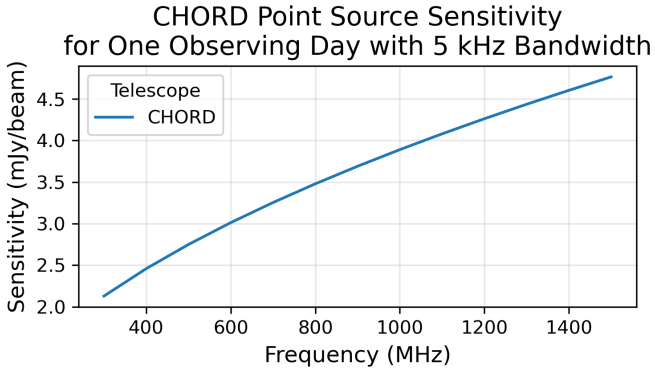


Figure 3.3: CHORD point source sensitivity versus observing frequency for one observing day with 5 kHz bandwidth. Sensitivity is improved at lower frequencies due to the wider beam and longer effective integration time.

The total RMS noise is obtained by combining equations 3.4 and 3.16 which yields:

$$\sigma_{\mathrm{RMS}} = \frac{\mathrm{SEFD}}{\sqrt{2\ \Delta\nu\, N(N-1)}} \sqrt{\frac{2\sqrt{2\ln(2)}\,\omega\ \cos\phi_0}{\sqrt{\pi}\,\rho_{\mathrm{FWHM}}\, D(\phi)}} \tag{3.17}$$

This can be simplified and expressed in radians as:

$$\sigma_{\mathrm{RMS}}\left[\mathrm{Jy\,beam^{-1}}\right] = \frac{\mathrm{SEFD}}{143} \sqrt{\frac{\cos\phi_0}{\Delta\nu[\mathrm{Hz}]\, N(N-1)\, \rho_{\mathrm{FWHM}}[\mathrm{rad}]\, D(\phi)}} \tag{3.18}$$

3.2 Angular Resolution

The angular resolution is the smallest angular scale that can be resolved by the telescope. Angular resolution is derived from the synthesized beam, which is the combined sensitivity profile after correlating all antennas. The standard angular resolution metric is the synthesized beam FWHM, which is given by

$$\theta_{\mathrm{FWHM}} = \frac{\lambda}{b_{\mathrm{EW}}}, \quad \phi_{\mathrm{FWHM}} = \frac{\lambda}{b_{\mathrm{NS}}}, \tag{3.19}$$

where b_{EW} and b_{NS} are the distances of the longest baselines in the east-west and north-south directions, respectively, and λ is the observing wavelength

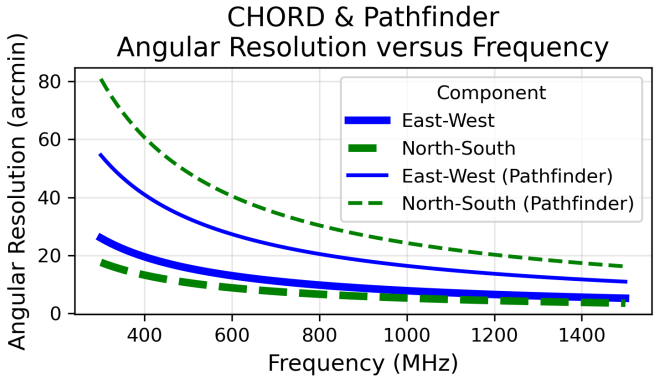


Figure 3.4: CHORD and Pathfinder angular resolution as a function of observing frequency.

that can be converted to frequency by $\nu = c/\lambda$. We plot the angular resolution versus frequency for both CHORD and Pathfinder in figure 3.4 showing that angular resolution improves at higher frequencies due to the shorter wavelength.

We approximate the synthesized beam as a 2-dimensional Gaussian with FWHM equal to the angular resolution in the azimuth (θ_{FWHM}) and zenith (ϕ_{FWHM}) directions. The synthesized beam's Gaussian profile is given by

$$A(\theta,\phi) = \exp\left[-4\ln 2\,\left(\frac{\theta}{\theta_{\mathrm{FWHM}}}\right)^2\right]\exp\left[-4\ln 2\,\left(\frac{\phi}{\phi_{\mathrm{FWHM}}}\right)^2\right]. \quad (3.20)$$

The synthesized beam solid angle Ω_s is the area subtended by the beam. In spherical coordinates, the differential solid angle is $d\Omega_s = \sin\theta\, d\theta\, d\phi$. For integrations over small solid angles, we can ignore the sky curvature and treat θ and ϕ as Cartesian coordinates (the synthesized beam ranges from $\sim 1^\circ$ to $\sim 0.1^\circ$ making the small angle approximation valid). The differential solid angle without curvature is $d\Omega_s \approx d\theta\, d\phi$. The synthesized beam solid angle is obtained by integrating the Gaussian beam over all angles, given by

$$\Omega_s \approx \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(\theta,\phi)\, d\theta d\phi \quad (3.21)$$

$$\approx \int_{-\infty}^{\infty}\exp\left[-4\ln 2\,\left(\frac{\theta}{\theta_{\mathrm{FWHM}}}\right)^2\right] d\theta \int_{-\infty}^{\infty}\exp\left[-4\ln 2\,\left(\frac{\phi}{\phi_{\mathrm{FWHM}}}\right)^2\right] d\phi \quad (3.22)$$

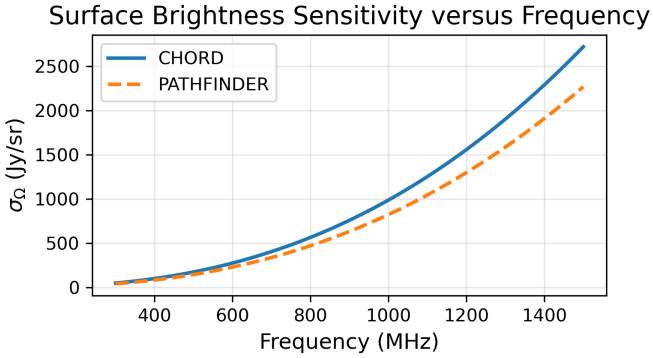


Figure 3.5: CHORD and Pathfinder surface brightness sensitivity (σ_Ω) using a 5 kHz bandwidth and one observation day. Sensitivity is worse at higher frequencies due to the smaller beam size, which reduces the effective integration time and synthesized beam solid angle.

$$\approx \frac{\pi\,\theta_{\mathrm{FWHM}}\,\phi_{\mathrm{FWHM}}}{4\ln 2}. \tag{3.23}$$

3.3 Brightness Sensitivity

Surface brightness sensitivity, which is a measure of the telescope's ability to detect extended sources, is obtained by dividing the point source sensitivity by the synthesized beam solid angle

$$\sigma_\Omega[\mathrm{Jy/sr}] = \frac{\sigma_{\mathrm{RMS}}}{\Omega_s}. \tag{3.24}$$

It is then converted to a brightness temperature by applying the Rayleigh-Jeans limit

$$\sigma_T[\mathrm{K}] = \frac{\sigma_\Omega c^2}{2k\nu^2}. \tag{3.25}$$

Figures 3.4 and 3.6 show the inverse relationship between angular resolution and surface temperature sensitivity across the 300–1500 MHz band, representing the trade-off between beam size and integration time. Sensitivity at higher frequency can be improved at the cost of angular resolution by convolving the final image to a larger beam, however low-frequency angular resolution cannot be improved without increasing the maximum baseline length. Sensitivity is also improved at higher declinations due to the longer effective integration time, as sources move slower across the sky and spend more time within the primary beam. The sensitivity maps for a single observing day are shown in figure 3.7.

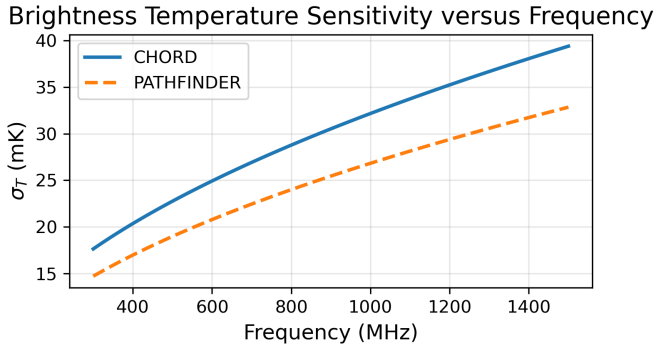


Figure 3.6: CHORD and Pathfinder brightness temperature sensitivity (σ_T) using a 5 kHz bandwidth and one observation day. Sensitivity is still worse at higher frequencies but the trend is less pronounced compared to surface brightness sensitivity.

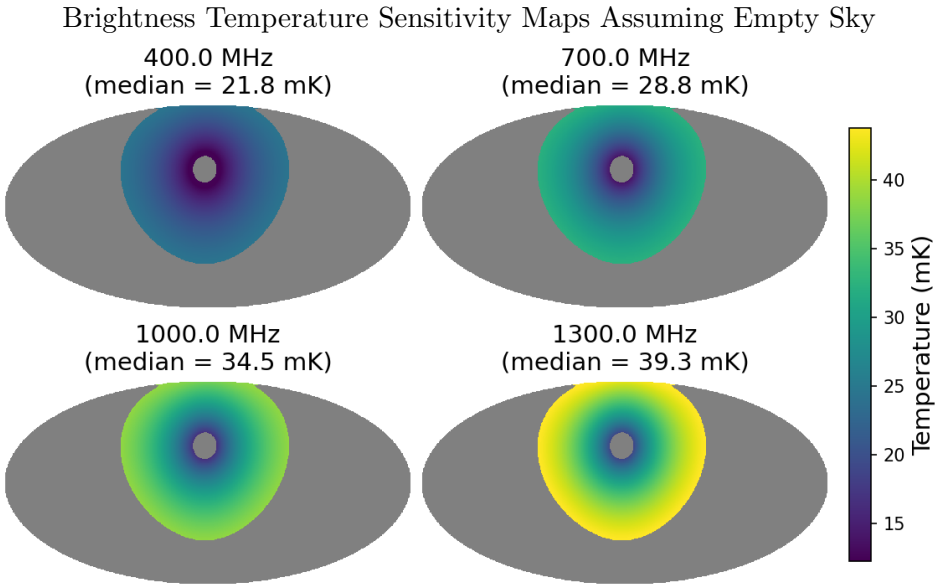


Figure 3.7: Brightness temperature sensitivity (σ_T) in the CHORD declination range for different frequencies assuming empty sky. Sensitivity improves at high declination due to the longer effective integration time, and at lower frequencies due to the wider primary beam.

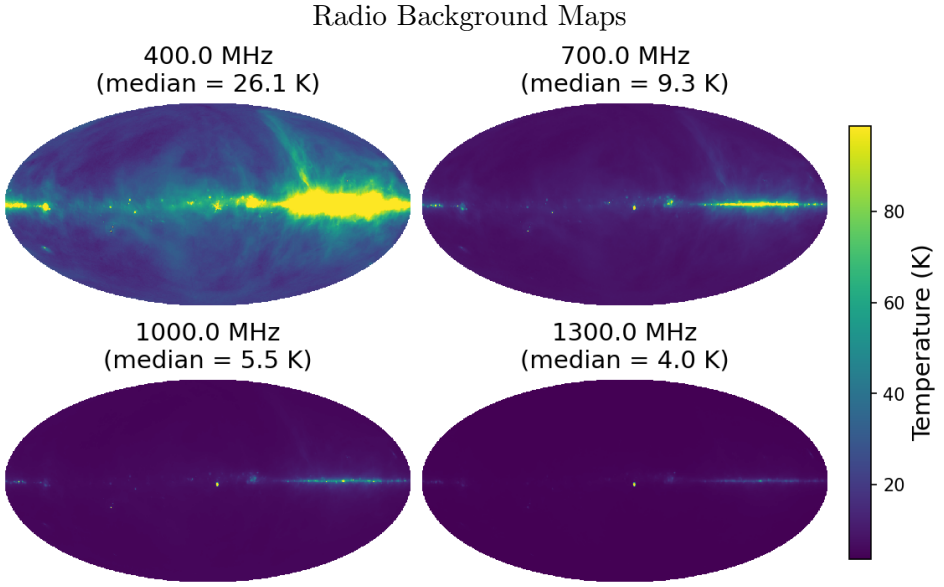


Figure 3.8: Radio background maps in Galactic coordinates centred on the CHORD field of view, measured in brightness temperature (K) for four different frequencies. Background is highest at lower frequency and along the Galactic plane, particularly towards the Galactic centre.

3.4 Sensitivity with Radio Background

Radio background contributes unwanted noise, which degrades sensitivity. Radio background is received by the antenna and adds linearly to the total antenna temperature. Therefore, RMS noise from background is calculated in the same method as equation (3.18). In other words, SEFD for combined background and internal RMS noise is taken to be the sum of antenna temperature and the background temperature,

$$\mathrm{SEFD} = \frac{2k(T_{\mathrm{ant}} + T_{\mathrm{sky}})}{A_e}. \tag{3.26}$$

The radio background is obtained from the Global Sky Model [20]. This model incorporates all dominant sources of radio emission into a single all-sky map of brightness temperature from 10 MHz to 5 THz. The model is accessible through the `pygdsm` package [15]. Plots are created using Mollweide projections in Galactic coordinates using the `healpy` package [21]. The full-sky background maps are shown in figure 3.8.

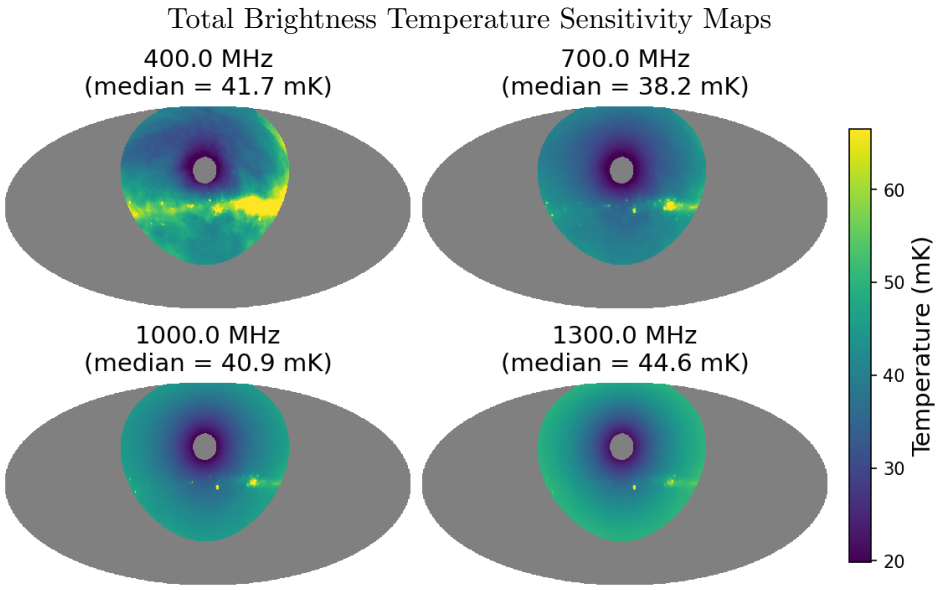


Figure 3.9: Total CHORD brightness temperature sensitivity maps measured in millikelvins (mK) for four different frequencies, showing the spatial and frequency dependence of sensitivity.

Maps of total sensitivity, which incorporates both system RMS noise and radio background RMS noise, are shown in figure 3.9. The total sensitivity has both spatial and frequency dependence, with more sensitivity at high declination and less along the midplane.

