

# **Exploring hydrogen-rich galaxy recovery with CHORD**

Rebecca Ceppas de Castro

McGill University, Department of Physics

Montréal, Canada

April, 2024

A thesis submitted to McGill University in partial fulfillment of the  
requirements of the degree of

Masters of Science, Physics

# Abstract

Galaxy surveys are essential tools in the pursuit of mapping out the distribution of matter in the nearby universe. These maps serve as a direct probe of the physics and cosmology of the early universe. In particular, 21 cm galaxy surveys detect galaxies rich in neutral hydrogen (HI) that emit light at the signature wavelength produced by the hyperfine transition in HI, 21 cm. The Canadian Hydrogen Observatory and Radio-Transient Detector (CHORD) is a next-generation radio interferometer designed to detect fast radio transients, probe fundamental physics, and map out the distribution of matter on cosmological scales. The combination of CHORD’s angular resolution, large field of view, sensitivity, and spectral properties results in its potential to detect up to 10 million individual 21 cm galaxies, a 300-fold improvement compared to current state-of-the-art surveys. Radio interferometry overcomes many of the limitations of single-dish instruments, but it presents a new set of challenges in the form of complex instrumental systematics. This thesis explores and quantifies the impacts of these systematics on the future CHORD 21 cm galaxy survey through an end-to-end simulation pipeline. We use mock galaxy catalogs to simulate CHORD observations of the sky and identify spectral and spatial instrumental effects. The results provide insight into the prospects and complexities of interferometric 21 cm galaxy surveys.

## Abrégé

Les relevés de galaxies sont des outils essentiels pour cartographier la distribution de matière dans l'univers proche. Ces cartes permettent d'explorer et d'étudier les lois de la physique et de la cosmologie de l'univers primordial. En particulier, les relevés des raies à 21 cm des galaxies détectent les galaxies riches en hydrogène neutre (HI) qui émettent des photons à la longueur d'onde de 21cm issus de la transition hyperfine dans HI. L'Observatoire Canadien de l'Hydrogène et le DéTECTeur Radio-Transitoire (CHORD) est un interféromètre radio de nouvelle génération conçu pour détecter les sursauts radios rapides, sonder la physique fondamentale et cartographier la distribution de la matière à l'échelle cosmologique. La combinaison de la résolution angulaire de CHORD, du grand champ de vision, de la sensibilité et des propriétés spectrales permet de détecter jusqu'à 10 millions de galaxies individuelles, soit une amélioration de 300 fois par rapport aux relevés de pointe actuels. L'interférométrie radio dépasse de nombreuses limites des instruments à une seule antenne, mais présente un nouvel ensemble de défis sous la forme d'une systématique instrumentale complexe. Cette thèse explore et quantifie les impacts de cette systématique sur le futur relevé de galaxie de CHORD à travers un pipeline de simulation de bout en bout. Nous utilisons de faux catalogues de galaxies pour simuler les observations CHORD du ciel et identifier les effets spectraux et spatiaux des instruments. Les résultats donnent un aperçu des perspectives et de la complexité des relevés interférométriques des raies à 21 cm des galaxies.

# Acknowledgements

Firstly, I would like to thank my supervisor Adrian Liu, without whom this thesis would not have been possible. His constant generosity in sharing his time, knowledge, and advice has been essential in my time as a graduate student. He has been an incredible mentor, guiding me through the challenges I faced over the past two years, and I am very grateful. I have had the pleasure of working closely with Akanksha Bij, Hans Hopkins, Kristine Spekkens, Olivia Pereira, and others who participated in our bi-weekly galaxy group calls. Your assistance at every stage of this project has kept me excited and motivated. To many of my colleagues in the McGill physics department, especially to the Cosmic Dawn research group, thank you for making my experience these past two years so rich. It has been a joy to experience your genuine excitement towards what we do. I also want to extend a special thanks to Arnab Chakraborty and Seth Siegel for your help and all the insightful comments and suggestions about this work. I owe a huge thank you to my partner, Ben. You have been a steady shoulder to lean on over the years and have kept me sane throughout this journey. Lastly, I want to extend my deepest thanks to my family for their continuous belief in me. You have encouraged me to pursue my ambitions every step of the way, and I wouldn't be here without you. Pai, M  e, Sa, e Du, obrigada por tudo.