Chapter 4

Radio Line Stacking

The principle of line stacking is that each RRL is a measurement of the same underlying physical quantities, namely the emission measure and electron temperature of the ionized gas. We can therefore treat each RRL measurement as an independent observation and reduce variance by taking the average[3]. We model each observation x_i as an independent draw from a normal distribution, $x_i \sim \mathcal{N}(T_L, \sigma_i^2)$, where T_L is the line strength and σ_i is the brightness temperature sensitivity[4] of line i. In the ideal case, for N independent samples with mean μ and common variance σ^2, the variance of the mean is σ^2/N. For RRLs, however, the RMS noise varies with frequency across the lines, so the optimal estimator is not a simple average but a weighted average as we discuss in section 4.1. In section 4.2, we provide optimal stacking strategies by Galactic-latitude band.

4.1 Inverse-Variance-Weighted Average

The inverse-variance-weighted average minimizes the variance of the mean when sample variances are not constant. The weighted mean $\hat{\mu}$ and variance of the mean $\mathrm{Var}(\hat{\mu})$ are given by

$$\hat{\mu} = \frac{\sum_i x_i w_i}{\sum_i w_i}, \quad \mathrm{Var}(\hat{\mu}) = \frac{1}{\sum_i w_i} \tag{4.1}$$

$$w_i = \frac{1}{\sigma_i^2}. \tag{4.2}$$

When all sample variances are equal ($\sigma_i^2 = \sigma^2 \ \forall i$), equation (4.1) reduces to the simple arithmetic mean.

[3]The Voigt profile of each line must be aligned in velocity space before averaging, which will have the effect of over sampling high frequency lines relative to low frequency lines.

[4]Brightness temperature sensitivity is equivalent to the RMS noise in the measurement, and the two terms are used interchangeably.

The inverse-variance-weighted average assumes that each sample has the same expected mean. However, peak line strength is inversely proportional to frequency, $T_L \propto \nu^{-1}$, meaning the expected mean is not constant. To account for this, we rescale each line and its RMS noise to a common reference frequency ν_0. After rescaling, each observation has the same expected mean but different expected sample variances, allowing us to apply inverse-variance weighting. The rescaled line strength x', RMS noise σ', and weights w' are given by

$$x'_i = x_i \frac{\nu_i}{\nu_0}, \quad \sigma'_i = \sigma_i \frac{\nu_i}{\nu_0}, \quad w'_i = \frac{\nu_0^2}{\sigma_i^2 \nu_i^2}. \tag{4.3}$$

The choice of ν_0 is arbitrary and simply changes the reference point of the estimated mean. The signal-to-noise ratio (SNR) of the stacked estimate is independent of the choice of ν_0 and is given by

$$\mathrm{SNR} = \frac{\hat{\mu}}{\sqrt{\mathrm{Var}(\hat{\mu})}} = \frac{\sum_i x_i w_i}{\sqrt{\sum_i w_i}}. \tag{4.4}$$

4.2 Signal-to-Noise Ratio by Sky Region

Peak line intensity T_L is inversely proportional to frequency, meaning that lower-frequency lines are expected to be strongest and easiest to detect. However, CHORD is also less sensitive at low frequency, as seen in Figure 3.9, due to the dominant impact of background noise. It is therefore important to understand how the SNR varies across the CHORD band depending on the sky location. To compute the SNR, we take the ratio of the expected signal, which is the line intensity T_L from equation (1.2), to the RMS noise, which is the total sensitivity σ_T from section 3.4. In this section we use the native angular resolution per line, which is the smallest possible synthesized beam for the given frequency, and a common frequency resolution of 5 kHz, representing the maximum possible precision per line. We analyze smoothing in section 4.2.1.

To visualize SNR by sky region, we assume a uniform line intensity of $T_L = 10\,\mathrm{mK}$ at 1500 MHz across the entire sky. In reality, line intensity will vary significantly across the sky, so the SNR scale is not accurate. However, the shape and maxima of the SNR-per-frequency plots are still valid and can be used to understand which lines are most important for a given sky location. The locations used are shown in Figure 4.1, and the SNR versus frequency for each location is shown in Figure 4.2.

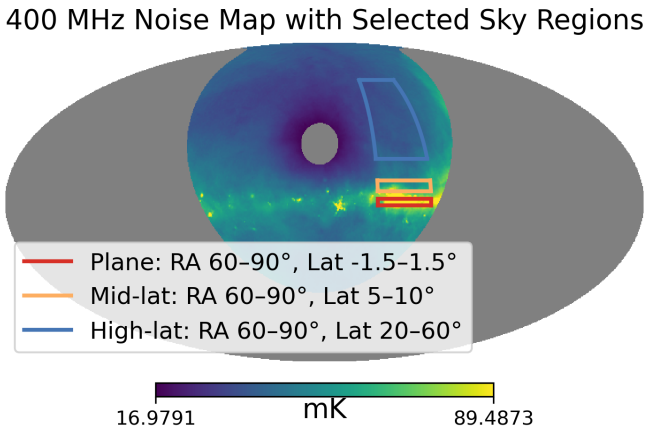


Figure 4.1: 400 MHz noise map showing the three selected regions used to demonstrate SNR variations for high Galactic latitude, mid Galactic latitude, and Galactic plane cases.

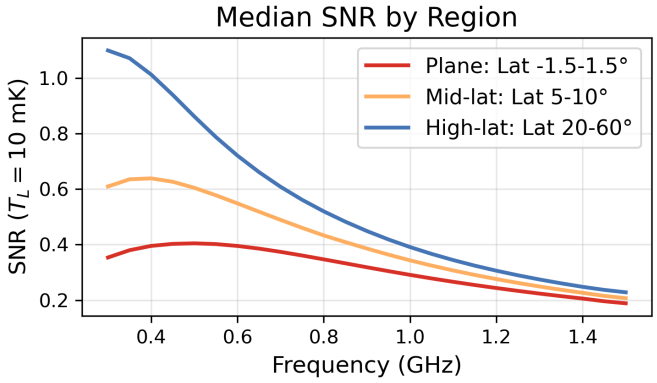
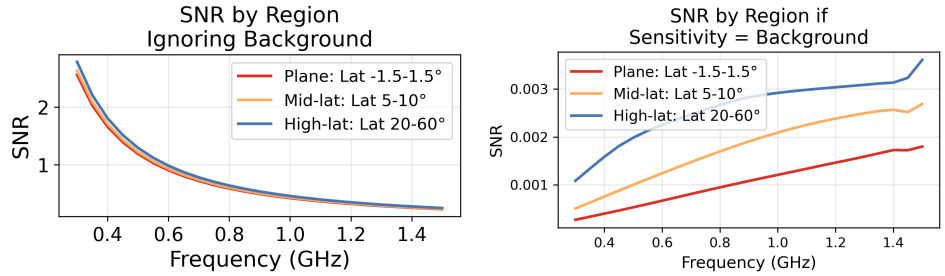


Figure 4.2: SNR per line by sky location for a common 5 kHz channel width and native angular resolution, demonstrating that SNR peaks around 500 MHz in the Galactic plane, around 400 MHz at mid Galactic latitudes, and around 300 MHz at high Galactic latitudes.



(a) System noise only

(b) Background noise only

Figure 4.3: In the system-noise-only case, SNR peaks at low frequency due to the wider beam and higher signals. When sensitivity is assumed equal to background temperature, SNR peaks at high frequency.

To understand what is driving the SNR trends by sky region, we can separate the impact of system effects (beam shape, integration time, declination) from the impact of background noise. In Figure 4.3 we plot SNR ignoring all background and SNR assuming sensitivity is equal to the background. Based on system effects alone, SNR is significantly higher at low frequency due to the compound effect of higher signal and higher sensitivity. In contrast, when sensitivity is directly proportional to background temperature, SNR is highest at high frequency where background noise is lowest. In reality, sensitivity is a combination of system effects, beam shape, and background noise, which is why SNR peaks differently depending on the sky location.

4.2.1 Signal-to-Noise Ratio after Smoothing

The results of Figure 4.2 use a common bandwidth of 5 kHz and apply each line's native angular resolution. We showed in section 3.2 that the synthesized beam varies from $\sim 21'$ at 300 MHz to $\sim 4.2'$ at 1500 MHz. Therefore, angular resolution varies from line to line. Furthermore, a 5 kHz bandwidth corresponds to a $\sim 5\,\mathrm{km\,s^{-1}}$ velocity resolution at 300 MHz and a $\sim 1\,\mathrm{km\,s^{-1}}$ resolution at 1500 MHz, meaning there is a unique velocity sampling for each line.

A common approach employed by L. D. Anderson et al. [1] and B. Liu et al. [11] is to smooth each line to a common velocity and angular resolution. We smoothed velocity resolution by applying a custom bandwidth ($\Delta\nu$) to

each line corresponding to a common velocity resolution Δv,

$$\Delta\nu = \frac{\Delta v}{c}\nu. \tag{4.5}$$

Velocity smoothing reduces RMS noise of high frequency lines by increasing the effective bandwidth, which impacts the point-source sensitivity of equation 3.18.

We smoothed angular resolution by averaging the native synthesized beam over the larger smoothed synthesized beam. For a smoothed solid angle Ω_{smoothed}, and native solid angle Ω_{native}, the number of independent samples is

$$N = \frac{\Omega_{\text{smoothed}}}{\Omega_{\text{native}}}. \tag{4.6}$$

We scale the brightness temperature sensitivity from equation 3.25 by $1/\sqrt{N}$ to account for the averaging of independent samples, giving

$$\sigma_T[\text{K}] = \frac{\sigma_{\,\text{RMS}} c^2}{\Omega_{\text{native}} 2k\nu^2}\sqrt{\frac{\Omega_{\text{native}}}{\Omega_{\text{smoothed}}}} \tag{4.7}$$

$$= \frac{\sigma_{\,\text{RMS}} c^2}{2k\nu^2\sqrt{\Omega_{\text{native}}\Omega_{\text{smoothed}}}}. \tag{4.8}$$

It is interesting to note that the smoothed brightness temperature sensitivity of equation 4.8 is equivalent to treating the solid angle as the geometric mean of the native and smoothed solid angles. One caveat of this approach is that it assumes each synthesized beam sample is independent, when in reality noise may be correlated over spatially adjacent samples. A second caveat is that previous surveys such as [1] and [11] perform angular smoothing by convolving native-resolution maps with a Gaussian kernel, whereas we have performed a non-weighted boxcar average. Despite these differences, the overall relationship between smoothing and SNR remains the same.

We computed SNR by region using a common velocity resolution of $5\,\text{km}\,\text{s}^{-1}$ and a common angular resolution of $22'$. The results are shown in Figure 4.4. SNR is significantly improved at high frequency, which is expected since both smoothing effects improve sensitivity at high frequency. Low-frequency lines still benefit from a larger primary beam and increased RRL signal strength. However, these results become outweighed by the stronger radio background at low frequency.

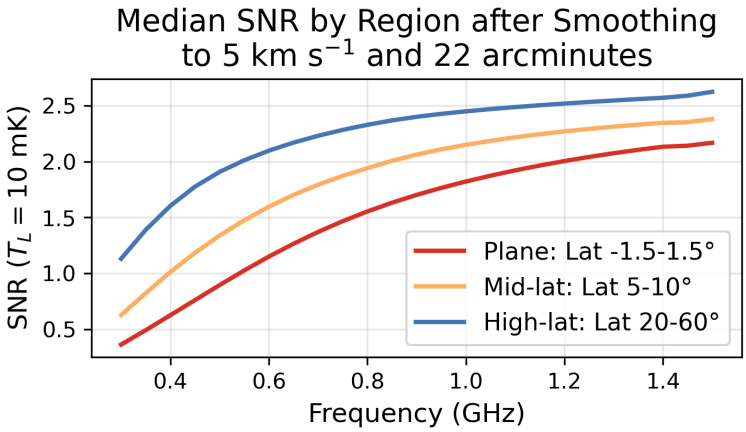


Figure 4.4: SNR per line by sky location after smoothing to a common velocity resolution of 5 km s^{-1} and a common angular resolution of 22 arcminutes. SNR is significantly improved at high frequency due to the smoothing effects.

We have shown that SNR per line is not a static quantity. It varies spatially based on changing background noise, and it varies with design choices such as velocity and angular smoothing. Therefore, the optimal stacking strategy, and the decision of which lines to target, depends heavily on the specific science goals.

Chapter 5

Extension CHORD

5.1 Candidate Site Selection

The angular resolution of an interferometer is determined by the maximum baseline, which is the distance between the two furthest dishes. CHORD plans to include very long baselines by adding outrigger stations 3300 km and 1000 km away at Green Bank Observatory and Hat Creek Radio Observatory, respectively, to perform very long baseline interferometry (VLBI). These outriggers are designed to provide precise localization of bright transient events like fast radio bursts, but do not provide smaller angular resolution for extended sources such as RRLs. Instead, tens of additional dishes can be added to the Dominion Radio Astrophysical Observatory (DRAO) site, which can supplement the core array with longer baselines. The purpose of these additional dishes, dubbed Extension CHORD, is to improve the angular resolution of the synthesized beam, while maintaining the highly sensitive redundant baseline architecture of CHORD. The impact on overall point source and surface brightness sensitivity will be minimal, as increasing collecting area from 512 dishes to 544 dishes only improves the sensitivity by 6 percent, and so it is ignored in the sensitivity analysis of section 3.1.

The DRAO site is located on ecologically sensitive land, which is developed in partnership with Indigenous stakeholders, and only certain areas of the site have the power infrastructure to support additional dishes. We selected candidate locations to model Extension CHORD by identifying the areas most suitable for development, such as places that already have disturbed ground and preexisting power connections (K. Phillips, private communication). A map of the site showing the proposed labelled locations is shown in Figure 5.1. For each candidate location we assume 16 additional dishes could be added, with the exception of location G, which is only large enough for 8 additional dishes. Location A2 is an alternate placement of location A which avoids dismantling the CHORD fabrication tent.

Location	Description	b_{EW}	b_{NS}
A	CHORD fabrication tent and gravel parking area.	600 m	250 m
A2	Alternate placement of location A avoiding the CHORD fabrication tent	600 m	250 m
B	Gravel immediately south of Synthesis Telescope track	220 m	120 m
C	Disturbed land immediately west of CHIME array	200 m	0 m
D	Disturbed land south of CHIME array	180 m	150 m
E	Disturbed land northeast of blockhouse	20 m	220 m
F	CHIME Pathfinder	250 m	70 m
G	Natural land near north water pump station	250 m	550 m

Table 5.1: Extension CHORD candidate locations at the DRAO site. b_{EW} and b_{NS} represent the approximate East-West and North-South distances from the centre of the location array to the centre of the core array.

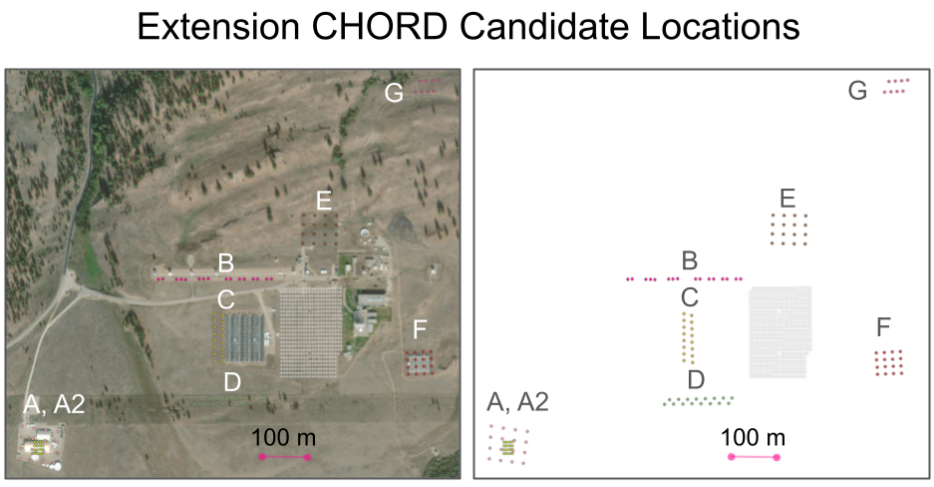


Figure 5.1: Site map for Extension CHORD. Each potential location is labelled A through G. The locations are selected to minimize construction effort by utilizing areas of disturbed ground and preexisting power connections.

5.2 (u, v) Coverage and Synthesized Beam

The Fourier transform of an image decomposes the image into a sum of spatial sinusoids of different frequencies. Spatial frequencies pointing East−West in the image are labelled u, while spatial frequencies pointing North−South are labelled v. Each baseline (the vector separating two dishes) of an interferometer is sensitive to a specific spatial frequency, set by the distances b_{EW} and b_{NS} between the two dishes and the wavelength of observation λ, as given by the equations

$$u = \frac{b_{\mathrm{EW}}}{\lambda}, v = \frac{b_{\mathrm{NS}}}{\lambda}. \tag{5.1}$$

Shorter baselines are sensitive to large-scale spatial structures on the sky, while longer baselines provide fine details. The collection of all u and v values sampled is called the (u, v) coverage. The synthesized beam, sometimes called the 'dirty beam', is the shape on the sky that a telescope is sensitive to given its (u, v) coverage. To perfectly reconstruct an image of the sky, every spatial frequency would need to be measured. An infinitely sampled (u, v) coverage would produce a delta-function synthesized beam, meaning a single point on the sky could be perfectly measured. In practice there are a finite number of baselines, creating a finite (u, v) coverage, so the reconstructed image will contain distortions. The shape of the synthesized beam determines the type of distortions and the overall angular resolution.

The (u, v) coverage and synthesized beam for the core CHORD array without any proposed extension dishes are shown in Figure 5.2. The (u, v) coverage is extremely evenly sampled, which is reflected in the Gaussian shape of the synthesized beam. Angular resolution is defined as the FWHM of the synthesized beam and is plotted as a red contour. The angular resolution matches the theoretical values of $7.41'$ and $5.05'$ in E−W and N−S directions, respectively, with the asymmetry due to the longer baselines in the v direction, as seen in the (u, v) coverage plot.

5.3 Weighting Methods

During imaging the (u, v) plane is discretely sampled by dividing the plane into a grid of cells and summing the contributions from each baseline within a cell. Sensitivity is maximized by weighting each cell equally, which is called natural weighting. However, spatial resolution can be improved by applying more weight to the longer baseline cells, which is called uniform

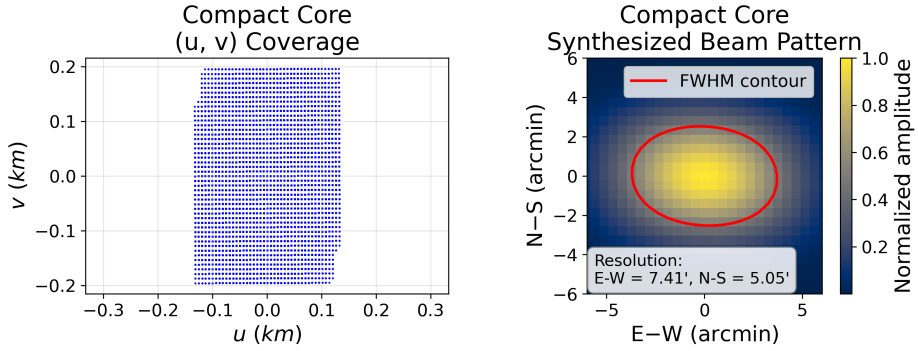


Figure 5.2: (u, v) coverage and synthesized beam for the full 512-element core CHORD array at 900 MHz. The angular resolution is plotted as a red contour of the FWHM of the synthesized beam.

weighting. A middle ground between the two is called robust weighting [3], which applies a tunable weight to balance sensitivity and resolution. The weight value w_k for cell k given a robust value R is

$$w_k = \frac{1}{1 + W_k f^2}, \qquad f^2 = \left(5 \cdot 10^{-R}\right)^2 \frac{\sum_k W_k}{\sum_k W_k^2}, \tag{5.2}$$

where W_k is the number of baselines that fall in cell k and the robust value R takes the suggested range of $[-2, +2]$. For $R \to +2$ the factor $f^2 \to 0$ and robust weighting approaches natural weighting. For $R \to -2$ the factor f^2 becomes large, so $w_k \to 1/(W_k f^2) \propto 1/W_k$, which down-weights densely sampled cells and approaches uniform weighting. Intermediate values of R interpolate smoothly between the two extremes.

Improving angular resolution comes at the cost of sensitivity, since signal from densely sampled cells is down-weighted. The amount of sensitivity lost can be quantified by the noise factor [3], which is the ratio of the sensitivity of the applied weighting to the sensitivity of natural weighting. The noise factor is given by

$$\text{Noise Factor} = \frac{\sigma_R}{\sigma_{\text{nat}}} = \frac{\sqrt{\sum_k w_k^2} / \sum_k w_k}{\sqrt{\sum_k W_k^2} / \sum_k W_k}, \tag{5.3}$$

where σ_R and σ_{nat} are the image noise levels of the applied and natural weighting, respectively.

5.4 Evaluating Beam Patterns

Here we analyze in detail the impact of adding location A to the CHORD array to illustrate how irregularly sampled (u, v) coverage affects the synthesized beam. The (u, v) coverage in the top left of Figure 5.3 shows the added long baselines, which appear as densely sampled zones at ± 0.6 km in u and ± 0.3 km in v, while the rest of the plane remains unsampled. The top right plot shows the combined core plus location A beam using natural weighting, where the beam is almost unchanged compared to the core array alone shown in Figure 5.2. This is expected since adding only 16 dishes has a minimal impact compared to the 512 dish core array. The bottom left plot shows the beam using uniform weighting, which applies more weight to the long baselines of location A than to the core. The uniform beam has multiple peaks, which arise from the (u, v) plane being poorly sampled. While the central peak does achieve an angular resolution close to the theoretical resolution of $1.69'$ and $3.13'$ in the E−W and N−S directions, respectively, the multiple peaks make it impossible to apply this angular resolution to practical observations. Instead, we define the 'effective angular resolution' as the maximum extent of the FWHM contours in the E−W and N−S directions, plotted as an ellipse. Based on the effective angular resolution, the uniform beam achieves $4.92'$ and $3.95'$ in the E−W and N−S directions, respectively, at a noise factor cost of 2.83. To understand the trade-off between angular improvement and sensitivity loss, the bottom right plot shows effective resolution versus noise factor for a range of robust values, where a noise factor of exactly 1 corresponds to natural weighting and the maximum noise factor corresponds to uniform weighting. From the elbow curve we can see that a roughly 25% improvement in angular resolution can be achieved for a marginal increase in noise factor.

5.5 Optimal Extension CHORD Configurations

In this section we consider configurations of at most two extension locations and compare every permutation. We compare the uniform weighting results, as they represent the maximal possible improvement. Comparisons for other weighting methods are given in appendix B.1 and all arrive at the same conclusion. We compare configurations using the geometric mean resolution, the diameter of the circle whose area matches that of the effective resolution ellipse, with results for all permutations shown in Figure 5.4. The best individual locations are B, C, and F at $3.16'$ each, and the worst indi-

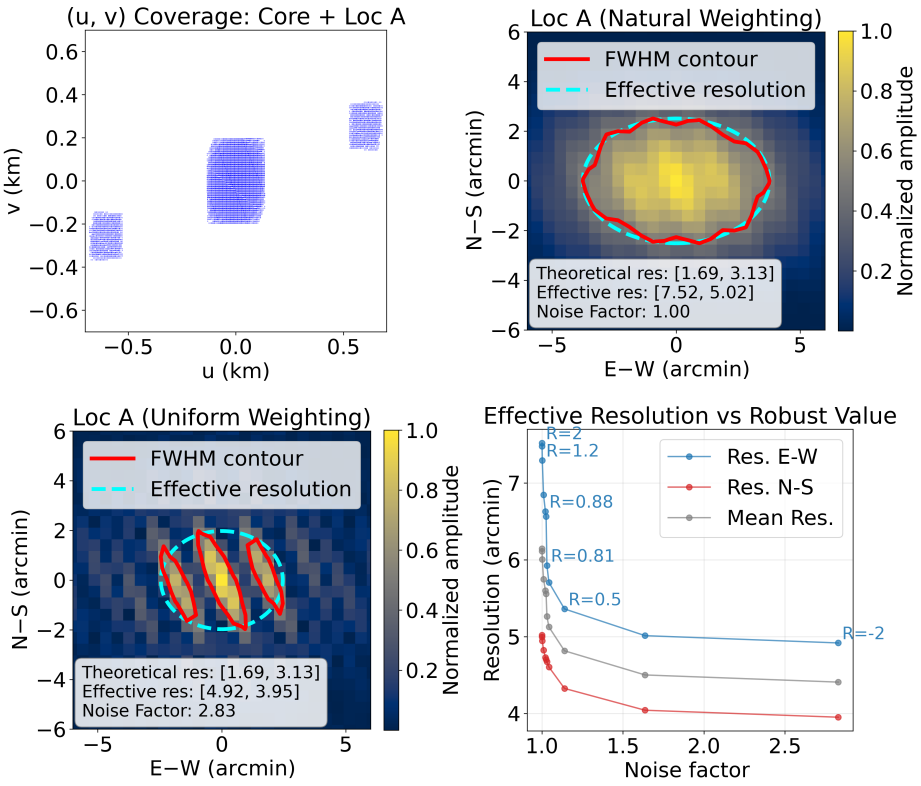


Figure 5.3: Effect of adding location A to the core CHORD array. Top left: (u, v) coverage. Top right: synthesized beam under natural weighting. Bottom left: synthesized beam under uniform weighting, with the effective angular resolution plotted as an ellipse. Bottom right: effective resolution versus noise factor over a range of robust values.



Figure 5.4: Geometric mean effective resolution for single locations (top diagonal) and all pairs (lower triangle) of extension CHORD locations under uniform weighting. The best individual locations are B, C, and F at 3.16′ and the best combination is B+D at 2.26′.

vidual locations are A/A2 and G at 4.06′ and above. The individual location findings are counterintuitive since the longest-baseline sites (A/A2 and G) should have the best theoretical resolution. However, the beam-quality impact of irregular (u, v) coverage outweighs this advantage. The best overall location is B+D at 2.26′.

After exploring all the combinations (Appendix B.1 Figures B.5 through B.11), the best overall configuration is B+D. The primary drivers for this selection are:

1. B is an excellent individual location, with a best effective resolution of 3.16′ and a regular beam shape, as seen in the top left of Figure 5.5, making it the best stand-alone choice.

2. D is a complementary shape to B, shown in the top right of Figure 5.5, which fills (u, v) coverage gaps in B as shown in appendix B.1.

3. The B+D combination can achieve a resolution of 3.45′ at a noise factor of just 1.26, which is a 44% improvement in angular resolution for a 26% loss in sensitivity, while having a perfectly symmetric beam shape, as seen in the bottom left of Figure 5.5.

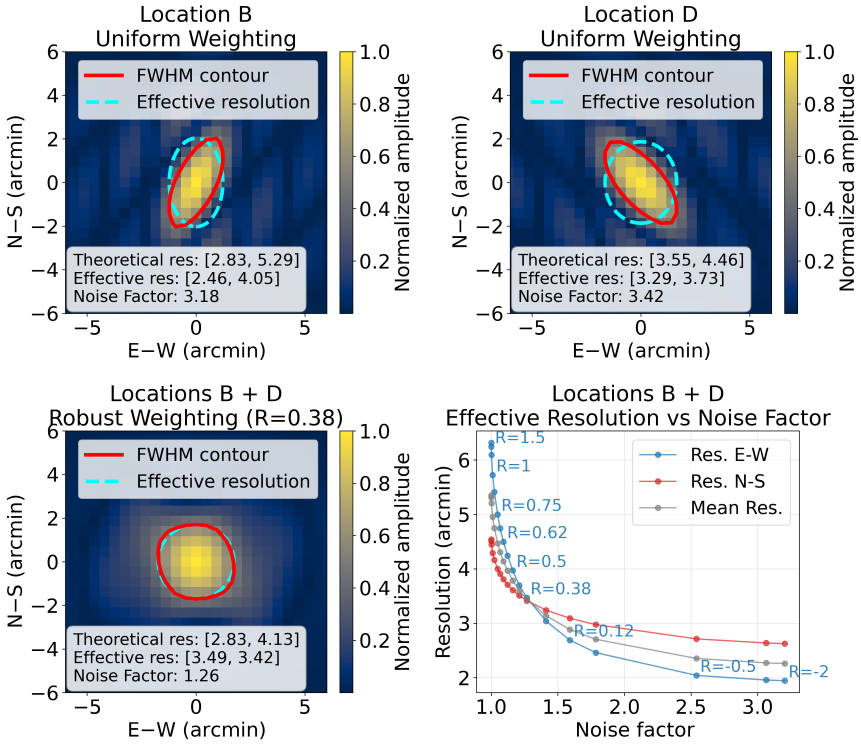


Figure 5.5: Synthesized beams for locations B, D, the combination B+D, and a plot of resolution versus noise factor. The beam shapes for B and D are complementary; when added together they produce a symmetric beam shape. The B+D combination can achieve a resolution of 3.45′ at a noise factor of 1.26, or a resolution of 2.26′ at a noise factor of 3.

4. For situations where sensitivity is not important, the B+D combination achieves the best overall effective resolution of 2.26′ at a noise factor of 3, which is a 63% improvement in angular resolution for a 200% loss in sensitivity, as seen in the bottom right of Figure 5.5.

Chapter 6

Simulations

6.1 Simulations of SIGGMA Observations

In this section we describe the RRL measurements from the Survey of Ionized Gas in the Galaxy, Made with the Arecibo Telescope (SIGGMA) [11] and estimate CHORD's ability to detect the H II regions reported by SIGGMA.

The SIGGMA survey covers a thin strip of the Galactic plane between Galactic longitudes $32\,^\circ \leq \ell \leq 70\,^\circ$ and Galactic latitudes $|b| \leq 1.5\,^\circ$ with a frequency range of 1225 MHz to 1525 MHz. A plot of the SIGGMA data showing the overlap with CHORD is shown in Figure 6.1. The survey averages the intensity profile of 12 RRLs (H163α to H174α) and covers a velocity range of $\pm 300\,\mathrm{km\,s^{-1}}$ with a $5.1\,\mathrm{km\,s^{-1}}$ channel resolution and a 6-arcminute angular resolution. An inner-Galaxy data release [10] provides measurements of 244 known H II regions and an additional 79 candidate regions[5], of which 14 and 12, respectively, are within the CHORD survey range. The H II regions visible to CHORD are plotted over the CHORD sensitivity map at 900 MHz shown in Figure 6.2, which shows that most H II regions are located in areas with high background noise[6] A subset of the visible H II regions is provided in table 6.1 and a full table is available in appendix C.1.

[5]Data available at `https://vizier.cds.unistra.fr/viz-bin/VizieR?-source=J/ApJS/240/14`

[6]RRL intensity is expected to be strongly correlated with background noise since the factors that produce RRL emission, temperature and electron density, are also responsible for bright free-free emission.