# Statement of contributions

The work presented in this thesis is a culmination of collaboration with many people. I will highlight their individual contributions in this statement. The code infrastructure used to simulate the effects of up-channelization in Section 5.2 and throughout this thesis was written by Olivia Pereira. We worked together to adapt her code to fit within the constraints of my pipeline. In Chapter 7, I use packages from the RADIOCOSMOLOGY code as well as code newly developed by myself to simulate the data products. The sky map components for the 21cm unresolved background, the synchrotron emission, and the extra-galactic point sources were created with CORA. The telescope object and beam transfer matrices were simulated with DRIFTSCAN. The visibilities and dirty maps were simulated with DRACO. There have been many contributors to these packages over the years. The mock HI galaxy catalogs, as well as the code used to simulate their spectra were provided by Akanksha Bij. The calibration error distributions we sample from in Section 7.5 were estimated by Bobby Pascua and the code to properly sample errors from these was developed in partnership with Olivia Pereira. The analysis presented in Chapter 8 was performed by myself. I developed the code for the spatial matched filter and CLEAN algorithm and used code from Hans Hopkins for the frequency matched filter in Section 8.3.

# Contents

Abstract . . . . .	i
Abrégé . . . . .	ii
Acknowledgements . . . . .	iii
Statement of contributions . . . . .	iv
List of Figures . . . . .	xiv
List of Tables . . . . .	xv
<b>1 Introduction</b>	<b>1</b>
<b>2 HI galaxies and their applications</b>	<b>4</b>
2.1 HI galaxies . . . . .	4
2.1.1 Busy function . . . . .	8
2.1.2 Stacking . . . . .	9
2.2 Astrophysics and cosmology applications . . . . .	10
<b>3 Status of HI galaxy surveys</b>	<b>13</b>
3.1 Current and past HI surveys . . . . .	13
3.1.1 HIPASS . . . . .	15
3.1.2 ALFALFA . . . . .	15
3.1.3 FAST . . . . .	17

3.2 Future HI surveys: The SKA . . . . .	17
3.2.1 MIGHTEE . . . . .	18
3.2.2 ASKAP: WALLABY and DINGO . . . . .	19
<b>4 The Canadian Hydrogen Observatory and Radio Transient Detector</b>	<b>20</b>
<b>5 Radio Interferometry</b>	<b>24</b>
5.1 Mathematical Description . . . . .	24
5.2 Up-channelization . . . . .	29
<b>6 Map-making</b>	<b>36</b>
6.1 Mathematical Description . . . . .	36
6.2 Maximum Likelihood . . . . .	38
6.3 The $m$ -mode formalism . . . . .	39
6.4 CLEAN . . . . .	43
<b>7 The Pipeline</b>	<b>45</b>
7.1 Sky map simulations . . . . .	45
7.1.1 21cm cosmological unresolved background . . . . .	46
7.1.2 Synchrotron emission from the Milky Way . . . . .	47
7.1.3 Extra-galactic point sources . . . . .	48
7.2 Mock HI galaxy catalogs . . . . .	49
7.3 Instrument set-up: beams . . . . .	54
7.4 Visibility Simulation . . . . .	57
7.5 Noise and Calibration Systematics . . . . .	60
7.5.1 Thermal noise . . . . .	61
7.5.2 Calibration errors . . . . .	62

7.6	Dirty maps . . . . .	64
<b>8</b>	<b>HI galaxy recovery</b>	<b>70</b>
8.1	Spatial matched filter . . . . .	70
8.2	CLEAN . . . . .	77
8.3	Frequency matched filter . . . . .	82
<b>9</b>	<b>Conclusion</b>	<b>90</b>