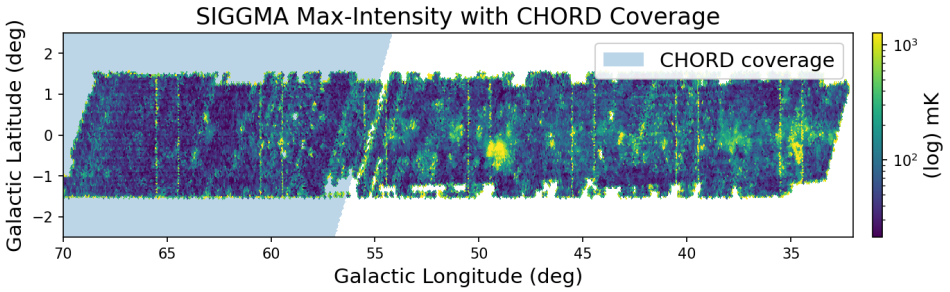


Figure 6.1: The SIGGMA data products, showing peak intensity per pixel. The area of declination $\delta \geq 20^\circ$ observable by CHORD is highlighted in blue.

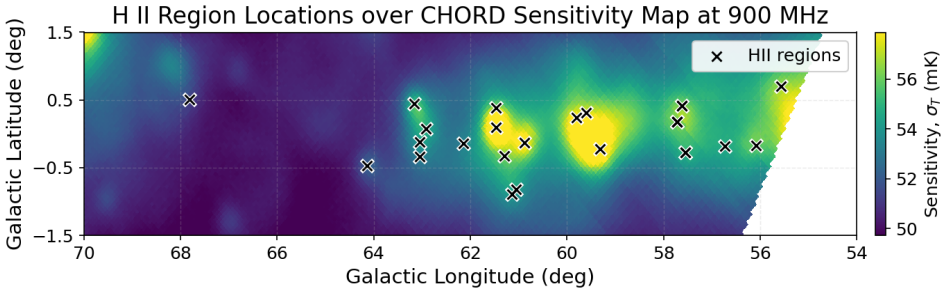


Figure 6.2: H II regions detected by SIGGMA plotted over the CHORD sensitivity map at 900 MHz, showing that most H II regions are in places with high background noise.

Order	Name	Peak [mK]	SNR	Type
1	G063.164+00.449	694.5 ± 40.6	16.9	Known
2	G061.473+00.094	348.6 ± 26.9	12.8	Known
3	G059.796+00.241	124.5 ± 11.4	10.7	Known
⋮	⋮	⋮	⋮	⋮
25	G057.715+00.173	17.1 ± 1.7	16.6	Candidate
26	G062.142−00.144	16.8 ± 2.6	6.2	Candidate

Table 6.1: A subset of the H II region results reported by SIGGMA that are within the CHORD coverage, ordered from highest to lowest peak line strength. A full table is available in appendix C.1.

The 12 stacked lines have an average frequency of 1375 MHz, so we define the peak line strength as the brightness temperature at 1375 MHz, $T_L(1375\ \text{MHz})$. We can then obtain the peak line strength for each line in the CHORD frequency range by scaling from this reference frequency,

$$T_L(\nu) = T_L(1375\ \text{MHz})\frac{1375\,\text{MHz}}{\nu}. \tag{6.1}$$

We computed the single-day per-line SNR of each of the 26 H II regions by dividing peak line strength by the CHORD sensitivity. We applied a common frequency resolution of 5 kHz and used CHORD's native angular resolution for each line, based on the precise spatial location of the given H II region. A plot of per-line SNR for one observing day is shown in Figure 6.3. SNR is maximized between 500 and 600 MHz, which is expected for observations along the Galactic plane. The lowest SNR is around 0.5, while the highest is 24. The two very bright H II regions 'G063.164+00.449' and 'G061.473+00.094' are significant outliers, with SNRs of 24 and 10, respectively, and are omitted from Figure 6.3.

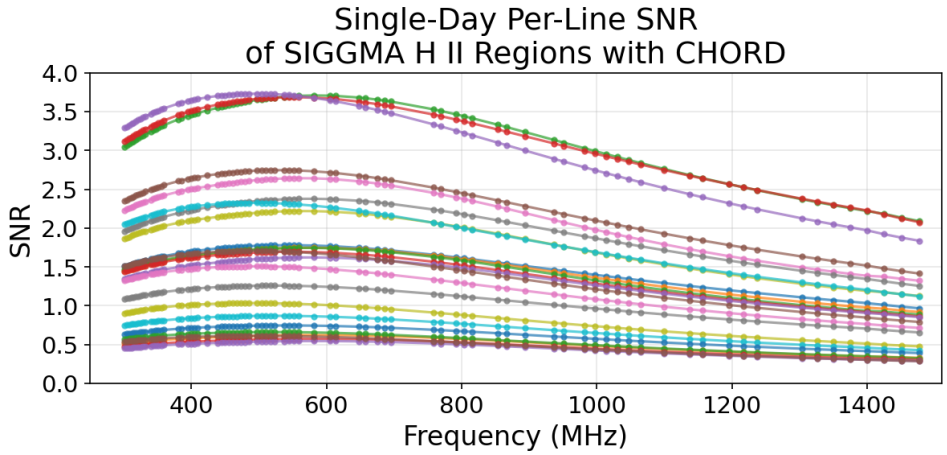


Figure 6.3: Single-day per-line SNR with a 5 kHz frequency resolution and native angular resolution for each H II region visible to CHORD, excluding the two very bright outliers, showing that SNR peaks between 500 and 600 MHz and is in the range 0.5 to 3.7.

Next we analyzed the impact of smoothing on SNR using a velocity resolution of $5\,\text{km}\,\text{s}^{-1}$ and a common angular resolution of 22 arcminutes as described in section 4.2.1. The resultant SNR is shown in Figure 6.4.

We also include the locations of the 5, 6, 8, 12 and 22 arcminute angular resolution cutoffs as dashed vertical lines. The number of lines within each cutoff is 8, 14, 27, 42, and 79, respectively. When smoothing is applied, the high-frequency lines have the highest SNR, so we prioritize them for stacking.

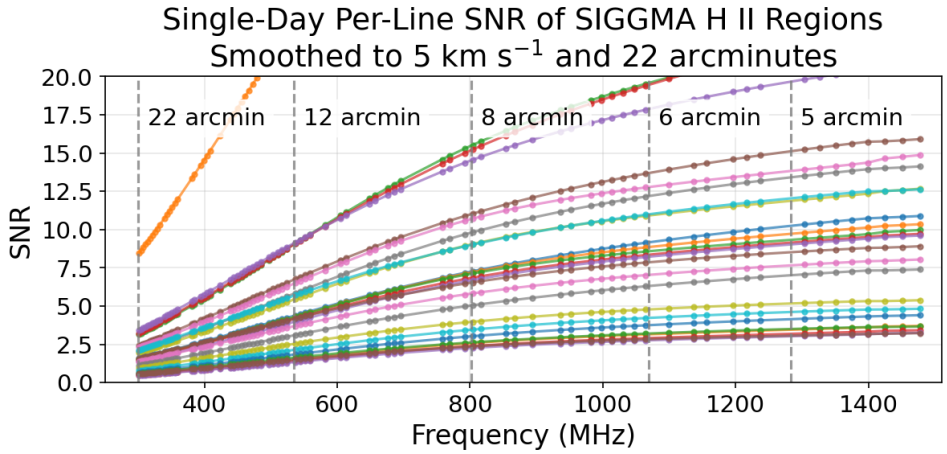


Figure 6.4: Single-day per-line SNR for 24 H II regions within the CHORD coverage, smoothed to a velocity resolution of $5\,\mathrm{km\,s^{-1}}$ and an angular resolution of 22 arcminutes. The frequency corresponding to an average CHORD angular resolution of 5, 6, 8, 12 and 22 arcminutes is shown as dashed vertical lines.

We simulated the line-stacking process described in section 4 using both the native-resolution SNR of Figure 6.3 and different smoothed resolutions. In the case of native resolution, lines are ordered from highest to lowest SNR, the best 1, 15 and 60 lines are stacked for each H II region. The resultant SNR predictions for one observing day are shown in table 6.2. The lowest single-day SNR with 60 stacked lines is 3.93 ± 0.39, which demonstrates that each region should be easily detectable by CHORD even before smoothing. Furthermore, each observing day is an independent sample, so SNR increases in proportion to the square root of the number of observing days. A true survey will observe each region between 9 and 36 days, meaning SNR is expected to increase by a factor of 3 to 6.

Order	Name	SNR best 1	SNR best 15	SNR best 60
1	G063.164+00.449	23.5 ± 1.4	90.98 ± 5.32	171.54 ± 10.03
2	G061.473+00.094	9.72 ± 0.75	37.47 ± 2.89	70.73 ± 5.46
3	G059.796+00.241	3.71 ± 0.34	14.26 ± 1.31	26.71 ± 2.45
⋮	⋮	⋮	⋮	⋮
25	G057.715+00.173	0.55 ± 0.05	2.10 ± 0.21	3.93 ± 0.39
26	G062.142−00.144	0.61 ± 0.09	2.37 ± 0.37	4.45 ± 0.69

Table 6.2: Predicted CHORD SNR after ordering and stacking the best 1, 15 and 60 lines from each SIGGMA H II region. Results show SNR after one observing day using native angular resolutions. A full results table is in appendix C.1.

We also considered the case of prioritizing angular resolution over sensitivity by only including the lines that meet a given angular resolution cutoff. We selected angular resolutions of 5, 6, and 8 arcminutes, and smoothed the available lines to that cutoff. We also smoothed the velocity resolution to $5\,\mathrm{km\,s^{-1}}$. The results are shown in table 6.3. The results show that with a 6-arcminute angular resolution, the predicted stacked SNR per day is 3 or greater, which matches SIGGMA results if using 4 or more observing days. For an improved angular resolution of 5 arcminutes, at least 9 observing days are required to obtain an SNR of 6 or greater.

Order	Name	SNR $5'$	SNR $6'$	SNR $8'$
1	G063.164+00.449	85.04 ± 4.97	130.52 ± 7.63	222.38 ± 13.00
2	G061.473+00.094	41.86 ± 3.23	63.49 ± 4.90	105.47 ± 8.14
3	G059.796+00.241	14.70 ± 1.35	22.45 ± 2.06	37.98 ± 3.48
⋮	⋮	⋮	⋮	⋮
25	G057.715+00.173	2.02 ± 0.20	3.11 ± 0.31	5.32 ± 0.53
26	G062.142−00.144	2.03 ± 0.31	3.13 ± 0.49	5.42 ± 0.84

Table 6.3: Predicted SNR for CHORD observations of SIGGMA H II regions with a smoothed angular resolution of 5, 6 and 8 arcminutes after stacking 8, 14, and 27 lines, respectively. A full results table is in appendix C.1.

A limitation of this analysis is that we estimate the radio background using the global sky model [20], which has an angular resolution of 56 arcminutes while the SIGGMA survey has an angular resolution of 6 arcminutes.

The background model suffers from beam dilution, when a bright but compact source is averaged over the larger beam size, the specific source location will be understated. RRL emission is strongly correlated with background emission. If the size of an H II region reported by SIGGMA is much smaller than the 56 arcminute beam size of the global sky model, then the background emission for that precise location is likely to be understated by the global sky model, causing our SNR estimates to be overstated.

Chapter 7

Conclusion

7.1 Conclusions

In this work, I analyzed the RFI environment at the DRAO using a one-hour segment of data from the DRAO RFI monitor and estimated the number of detectable H II lines. As part of the analysis I found that each calibration segment necessitated an independent RFI label to account for changes in the gain factor. I applied spectral kurtosis to integrated power and, after proper calibration, found that it closely matched theory. The generated RFI labels were compared against known RFI events and found strong agreement. We considered a frequency bin clean provided that it was RFI-free in at least 50% of the time segments, and we applied a Gaussian-weighted average over frequency bins to classify whether a given RRL would be visible. I found that a total of 79 RRLs in the range of 300 – 1500 MHz are detectable and 37 are contaminated by RFI. The 79 detectable lines are used as the basis for all future detection analysis.

I derived the theoretical point source sensitivity and brightness temperature sensitivity of CHORD and found that intrinsic sensitivity increases at high declination and at low frequency. Radio background adds linearly to the system temperature and is strongest at low frequency and along the Galactic plane. I found that sensitivity degrades at low frequency due to the strong background and that the overall sensitivity maps are highly spatially dependent and frequency dependent. Rather then estimating a single CHORD sensitivity value, this work presents sky maps which provide sensitivity for any given frequency or sky location.

I showed how an inverse-variance-weighted average can be used to maximize the signal-to-noise ratio of multiple RRLs having different sensitivities. I explored how the optimal RRL stacking order varies by sky location due to the changing SNR per line. I analyzed both velocity and angular smoothing to show that lines can be stacked more effectively at the cost to angular precision. As with sensitivity, the optimal stacking strategy depends on sky

location and on scientific goals, namely which of sensitivity, velocity resolution, or angular resolution is most valuable.

I modelled adding dishes to the main CHORD array to create the Extension CHORD array. We identified 8 locations as candidate sites and modelled all combinations of 1 or 2 locations, for 40 total combinations. I found that the Synthesis Telescope track in combination with the disturbed land south of CHIME represented the optimal locations for constructing Extension CHORD. While these two locations do not represent the longest possible baselines, they work together to produce excellent (u, v) coverage and a symmetric synthesized beam without any distortions. I found that adding 32 dishes to the recommended locations could improve angular resolution by 44% while reducing sensitivity by 25%, or could improve angular resolution by 63% for a loss in sensitivity of 200%.

Finally, I estimated CHORD's detection ability of real H II regions by simulating the CHORD response on data published by the SIGGMA survey. In the worst case, CHORD could match SIGGMA SNR levels with a common 6 arcminute angular resolution after 4 observing days, and could improve angular resolution to 5 arcminutes with matching SNR levels after 9 days. In other cases CHORD can detect the bright H II regions measured by SIGGMA after just one day and in most cases using only one RRL. In conclusion, I've shown that an RRL survey with CHORD could both improve detections of known H II regions and likely measure new ones, building a more complete picture of the ionized interstellar medium.

Bibliography

[1] Anderson, L. D., Luisi, M., Liu, B., et al. 2021, The GBT Diffuse Ionized Gas Survey (GDIGS): Survey Overview and First Data Release, The Astrophysical Journal Supplement Series, 254, 28, doi: 10.3847/1538-4365/abef65

[2] Bij, A., & The CHORD Collaboration. 2026, CHORD HI-Galaxy Survey Forecasts: Searching for Nearby Dark Galaxies and High Redshift Giants, In Prep

[3] Briggs, D. S. 1995, High Fidelity Deconvolution of Moderately Resolved Sources, PhD thesis

[4] Bruce, N., Baard, C., Harrison, S., et al. 2026, Wideband RFI Monitor Requirements, Design, and Commissioning at DRAO, arXiv, doi: 10.48550/arXiv.2605.08479

[5] Carroll, B. W., & Ostlie, D. A. 2017, An Introduction to Modern Astrophysics:, 2nd edn. (Cambridge University Press), doi: 10.1017/9781108380980

[6] Condon, J. J., & Ransom, S. M. 2016, Essential Radio Astronomy (Princeton University Press), doi: 10.2307/j.ctv5vdcww

[7] Hopkins, H. S., Lang, D., Smith, K., et al. 2026, Matched Filtering for the Canadian Hydrogen Observatory and Radio-Transient Detector Galaxy Search, arXiv, doi: 10.48550/arXiv.2603.03520

[8] Jo, A., Bm, G. K., & F, G. 2016, Performances of Several Univariate Tests of Normality: An Empirical Study, Journal of Biometrics & Biostatistics, 07, doi: 10.4172/2155-6180.1000322

[9] Kendall, M. G. M. G. 1969, The Advanced Theory of Statistics (London : Griffin)

[10] Liu, B., Anderson, L. D., McIntyre, T., et al. 2019, Survey of Ionized Gas of the Galaxy, Made with the Arecibo Telescope (SIGGMA): Inner Galaxy Data Release, The Astrophysical Journal Supplement Series, 240, 14, doi: 10.3847/1538-4365/aaef8e

[11] Liu, B., McIntyre, T., Terzian, Y., et al. 2013, SIGGMA: A SURVEY OF IONIZED GAS IN THE GALAXY, MADE WITH THE ARECIBO TELESCOPE, The Astronomical Journal, 146, 80, doi: 10.1088/0004-6256/146/4/80

[12] Nita, G., Gary, D., & Hellbourg, G. 2017, Spectral kurtosis statistics of quantized signals, in Proceedings of 2016 Radio Frequency Interference: Coexisting with Radio Frequency Interference, RFI 2016 (Institute of Electrical and Electronics Engineers Inc.), 75–80, doi: 10.1109/RFINT.2016.7833535

[13] Nita, G. M., & Gary, D. E. 2010, Statistics of the Spectral Kurtosis Estimator, Publications of the Astronomical Society of the Pacific, 122, 595, doi: 10.1086/652409

[14] Nita, G. M., & Gary, D. E. 2010, The Generalized Spectral Kurtosis Estimator, Monthly Notices of the Royal Astronomical Society: Letters, 406, L60, doi: 10.1111/j.1745-3933.2010.00882.x

[15] Price, D. C. 2016, PyGDSM: Python Interface to Global Diffuse Sky Models, Astrophysics Source Code Library, ascl:1603.013

[16] Smith, E. T., Lynch, R., & Pisano, D. J. 2022, Simulating Spectral Kurtosis Mitigation Against Realistic RFI Signals, The Astronomical Journal, 164, 123, doi: 10.3847/1538-3881/ac7e47

[17] Taylor, J., Denman, N., Bandura, K., et al. 2019, Spectral Kurtosis-Based RFI Mitigation for CHIME, Journal of Astronomical Instrumentation, 08, 1940004, doi: 10.1142/S225117171940004X

[18] The CHORD Collaboration. 2026, Overview of the Canadian Hydrogen Observatory and Radio Transient Detector (CHORD) Project, in SPIE Astronomical Telescopes and Instrumentation, Vol. Submitted, Copenhagen, Denmark

[19] Vanderlinde, K., Liu, A., Gaensler, B., et al. 2019, The Canadian Hydrogen Observatory and Radio-transient Detector (CHORD), in Canadian Long Range Plan for Astronomy and Astrophysics White Papers, Vol. 2020, eprint: arXiv:1911.01777, 28, doi: 10.5281/zenodo.3765414

[20] Zheng, H., Tegmark, M., Dillon, J. S., et al. 2017, An Improved Model of Diffuse Galactic Radio Emission from 10 MHz to 5 THz, Monthly Notices of the Royal Astronomical Society, 464, 3486, doi: 10.1093/mnras/stw2525

[21] Zonca, A., Singer, L., Lenz, D., et al. 2019, Healpy: Equal Area Pixelization and Spherical Harmonics Transforms for Data on the Sphere in Python, Journal of Open Source Software, 4, 1298, doi: 10.21105/joss.01298

Appendix

Appendix A

RFI

A.1 Log-Gamma vs Gaussian Distributions

A.1.1 Gamma Distribution

Claim. A normalized gamma distribution approaches a Gaussian distribution for shape parameter $N \gg 1$.

Proof. Taking the log of a gamma PDF with shape parameter N and scale parameter θ (written as $f(x)$ instead of $f(x; N, \theta)$ for compactness) gives:

$$\ln f(x) = \ln \frac{x^{N-1} e^{-x/\theta}}{\theta^N \Gamma(N)} \tag{A.1}$$

$$= (N-1)\ln x - \frac{x}{\theta} - N \ln \theta - \ln \Gamma(N). \tag{A.2}$$

The mode can be found by taking the first derivative of the log PDF and setting it to zero:

$$\frac{d}{dx} \ln f(x)\bigg|_{x=x_0} = \frac{N-1}{x_0} - \frac{1}{\theta} = 0 \tag{A.3}$$

$$\Rightarrow x_0 = (N-1)\theta \approx N\theta. \tag{A.4}$$

The curvature at the mode can be found by taking the second derivative of the log PDF and evaluating it at the mode:

$$\frac{d^2}{dx^2} \ln f(x)\bigg|_{x=x_0} = -\frac{N-1}{x_0^2} \tag{A.5}$$

$$= -\frac{1}{(N-1)\theta^2} \approx -\frac{1}{N\theta^2}. \tag{A.6}$$

Taking a Taylor expansion of the log PDF around the mode x_0, noting that

the derivative of $\ln f(x)$ vanishes at $x = x_0$ since x_0 is the mode, gives:

$$\ln f(x) = \ln f(x_0) + \frac{1}{2}\,\frac{d^2}{dx^2} \ln f(x)\bigg|_{x=x_0} (x - x_0)^2 + \dots \tag{A.7}$$

$$\approx \ln f(x_0) - \frac{(x - x_0)^2}{2N\theta^2}. \tag{A.8}$$

Exponentiating,

$$f(x) \approx C \exp\left(-\frac{(x - x_0)^2}{2N\theta^2}\right) \tag{A.9}$$

$$\approx C \exp\left(-\frac{(x - N\theta)^2}{2N\theta^2}\right). \tag{A.10}$$

Consider a Gaussian distribution with mean $N\theta$ and variance $N\theta^2$:

$$g(x; N\theta, N\theta^2) = \frac{1}{\sqrt{2\pi N\theta^2}} \exp\left(-\frac{(x - N\theta)^2}{2N\theta^2}\right) \tag{A.11}$$

$$= C \exp\left(-\frac{(x - N\theta)^2}{2N\theta^2}\right). \tag{A.12}$$

This shows that the PDF of the gamma distribution approaches the PDF of the Gaussian distribution up to a constant C as N becomes large, so that $N - 1 \approx N$. It can be shown that the constant C is equal to $1/\sqrt{2\pi N\theta^2}$ by normalizing the gamma distribution and using Stirling's approximation for the gamma function. However, the normalizing constant does not impact the kurtosis, so the full derivation of it is omitted.

A.1.2 Log-Gamma Distribution

Claim. The log of a gamma distribution approaches a Gaussian distribution for shape parameter $N \gg 1$.

Proof. Let W be a log-gamma random variable defined as $W = \ln X$ where $X \sim \text{Gamma}(N, \theta)$. By a change of variables with $x = e^w$ and $dx = e^w\,dw$, the PDF of W is

$$f_W(w) = f_X(e^w)\,e^w = \frac{(e^w)^{N-1} e^{-e^w/\theta}}{\theta^N \Gamma(N)}\,e^w \tag{A.13}$$

$$= \frac{e^{Nw}\,e^{-e^w/\theta}}{\theta^N \Gamma(N)}. \tag{A.14}$$

The log PDF is

$$\ln f_W(w) = Nw - \frac{e^w}{\theta} - N\ln\theta - \ln\Gamma(N). \tag{A.15}$$

Applying the same procedure as above, the mode and curvature are obtained from the first two derivatives:

$$\frac{d}{dw}\ln f_W(w)\bigg|_{w=w_0} = N - \frac{e^{w_0}}{\theta} = 0 \;\Rightarrow\; w_0 = \ln(N\theta), \tag{A.16}$$

$$\frac{d^2}{dw^2}\ln f_W(w)\bigg|_{w=w_0} = -\frac{e^{w_0}}{\theta} = -N. \tag{A.17}$$

Taking the Taylor expansion about the mode, where again the first derivative term is zero, gives:

$$\ln f_W(w) = \ln f_W(w_0) + \frac{1}{2}\,\frac{d^2}{dw^2}\ln f_W(w)\bigg|_{w=w_0}(w-w_0)^2 + \ldots \tag{A.18}$$

$$\approx \ln f_W(w_0) - \frac{N}{2}(w-w_0)^2. \tag{A.19}$$

After exponentiating, the approximate form of the log-gamma PDF is

$$f_W(w;N,\theta) \approx C\exp\left(-\frac{N(w-w_0)^2}{2}\right). \tag{A.20}$$

This shows that the PDF of a log-gamma random variable $W = \ln X$ is approximately Gaussian to second order, with mean $w_0 = \ln(N\theta)$ and variance $1/N$.

A.2 Spectral Kurtosis Pair Plot

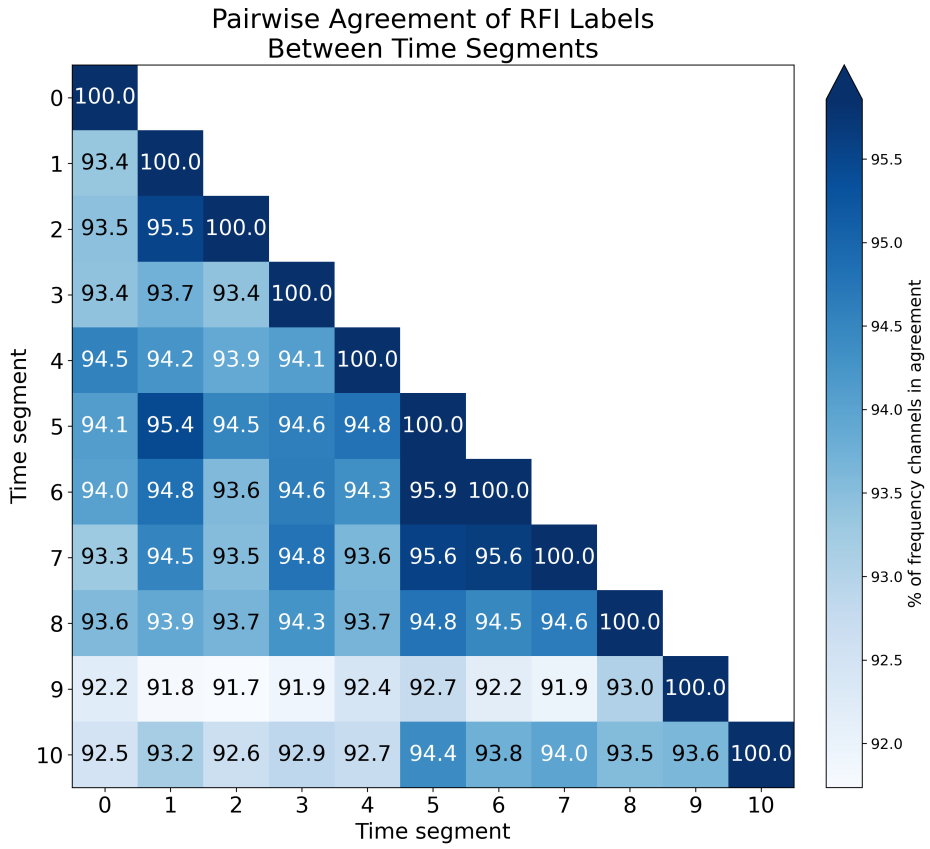


Figure A.1: Pair plot of agreement between RFI labels for each of the 11 time segments showing the extent of agreement between each pair of segments.

A.3 Alternative RFI Thresholds

Two subjective design decisions were made in the RFI analysis of section 2.4.1. The first was to use RFI thresholds of [0.1%, 99.9%] on calibrated spectral kurtosis. These thresholds were chosen as deliberately wide to reduce false positives (clean spectrum labelled RFI); however, they also increase false negatives (RFI spectrum labelled clean). Secondly, while aggregating the 11 independent time segments, we assumed that if at least 6 of the 11 segments are clean ('50%') we consider the frequency clean; alter-

native approaches are to require that at least one of the segments be clean ('any') or that all segments be clean ('all'), representing more permissive and more strict approaches respectively. The results of these alternative approaches are shown in table A.1.

Table A.1: Number of detectable hydrogen RRLs per frequency region, by cross-segment aggregation method (any/50%/all) and spectral kurtosis threshold (0.1%, 0.5%, 1%).

Aggregation	**any**			**50%**			**all**		
SK threshold (%)	0.1	0.5	1	0.1	0.5	1	0.1	0.5	1
Low frequency	32	32	32	29	29	29	0	0	0
Mid frequency	26	26	26	23	23	23	1	0	0
High frequency	32	32	31	27	26	25	0	0	0
Total	90	90	89	79	78	77	1	0	0

The results show that the choice of SK threshold has a minimal impact on total visible lines, impacting between just one or two lines. Meanwhile the aggregation method changes results significantly from 90 visible lines to just 1.

To understand the sensitivity to aggregation, we can imagine a spectrum that is completely RFI free. If we choose a 0.1% SK threshold then we expect a 0.2% false positive rate per channel per time segment (given that its a two sided test). Choosing the 'any' aggregation method, the channel is clean if at least one time segment is not a false positive, which has probability $1 - 0.002^{11} \approx 1$. The '50%' aggregation will be clean if at least 6 of 11 channels are not false positives, using the binomial distribution this has probability $P(X \geq 6) = \sum_{k=6}^{11} \binom{11}{k} 0.998^k \cdot 0.002^{11-k} \approx 1$. For both previous cases the probability of aggregating a false positive is essentially zero. Meanwhile, for the 'all' aggregation, the channel is clean when all time segments are clean which has probability $0.998^{11} = 0.978$, so the probability of a false positive per channel is 2.2%. We consider an RRL to be detectable by taking the Gaussian weighted average across a 400 km s^{-1} profile, which comprises 120 channels at 300 MHz or 592 channels at 1500 MHz. Given a 2.2% false positive rate, the probability that one or more of the 120 channels are false positive is $1 - 0.978^{120} = 0.93$, meaning that the line is likely to be characterized as undetectable based on false positives alone, explaining why the 'all' aggregation shows no detectable lines.

Since the probability of aggregating a false positive in both the 'any' aggregation and '50%' aggregation is extremely low, the differences we observe in detectable lines between them is likely the result of genuine intermittent RFI being detected and not just statistical noise.

Appendix B

Extension CHORD

B.1 Additional Extension CHORD Plots

B.1.1 Pairwise Comparisons

Pairwise comparisons were performed by looking at natural weighting, robust weighting [3] with $R = 0$, and uniform weighting for each permutation of extension CHORD locations. The combination B + D emerges as the best candidate in every weighting scheme. A combined metric of resolution times noise factor, which attempts to balance the trade-off between resolution and sensitivity, also shows that B + D is the best candidate across all weighting schemes. The pairwise comparisons are shown in figure B.1.

B.1.2 (u, v) Coverage of B + D

A plot of the (u, v) coverage of B, D, and B + D is shown in figure B.2, which shows the complementary nature of the two locations and that the combined (u, v) coverage plot is well-filled.

B.1.3 Beam Plots for Each Individual Location

The beam plots for each individual location plus the core CHORD array are shown in figures B.3 and B.4. The plots show the beam using natural weighting. Also included is a plot of angular resolution versus noise factor for a range of Briggs robust weighting parameter values from $R = +2$ (natural weighting) to $R = -2$ (uniform weighting).

B.1.4 Beam Plots for Each Location Pair

The beam plots for each location pair are shown in figures B.5 through B.11.

B.1.5 (u, v) Coverage and Beam Pattern for the Complete Array

(u, v) coverage and uniform-weighted beam pattern for the complete array (incorporating locations A, B, C, D, E, F, G and the core array) are shown in figure B.12. The resolution is 1.4 arcminutes in the E−W direction and 1.5 arcminutes in the N−S direction.