# List of Figures

2.1	Spin flip transition of an HI atom resulting the emission of light with a rest wavelength of 21cm. . . . .	5
2.2	Busy function fit to HI observed data of two galaxies, NGC 300 and NGC 3351 from HIPASS by the Parker Telescope. The data is shown in white circles while the colored lines are fits of different versions of the Busy function. Note that $B_1$ (blue), the generalised busy function used for this project, presents comparable fits to the other two forms $B_0$ , $B_2$ for both galaxy profiles. Figure taken from [1]. . . . .	8
3.1	Overlapping source detections from ALFALFA, HIPASS, and FASHI shown as a funcion of their position on the sky. Figure taken from [2]. . . . .	16
4.1	The distribution of detected galaxies as a function of redshift. The blue solid line shows the prediction for CHORD while the orange dashed line shows all the ALFALFA detections. CHORD shows a significant increase in source counts at all redshifts and peaks at a higher redshift. The red regions frequencies with challenging RFI (radio frequency interference) contamination. Adapted from [3]. . . . .	21
4.2	CHORD system overview. Image credit: Juan Mena. . . . .	22

5.1	Scales accessible to the CHORD instrument. The purple lines represent the $r_{\perp}$ feature sizes resolved by the pathfinder (light purple) and full-CHORD (dark purple). The green lines represent the $r_{\parallel}$ feature sizes in both the coarse CHORD spectral resolution of 586 kHz (light green) and a resolution about 16 times finer $\sim 36$ kHz (dark green). . . . .	28
5.2	Up-channelization matrix $R(c, u; \nu)$ for a frequency range 1414 – 1421 MHz with $U = 16$ . There is a bright diagonal region illustrating the maximum sensitivities of channels of certain frequencies. There are also features depicting spectral leakage and modulation (see text for explanation). . . . .	31
5.3	Normalization vector for a frequency range 1414 – 1421 with $U = 16$ . The coarse channels are marked by the alternating gray and white background. There is an envelope around each coarse channel, with peak sensitivity in the centers of coarse channels and a significant drop in sensitivity around their edges. . . . .	33
5.4	Two examples illustrating the effects of up-channelization on a Gaussian function and a delta function. The coarse channels are marked by the alternating gray and white background. We show how the placement of the original signal produces different aliasing responses and symmetries in the output signal. . . . .	34
5.5	Similar to Figure 5.4 but with a simulated HI galaxy profile up-channelized with factors $U = 2, 16, 64$ (respectively in purple, orange, and green). The dotted black line shows the continuous profile. . . . .	35

7.1	Unresolved 21cm cosmological emission. (Left) Spatial features at 1420 MHz. (Right) Spectral features on the higher end of the CHORD band for an arbitrary pixel. Unlike in Figures 7.2-7.3, the brightness temperature $T_b$ is given in mK. . . . .	46
7.2	Synchrotron emission from the Milky Way. (Left) Spatial features at 1420 MHz. (Right) Spectral features on the higher end of the CHORD band for an arbitrary pixel. We see the characteristic galactic plane spatial structure coupled with a smooth frequency evolution. . . . .	47
7.3	Extra-galactic point sources signal contribution in the sky. (Left) Spatial features at 1420 MHz. (Right) Spectral features on the higher end of the CHORD band for an arbitrary pixel. The color bar for the brightness temperature $T_b$ has been modified to show a maximum of 1K for visualization. . . . .	49
7.4	Spatial distribution of sources from the ALFALFA constrained mock galaxy catalog used in this thesis. The color bar corresponds to the source's distance away from the observer in Mpc. The sizes of the points correspond to the $M_{\text{HI}}$ of the galaxies. The legend shows representative masses, however, the entire sample does contain a few more massive sources and multiple less massive sources than the ones referenced in the legend. . . . .	50
7.5	A sample of simulated HI galaxy spectra shown over a section of CHORD's frequency band. The variety of shapes and amplitudes correspond to the differences in physical and observational parameters outlined throughout this section and in section 2.1.1. . . . .	54