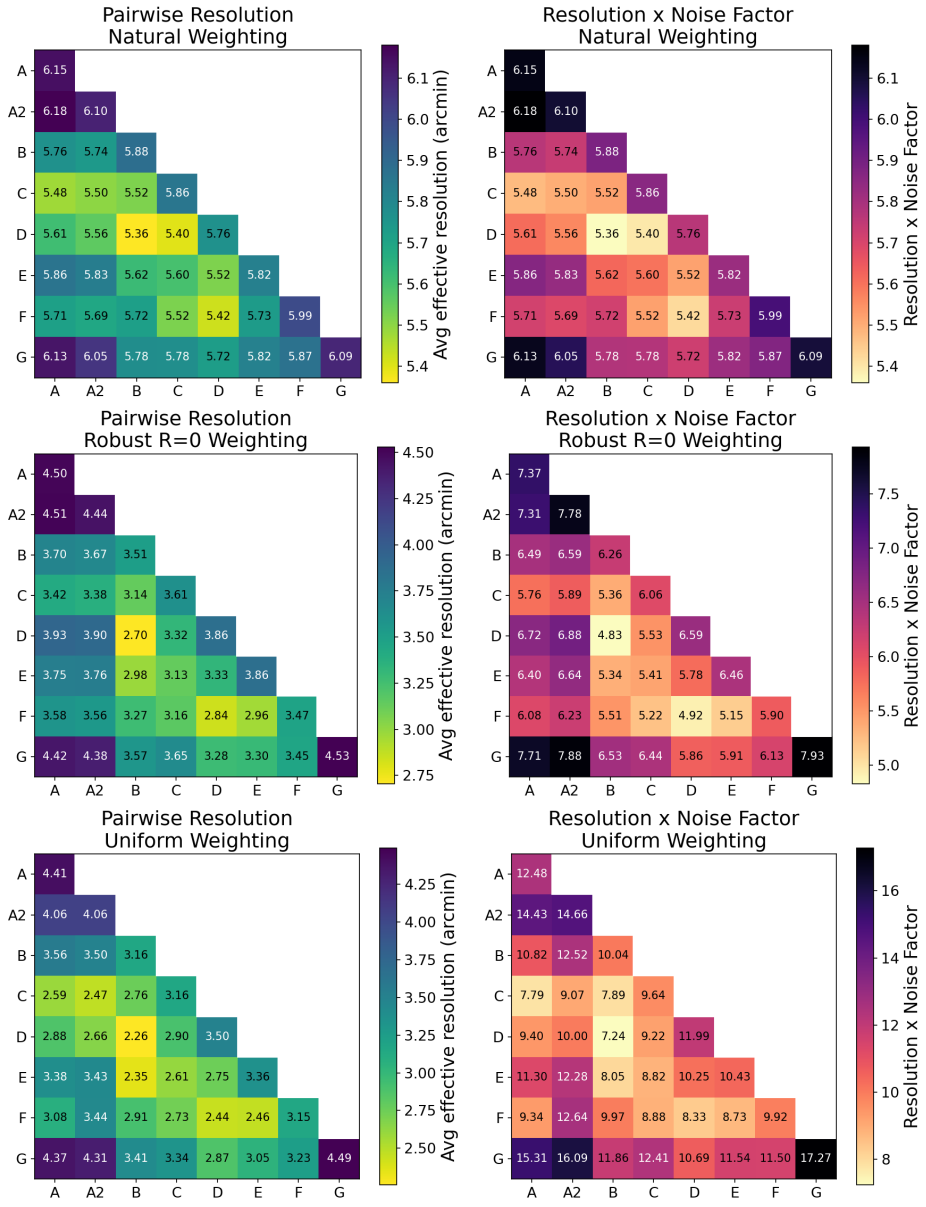


Figure B.1: Resolution and Resolution x Noise Factor pairwise comparisons for natural, robust ($R = 0$), and uniform weighting schemes. B + D emerges as the best candidate in every comparison method.

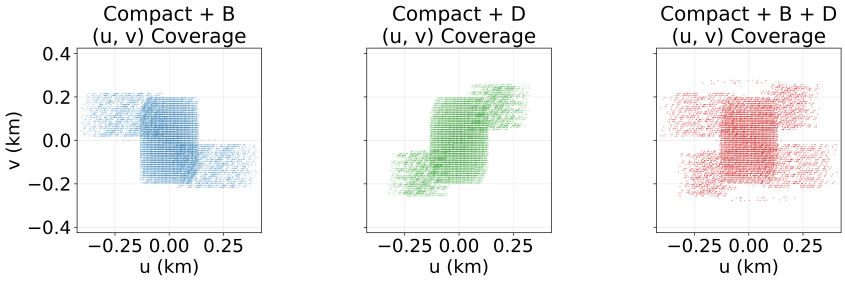


Figure B.2: (u, v) coverage of core + B, core + D, and core + B + D configurations showing the complementary nature of locations B and D.

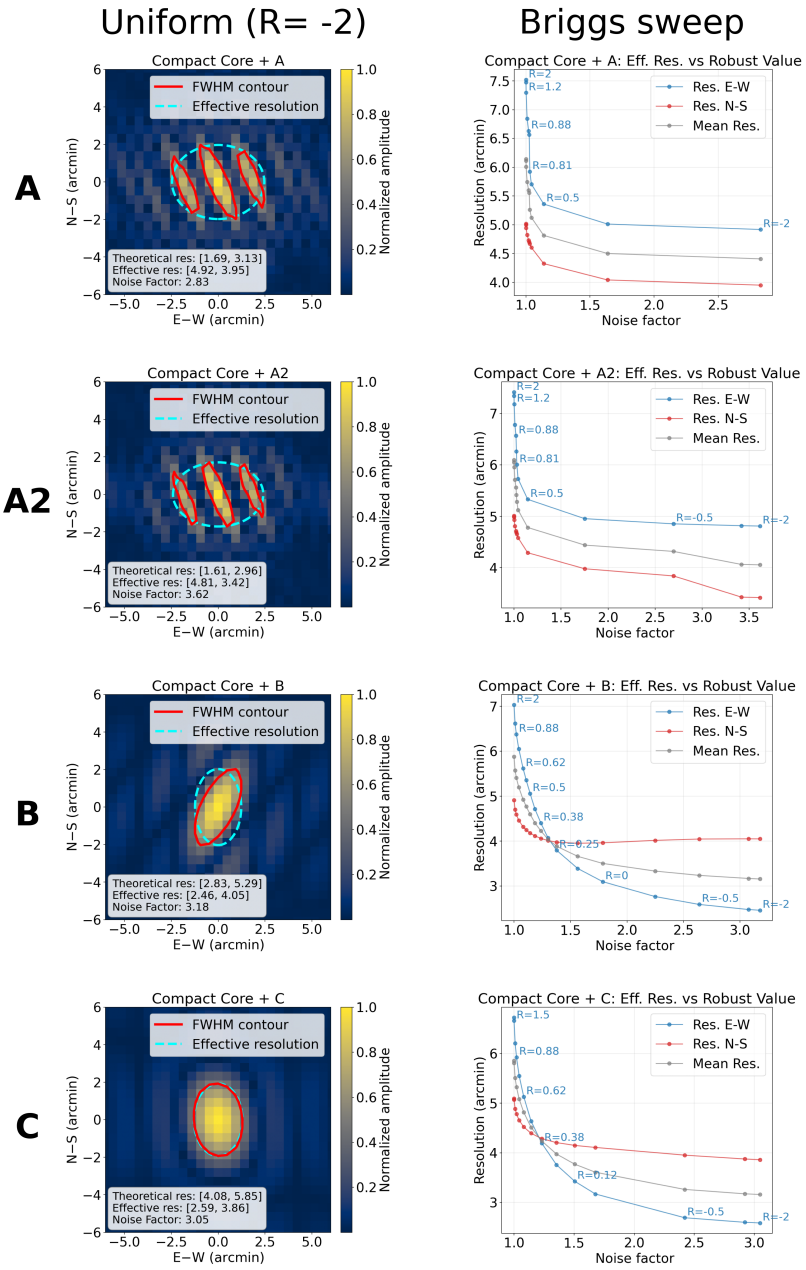


Figure B.3: Beam plots for locations A, A2, B and C showing uniform weighting, and a Briggs robust parameter value sweep.

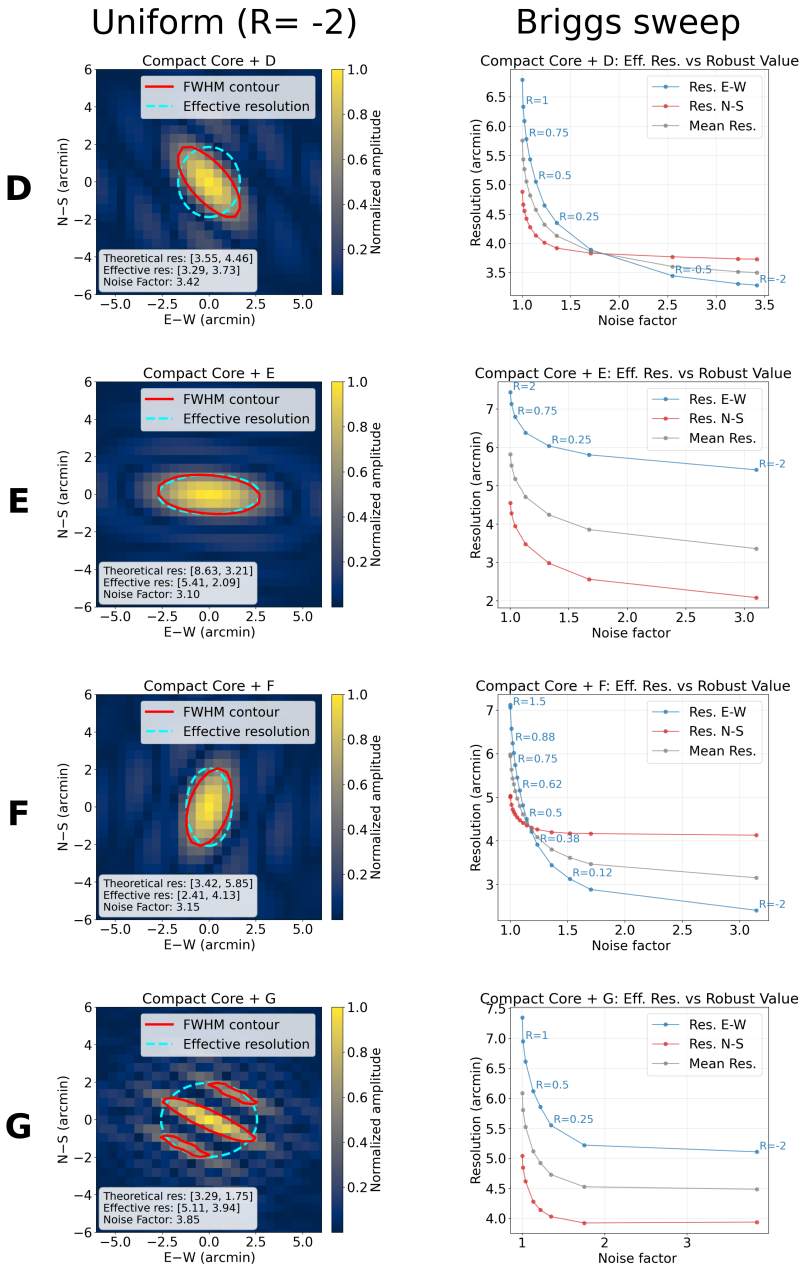


Figure B.4: Beam plots for locations D, E, F, and G showing uniform weighting, and a Briggs robust parameter value sweep.

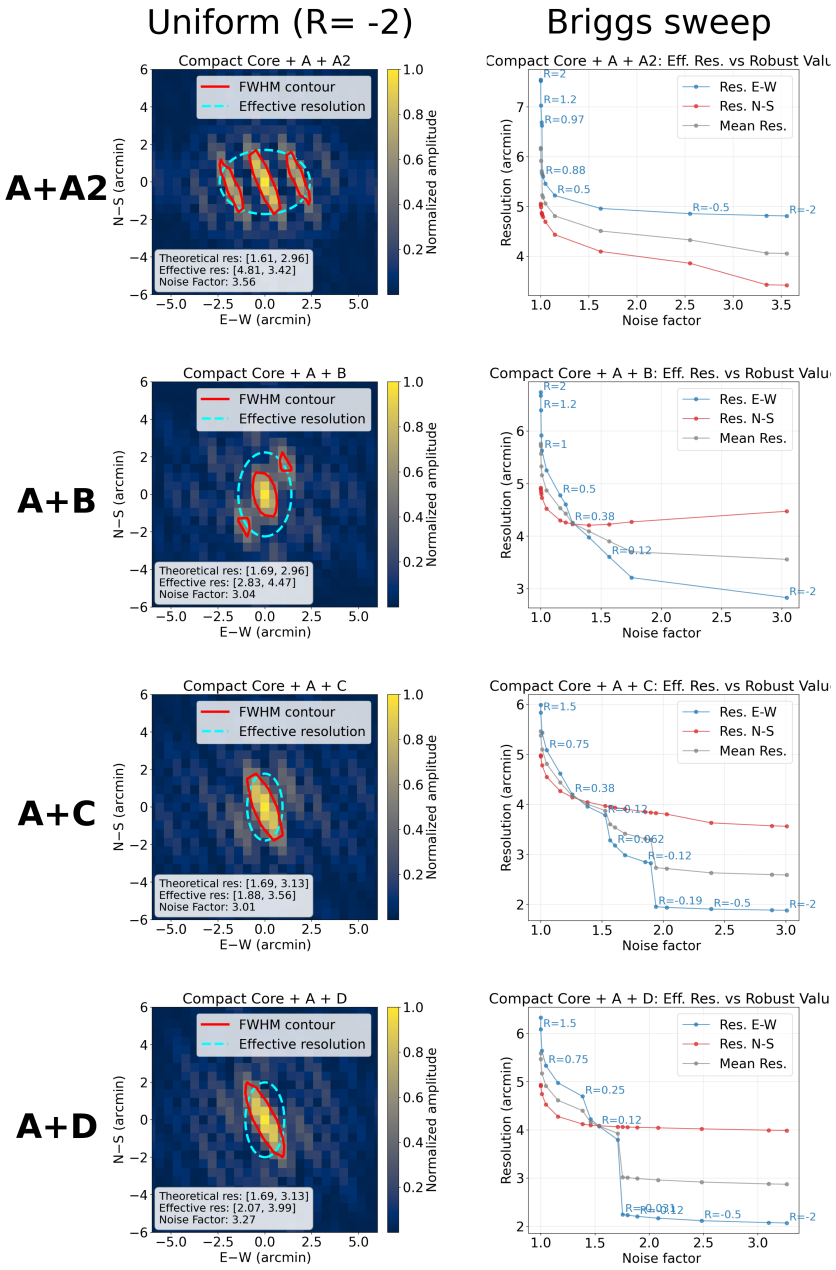


Figure B.5: Beam plots for location pairs A+A2, A+B, A+C, and A+D showing uniform weighting, and a Briggs robust parameter value sweep.

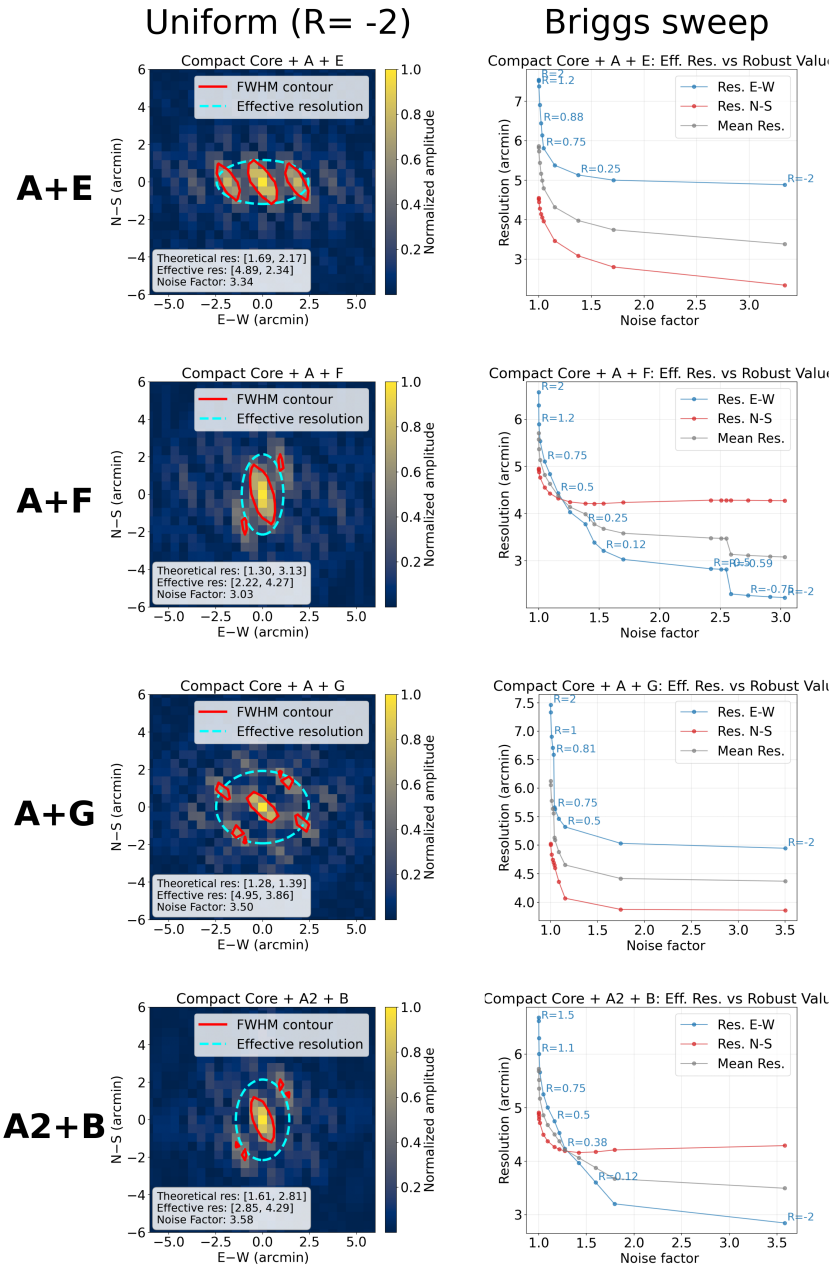


Figure B.6: Beam plots for location pairs A+E, A+F, A+G, and A2+B showing uniform weighting, and a Briggs robust parameter value sweep.

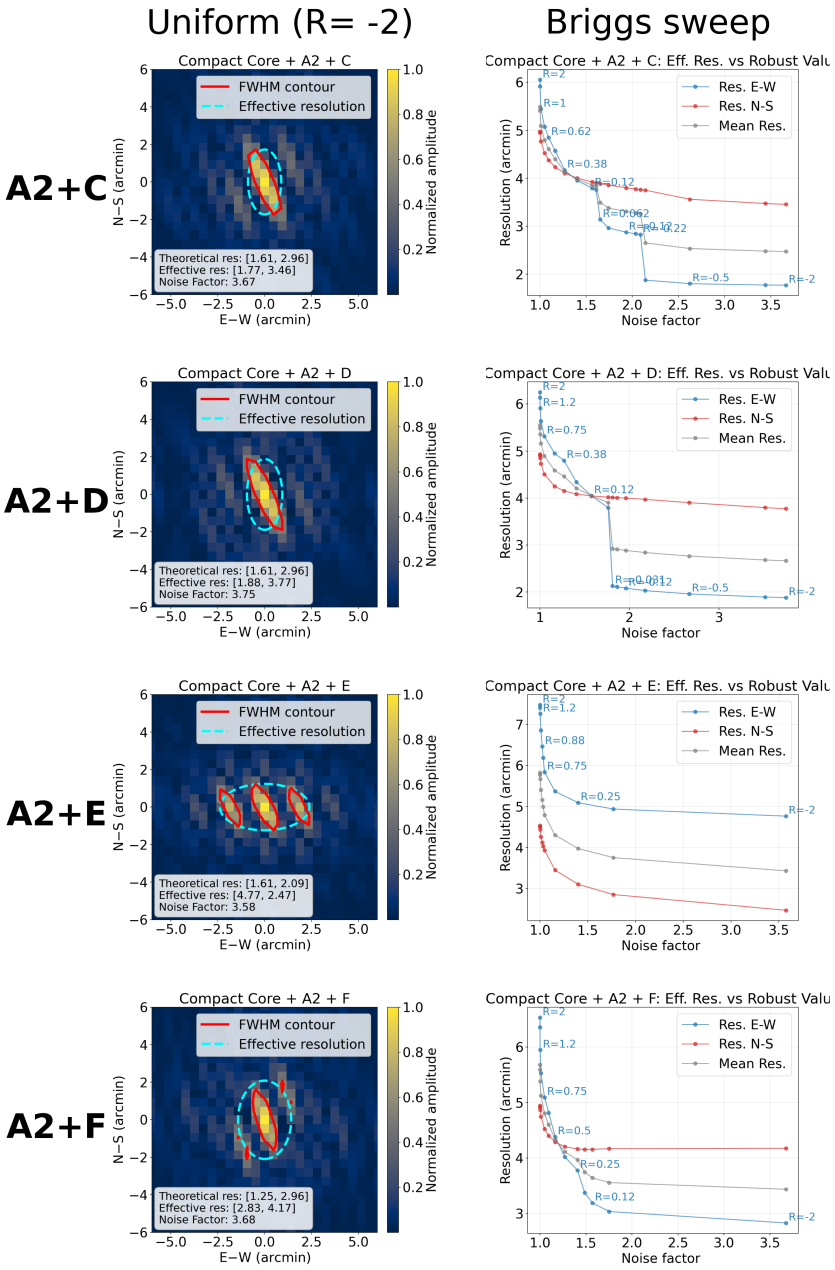


Figure B.7: Beam plots for location pairs A2+C, A2+D, A2+E, and A2+F showing uniform weighting, and a Briggs robust parameter value sweep.

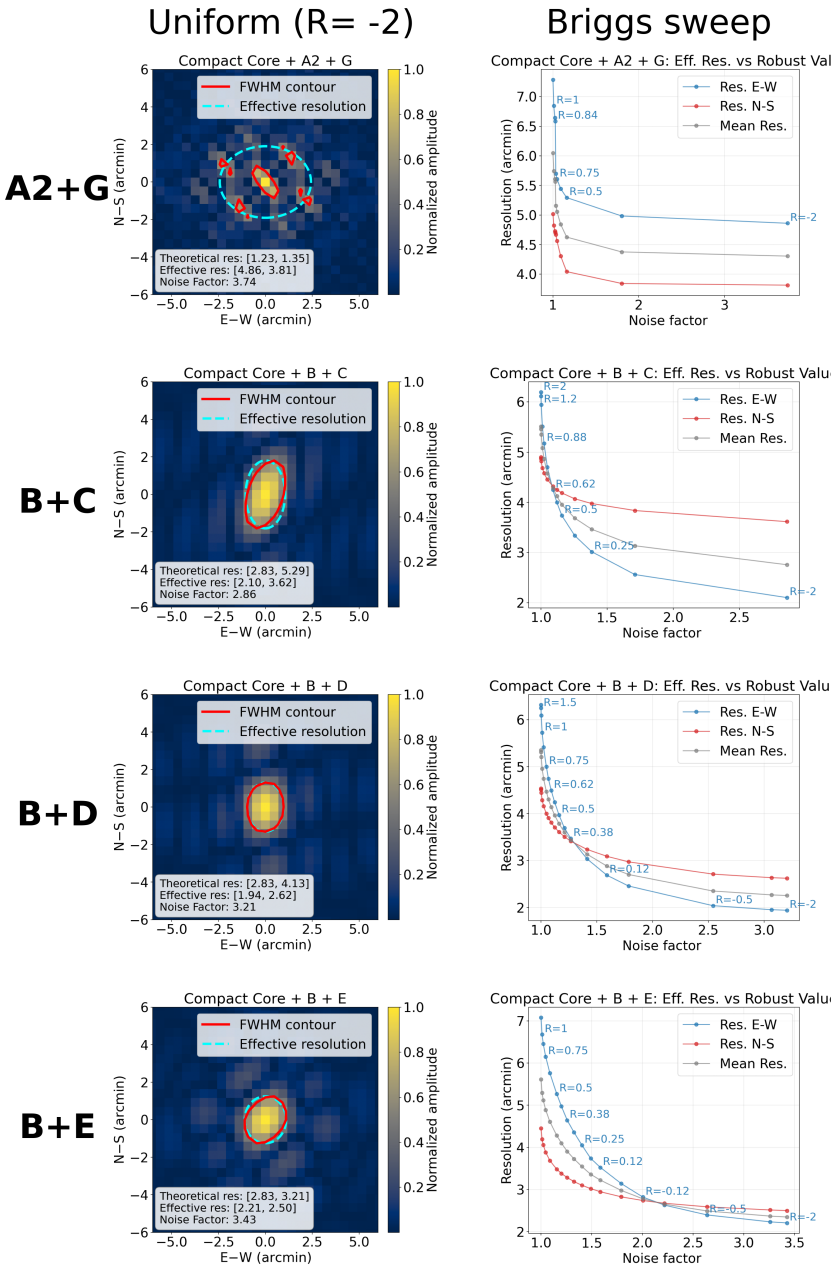


Figure B.8: Beam plots for location pairs A2+G, B+C, B+D, and B+E showing uniform weighting, and a Briggs robust parameter value sweep.

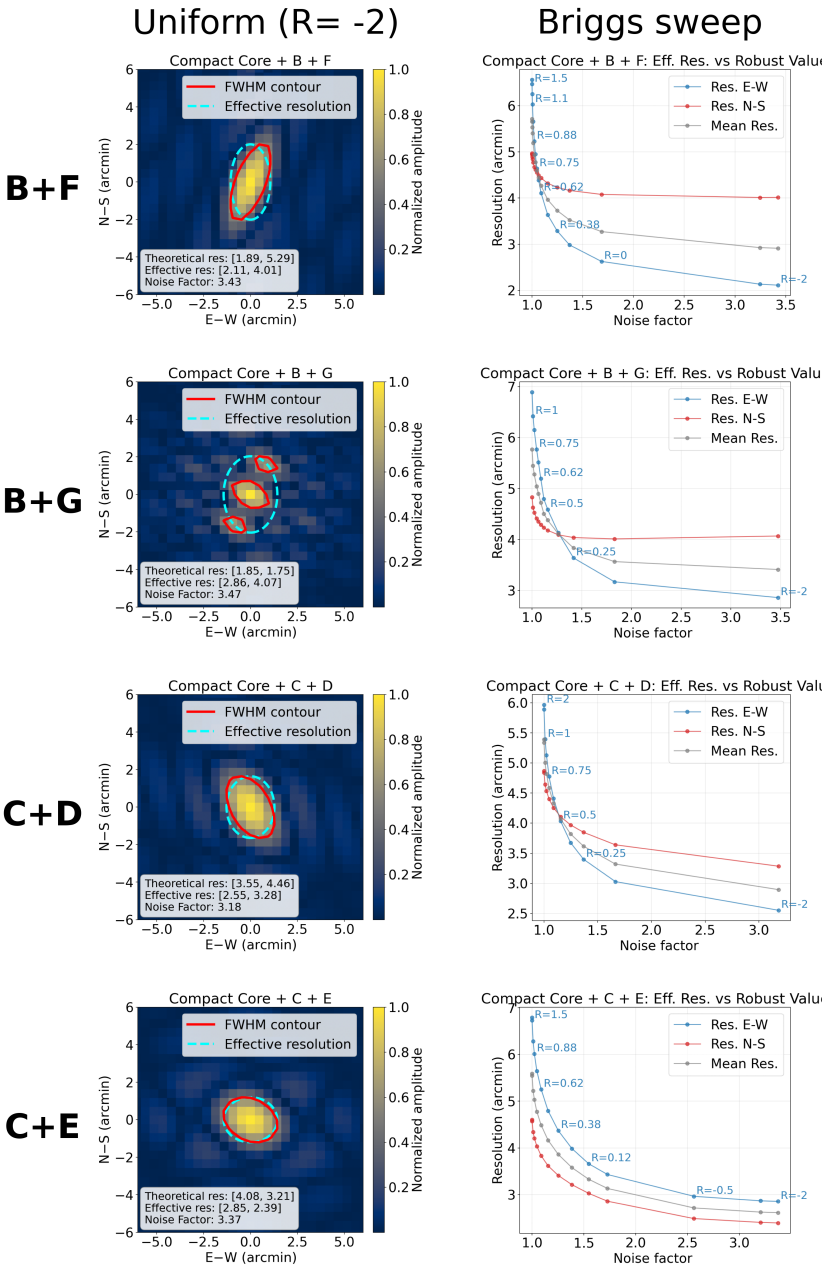


Figure B.9: Beam plots for location pairs B+F, B+G, C+D, and C+E showing uniform weighting, and a Briggs robust parameter value sweep.

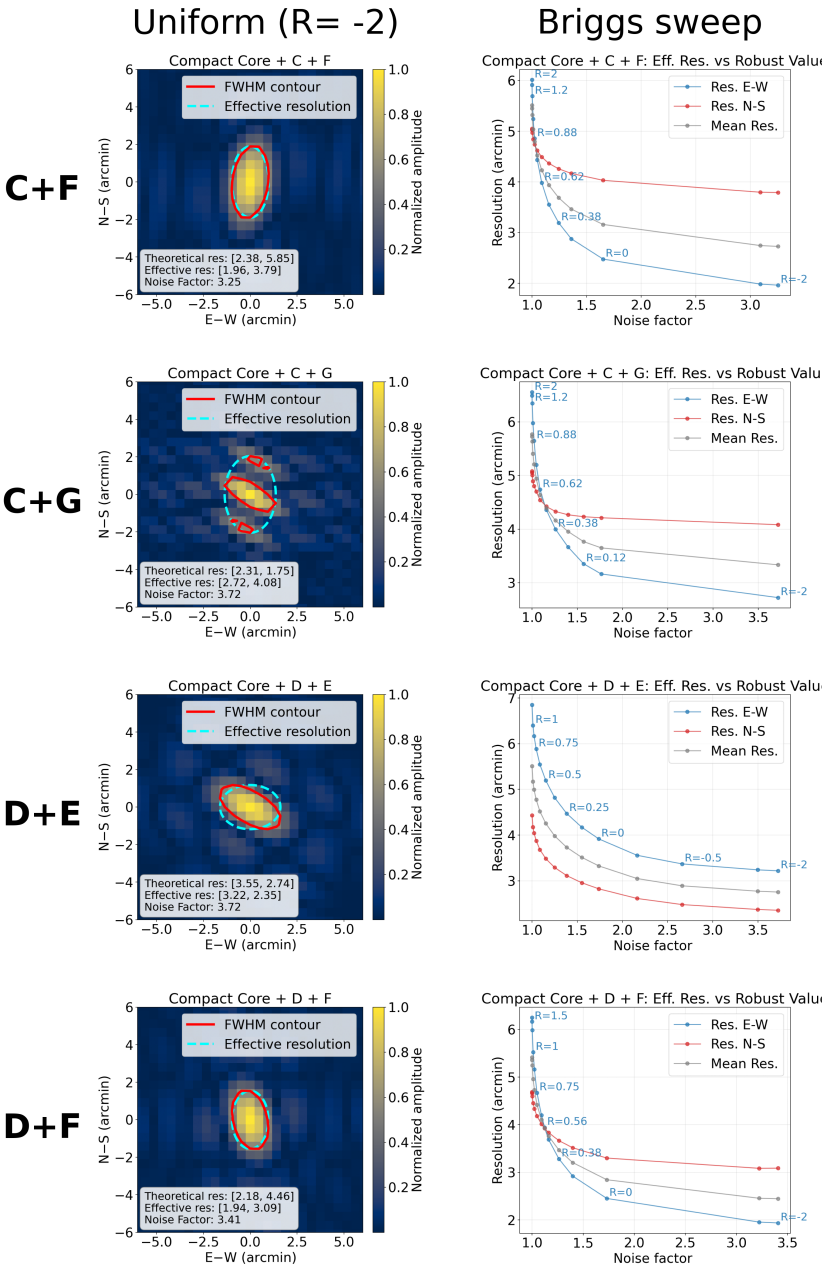


Figure B.10: Beam plots for location pairs C+F, C+G, D+E, and D+F showing uniform weighting, and a Briggs robust parameter value sweep.