---

7.6	The left panel shows the distribution of dishes for a $2 \times 2$ array. The middle panel shows the number of baselines of each given length. The right panel shows the simulated synthesized beam at 1418 MHz for this array. The color bar shows the relative sensitivity of the beam as a fraction of the main lobe sensitivity. A scale is shown to provide a sense of the size of the features. . . . .	55
7.7	Same as Figure 7.6 but for the CHORD pathfinder configuration. There are 6 dishes with a 6.3 m separation in the EW direction and 11 dishes with 8.5 m separation in the NS direction. . . . .	57
7.8	Visibilities of a full sky map (left panel) and an HI galaxies only map (right panel) for a single baseline as observed by a pathfinder-like array. The contribution from HI galaxies has been boosted by a factor of 10 in the full sky map for illustration purposes. . . . .	58
7.9	Slices along 3 arbitrary RA values from the full sky visibilities. The emission from HI galaxies in the map has not been boosted. Amongst the smooth spectral dependencies of the dominant signal, the spectral shapes of the galaxies are evident even for faint sources. . . . .	59
7.10	Frequency evolution of noise variance. The orange line shows thermal noise free of up-channelization effects while the purple line accounts for the modulation correction. The bottom panel shows the different in variance for the two cases with the white and gray alternating to represent the coarse channels. . . . .	62
7.11	CHORD estimated calibration error distribution. The left panel shows the amplitude percent error distribution and the right panel shows the phase error distribution in radians. Notably both are non-Gaussian with extended tails. . . . .	64

7.12	The top panel shows the dirty map computed from a simulated observation of an input sky containing only HI galaxies at 1418 MHz. Various features arise in the dirty map due to the synthesized beam structure. . . . .	65
7.13	HI profiles resulting from the pipeline. The top panels show an HI galaxy and a full sky map after up-channelization. The lower panels show the same profiles after the full observation process and dirty map computation. Characteristic features of spectral and spatial aliasing are further discussed in the text. . . . .	66
7.14	Schematic of the complete pipeline. The colors denote which packages are used for each step. CORA, DRIFT, and CAPUT are from the existing CHIME/CHORD infrastructure. CHORD H-GASP: HI Galaxy Simulation Pipeline is the package developed for this thesis. Different empirical constraints and systematics are included for more realistic outputs. The results are analysed using two main algorithms for source detection. . . . .	69
8.1	One-dimensional representations of the dirty beam template $t(x)$ (left) and data $d(x)$ (right) used for the matched filter. The dirty beam was simulated for a pathfinder configuration. The dirty map contains 10 injected galaxies shown in green. . . . .	72
8.2	Matched filter output $y(x)$ for the dirty map. The red points locate pixels where a detection with $\text{SNR} > 1$ was made while the true pixels of the injected galaxies are shown in green. We see both false-positive and false-negative detections. . . . .	74

8.3	Detection SNR as a function of the galaxy brightness and the Declination of the source. The left and right panels show results for noise levels equivalent to one and 200 day of observations respectively. The labels in the figure show the percentage of sources detected with a $5\sigma$ or higher significance. . . . .	75
8.4	Summary statistics for the detections made with the matched filter in dirty maps with $N_{\text{gals}} = 200, 500, 1000$ . The completeness (top panel) is higher for brighter sources and for less crowded maps. The reliability (bottom panel) remains under $\sim 25\%$ for all dirty maps and values of SNR. . . . .	76
8.5	Performance of CLEAN using three synthesized beams. The left panel shows the dirty map with the galaxies circled in white. The right panel shows the detections made with the simulated synthesized beam (black), the Gaussian beam (green), and the Airy beam (purple). . . . .	79
8.6	Three consequent iterations of CLEAN using the simulated synthesized beam, a Gaussian beam, and an Airy beam. Each row show the residuals after the instrument's response to the remaining brightest source is subtracted from the dirty map. . . . .	81
8.7	Examples of the matched filter outputs for three profiles that have been up-channelized with $U = 64$ . The colored bands represent frequency ranges flagged by the algorithm as a detection. The original galaxy profiles are shown in the gray dashed lines. . . . .	83
8.8	HI mass distribution of detected sources for $U = 2, 8, 16, 32, 64$ shown in comparison to the input HI mass distribution (black). Interesting features are described in the text. . . . .	85