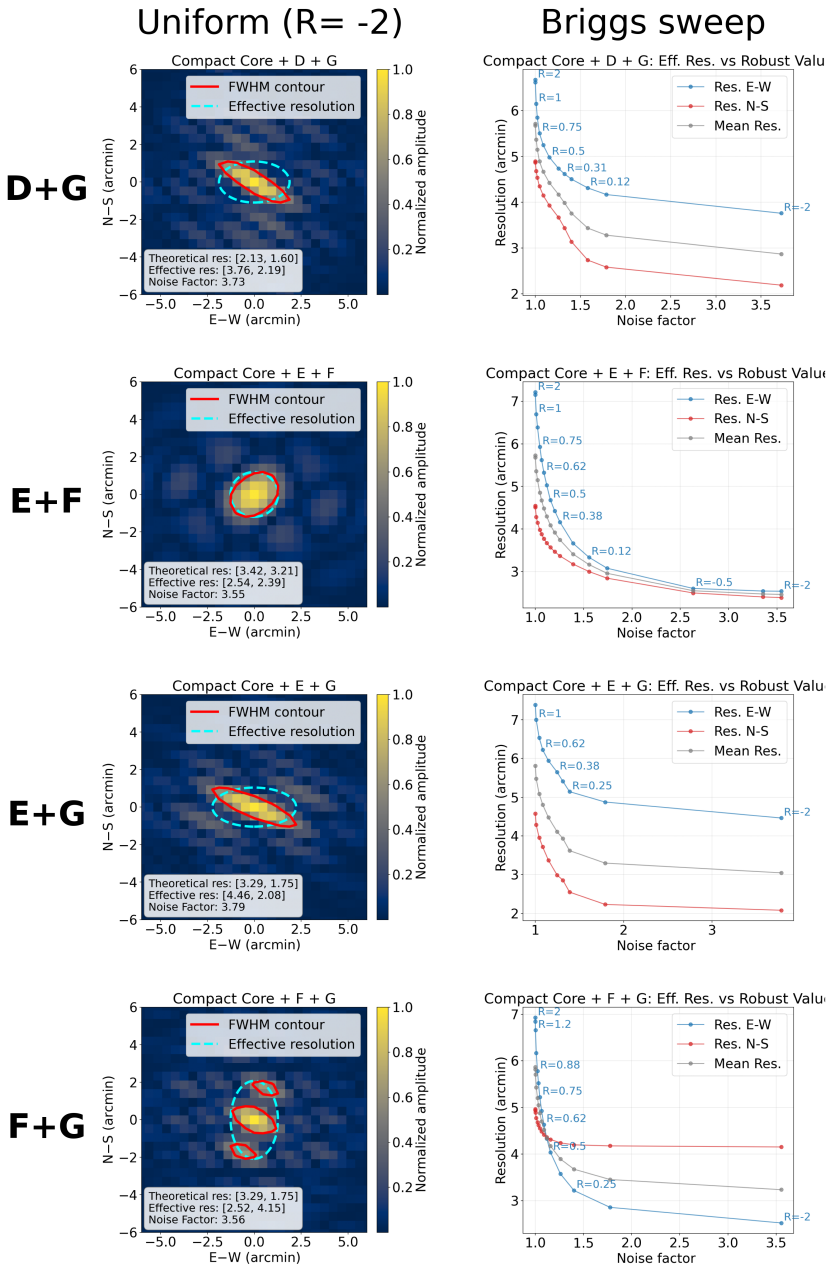


Figure B.11: Beam plots for location pairs D+G, E+F, E+G, and F+G showing uniform weighting, and a Briggs robust parameter value sweep.

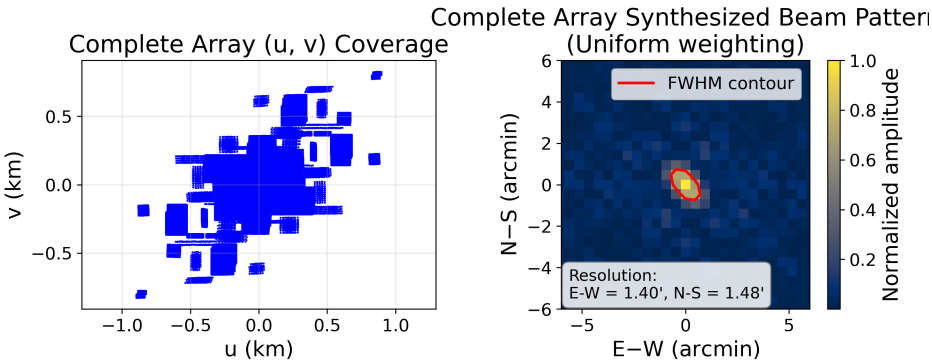


Figure B.12: (u, v) coverage and uniform-weighted beam pattern for the complete array incorporating all extension locations and the core array.

Appendix C

Appendix C: SIGGMA Results

C.1 Complete SIGGMA Tables

C.1.1 SIGGMA H II Regions

A complete set of the reported H II regions produced by SIGGMA which are also visible to CHORD.

Order	Name	Peak [mK]	FWHM [km/s]	S/N	Type
1	G063.164+00.449	694.5 ± 40.6	16.5 ± 1.1	16.9	Known
2	G061.473+00.094	348.6 ± 26.9	30.1 ± 2.7	12.8	Known
3	G059.796+00.241	124.5 ± 11.4	33.2 ± 3.5	10.7	Known
4	G059.321−00.223	124.1 ± 4.3	23.7 ± 1.0	28.5	Known
5	G062.921+00.079	107.2 ± 4.0	23.8 ± 1.0	26.7	Known
6	G061.292−00.331	83.8 ± 17.2	41.2 ± 9.8	5.0	Known
7	G056.725−00.179	78.1 ± 6.5	75.8 ± 7.3	11.8	Candidate
8	G055.567+00.701	75.3 ± 2.8	19.1 ± 0.8	26.6	Candidate
9	G057.541−00.279	66.6 ± 7.3	26.6 ± 3.4	12.9	Known
10	G063.050−00.124	65.3 ± 2.5	24.9 ± 1.1	26.0	Candidate
11	G060.881−00.135	57.5 ± 8.1	28.2 ± 6.0	12.7	Known
12	G057.624+00.410 b	54.6 ± 5.8	74.4 ± 12.1	21.8	Candidate
13	G057.545−00.275	52.4 ± 7.5	16.1 ± 3.8	7.5	Known
14	G056.083−00.176	51.3 ± 6.3	15.6 ± 2.2	8.0	Known
15	G057.624+00.410 b	50.7 ± 11.3	14.5 ± 4.0	9.0	Candidate
16	G064.130−00.475	45.4 ± 3.0	27.9 ± 2.1	14.9	Known
17	G063.052−00.332	41.5 ± 3.6	43.5 ± 4.4	11.5	Known
18	G061.467+00.380	38.9 ± 3.0	20.5 ± 1.8	12.9	Known
19	G067.816+00.511	27.2 ± 4.9	19.1 ± 4.0	5.4	Candidate
20	G061.144−00.891	25.1 ± 4.6	18.4 ± 3.9	5.4	Candidate
21	G059.600+00.316	23.4 ± 3.5	28.6 ± 5.0	6.6	Known
22	G057.715+00.173	19.5 ± 2.6	33.8 ± 7.2	13.3	Candidate
23	G061.053−00.825 b	19.1 ± 3.2	47.9 ± 9.3	5.9	Candidate
24	G057.715+00.173	18.0 ± 2.9	22.7 ± 4.9	10.0	Candidate
25	G057.715+00.173	17.1 ± 1.7	68.7 ± 11.3	16.6	Candidate
26	G062.142−00.144	16.8 ± 2.6	63.9 ± 11.7	6.2	Candidate

Table C.1: Set of SIGGMA H II regions and candidate H II regions visible to CHORD.

C.1.2 SIGGMA Stacked SNR Predictions

A complete set of the predicted SNR levels that CHORD can achieve on the H II regions reported by SIGGMA. Using a common frequency resolution of 5 kHz and each line's native angular resolution.

Order	Name	SNR top 1	SNR top 15	SNR top 60
1	G063.164+00.449	23.55 ± 1.38	90.98 ± 5.32	171.54 ± 10.03
2	G061.473+00.094	9.72 ± 0.75	37.47 ± 2.89	70.73 ± 5.46
3	G059.796+00.241	3.71 ± 0.34	14.26 ± 1.31	26.71 ± 2.45
4	G059.321−00.223	3.69 ± 0.13	14.21 ± 0.49	26.79 ± 0.93
5	G062.921+00.079	3.73 ± 0.14	14.42 ± 0.54	27.14 ± 1.01
6	G061.292−00.331	2.75 ± 0.56	10.60 ± 2.18	19.94 ± 4.09
7	G056.725−00.179	2.64 ± 0.22	10.18 ± 0.85	19.07 ± 1.59
8	G055.567+00.701	2.38 ± 0.09	9.15 ± 0.34	17.12 ± 0.64
9	G057.541−00.279	2.22 ± 0.24	8.55 ± 0.94	16.02 ± 1.76
10	G063.050−00.124	2.33 ± 0.09	8.99 ± 0.34	16.90 ± 0.65
11	G060.881−00.135	1.78 ± 0.25	6.87 ± 0.97	12.92 ± 1.82
12	G057.624+00.410 b	1.75 ± 0.19	6.72 ± 0.71	12.59 ± 1.34
13	G057.545−00.275	1.75 ± 0.25	6.72 ± 0.96	12.59 ± 1.80
14	G056.083−00.176	1.70 ± 0.21	6.54 ± 0.80	12.28 ± 1.51
15	G057.624+00.410 b	1.62 ± 0.36	6.24 ± 1.39	11.69 ± 2.60
16	G064.130−00.475	1.71 ± 0.11	6.60 ± 0.44	12.38 ± 0.82
17	G063.052−00.332	1.51 ± 0.13	5.83 ± 0.51	10.95 ± 0.95
18	G061.467+00.380	1.26 ± 0.10	4.87 ± 0.38	9.16 ± 0.71
19	G067.816+00.511	1.04 ± 0.19	4.00 ± 0.72	7.50 ± 1.35
20	G061.144−00.891	0.87 ± 0.16	3.36 ± 0.62	6.32 ± 1.16
21	G059.600+00.316	0.75 ± 0.11	2.88 ± 0.43	5.41 ± 0.81
22	G057.715+00.173	0.62 ± 0.08	2.40 ± 0.32	4.48 ± 0.60
23	G061.053−00.825 b	0.66 ± 0.11	2.56 ± 0.43	4.80 ± 0.80
24	G057.715+00.173	0.58 ± 0.09	2.21 ± 0.36	4.14 ± 0.67
25	G057.715+00.173	0.55 ± 0.05	2.10 ± 0.21	3.93 ± 0.39
26	G062.142−00.144	0.61 ± 0.09	2.37 ± 0.37	4.45 ± 0.69

Table C.2: Predicted SNR for CHORD observations of SIGGMA H II regions using one observing day, showing the SNR from stacking the highest 1, 15 and 60 stacked lines.

C.1.3 SIGGMA Stacked SNR Predictions with Smoothing

A complete set of the predicted SNR levels which CHORD can achieve on the H II regions reported by SIGGMA, after smoothing to a common velocity resolution of $5\,\mathrm{km\,s^{-1}}$ and a common angular resolution as reported in the table header.

Order	Name	SNR 5′	SNR 6′	SNR 8′
1	G063.164+00.449	85.04 ± 4.97	130.52 ± 7.63	222.38 ± 13.00
2	G061.473+00.094	41.86 ± 3.23	63.49 ± 4.90	105.47 ± 8.14
3	G059.796+00.241	14.70 ± 1.35	22.45 ± 2.06	37.98 ± 3.48
4	G059.321−00.223	14.70 ± 0.51	22.44 ± 0.78	37.79 ± 1.31
5	G062.921+00.079	13.05 ± 0.49	20.08 ± 0.75	34.37 ± 1.28
6	G061.292−00.331	10.04 ± 2.06	15.43 ± 3.17	26.35 ± 5.41
7	G056.725−00.179	9.24 ± 0.77	14.21 ± 1.18	24.57 ± 2.04
8	G055.567+00.701	8.91 ± 0.33	13.70 ± 0.51	23.42 ± 0.87
9	G057.541−00.279	7.92 ± 0.87	12.18 ± 1.34	20.99 ± 2.30
10	G063.050−00.124	7.98 ± 0.31	12.29 ± 0.47	21.11 ± 0.81
11	G060.881−00.135	6.84 ± 0.96	10.47 ± 1.47	17.74 ± 2.50
12	G057.624+00.410 b	6.48 ± 0.69	9.96 ± 1.06	17.05 ± 1.81
13	G057.545−00.275	6.22 ± 0.89	9.57 ± 1.37	16.50 ± 2.36
14	G056.083−00.176	6.09 ± 0.75	9.36 ± 1.15	16.07 ± 1.97
15	G057.624+00.410 b	6.02 ± 1.34	9.25 ± 2.06	15.83 ± 3.53
16	G064.130−00.475	5.62 ± 0.37	8.69 ± 0.57	15.03 ± 0.99
17	G063.052−00.332	5.07 ± 0.44	7.82 ± 0.68	13.48 ± 1.17
18	G061.467+00.380	4.66 ± 0.36	7.16 ± 0.55	12.17 ± 0.94
19	G067.816+00.511	3.40 ± 0.61	5.27 ± 0.95	9.12 ± 1.64
20	G061.144−00.891	3.05 ± 0.56	4.70 ± 0.86	8.09 ± 1.48
21	G059.600+00.316	2.76 ± 0.41	4.23 ± 0.63	7.23 ± 1.08
22	G057.715+00.173	2.31 ± 0.31	3.54 ± 0.47	6.07 ± 0.81
23	G061.053−00.825 b	2.31 ± 0.39	3.57 ± 0.60	6.14 ± 1.03
24	G057.715+00.173	2.13 ± 0.34	3.27 ± 0.53	5.60 ± 0.90
25	G057.715+00.173	2.02 ± 0.20	3.11 ± 0.31	5.32 ± 0.53
26	G062.142−00.144	2.03 ± 0.31	3.13 ± 0.49	5.42 ± 0.84

Table C.3: SNR for CHORD observations of SIGGMA H II regions with a smoothed angular resolution of 5, 6, and 8 arcminutes.