8.9 Distribution of HI masses for detections classified as real (solid lines) and false (dashed lines) for $U = 2$ (orange) and $U = 64$ (green). False detections generally have lower HI masses, though that effect is more present for higher $U$ factors. . . . .	86
8.10 Summary statistics for the detections of galaxies as a function of HI mass and $U$ . The top panel shows the completeness percentage, the bottom panel shows the reliability of detections. Samples corresponding to lower $U$ seem to be less complete but more reliable for all $M_{\text{HI}}$ . . . . .	87

# List of Tables

3.1	Details of select past, current, and future HI galaxy surveys as labeled in column 1. Column 2 ( $N_{\text{galaxies}}$ ) gives the number of HI galaxies released in catalogs at the time of writing this thesis. The numbers in parenthesis represent forecasts for future surveys. Column 3 ( $z_{\text{max}}$ ) is the maximum redshift out to which galaxies have been (are predicted to be) detected. Column 4 ( $A_{\text{survey}}$ ) gives the survey areas in square degrees and Column 5 details operation dates, with a + signaling ongoing and future surveys. More information on each survey is given throughout this chapter and in Chapter 4.	14
4.1	Overview of CHORD instrument parameters.	23
7.1	Sample of mock HI galaxy catalog. The properties shown are explained in the text and their values have been rounded for clarity. The full catalog from which these galaxies were taken contains $N_{\text{gals}} = 3,532$ galaxies and was simulated to cover a stripe centered at Dec = 45° of approximately 5° in width.	52

# Chapter 1

## Introduction

With major advances in technology and analysis techniques, radio astronomy has experienced rapid growth in recent years. Single dish radio telescopes like the Five-hundred-meter Aperture Spherical radio Telescope (FAST) [4], the Green Bank Telescope (GBT) [5], and the South Pole Telescope (SPT) [6], and innovative radio interferometers such as the Canadian Hydrogen Intensity Mapping Experiment (CHIME) [7], the Hydrogen Epoch of Reionization Array (HERA) [8], and the Murchison Widefield Array (MWA) [9] have paved the way in this revolution. From precision cosmology, to deep imaging of the galactic plane, and extensive galaxy surveys, the products of this new generation of radio instruments provide fresh insights into fundamental questions about the universe.

Building on the successes and improving on the limitations of existing instruments, the Canadian Hydrogen Observatory and Radio-Transient Detector (CHORD) will exceed current state-of-the-art efforts in mapping out the distribution of matter in the universe. It will help answer questions of fundamental physics and considerably add to the existing population of detected and localized Fast Radio Bursts (FRBs). A pathfinder will operate

prior to the full deployment of CHORD and one of the initial science goals is to perform a survey of hydrogen-rich galaxies (HI or 21 cm galaxies hereafter) in the nearby universe. Galaxy surveys are essential in the pursuit of understanding the distribution of matter and the underlying physical processes governing the formation and evolution of structures in the universe.

As a complex instrument depending on the precise and consistent operation of its multiple parts, CHORD’s systematics must be well understood. An incomplete awareness of these would hamper the instruments performance. In terms of the HI galaxy survey, this would lead to missed or erroneous detections and potential hidden biases in the resulting HI galaxy catalog. This thesis consists of an in-depth exploration of the systematics of the CHORD pathfinder, specifically relating to the HI galaxy survey. It provides intuition into the importance of various instrumental parameters and depicts the consequences of having incorrect assumptions about them. The work done in understanding the CHORD pathfinder will benefit not only the performance of full CHORD but will eventually pave the way for future instruments like the Square Kilometer Array (SKA).

For this purpose, we have built an end-to-end pipeline to simulate HI galaxy observations with CHORD. The pipeline includes semi-realistic simulations of the sky and of HI galaxy populations. CHORD’s response to the sky is simulated in detail so that the product of the pipeline is a high-fidelity prediction of what a CHORD observation will look like. We include various systematics such as spectral and spatial aliasing effects. We then test a number of tools for HI galaxy recovery to understand their strengths and limitations in the face of the complex systematics. This thesis is organized in the following way.

- In Chapter 2, I introduce HI galaxies and the astrophysical and cosmological applications of HI galaxy surveys. I describe peculiarities in their observations and how they can be modelled.

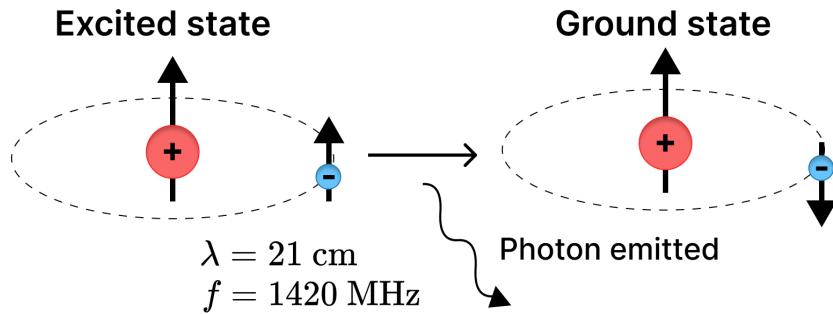
- In Chapter 3 I detail the current status of HI galaxy surveys. I provide a brief history of this field, highlighting the contributions of current state-of-the-art experiments and their future counterparts.
- Chapter 4 gives an introduction to the Canadian Hydrogen Observatory and Radio Transient Detector (CHORD). I dive into its instrumental features and capabilities and provide an order of magnitude comparison to the current best efforts in HI galaxy surveys.
- In Chapter 5 I discuss the basics of interferometry. I give a brief mathematical description and provide a comparison to more traditional single-dish instruments. Finally, I describe the essential computational technique of up-channelization and illustrate its significance to the science case at hand.
- Chapter 6 provides a mathematical framework for map-making and introduces the  $m$ -mode formalism used in the pipeline. I also describe the commonly used CLEAN algorithm within this context.
- Chapter 7 gives an overview of the elements of the pipeline and its step-by-step functioning. It includes input map simulations, beam calculations, and the mock observations. The systematics are portrayed further.
- In Chapter 8 I explore the use of CLEAN and matched filters for detecting HI galaxies from simulated CHORD data products. The impact of the systematics on the performance of these tools is discussed.
- Finally, the conclusions of this thesis are given in Chapter 9.

# Chapter 2

## HI galaxies and their applications

### 2.1 HI galaxies

Hydrogen is the most abundant element in the universe; it is the foundation of many astrophysical processes at multiple stages and scales through cosmic time. In particular, the multiple phases of hydrogen can be linked to essential steps in galaxy formation and evolution. In their formation process, galaxies accumulate ionized hydrogen (HII) from the intergalactic medium (IGM), circumgalactic medium (CGM) and mergers. Within galaxies, physical processes first convert HII to neutral hydrogen (HI) and then to molecular hydrogen ( $H_2$ ) which fuels star formation [10]. HI is traced by the 21 cm spectral line, emitted when the single electron in the hydrogen atom undergoes a hyperfine transition, a spin flip (see Figure 2.1). Even though this is a forbidden transition with a very low probability per unit time of  $2.6 \times 10^{-15} \text{ s}^{-1}$ , the huge abundance of HI makes the observation not only possible but extremely useful for tracing matter across the universe [11]. The existence of this emission line was first predicted in 1945 by Dutch astronomer Hendrick van de Hulst and first observed by Ewen and Purcell in 1951 [12].



**Figure 2.1:** Spin flip transition of an HI atom resulting in the emission of light with a rest wavelength of 21cm.

The 21 cm emission line provides an alternative avenue for observing galaxies compared to traditional optical surveys. All surveys have underlying biases and so a multi-faceted approach to galaxy search will lead to a more complete understanding of the galaxy population and its underlying properties. In a comparison between optical and HI surveys, West et al. (2010) [13] found that the latter often contain bluer, fainter and less massive galaxies. HI-selected samples are also more likely to contain lower surface brightness sources and dwarf galaxies. A complete census of these is essential for constraining models of galaxy formation and evolution. HI observations are strongly biased against early-type galaxies as those are often quenched or low star formation rate (SFR) systems with low reservoirs of neutral gas [14]. Large samples of red-sequence elliptical galaxies are therefore more likely to be found in optical surveys.

HI observations contain a wealth of information about galaxy morphology and spatial structure that may not be available through other means. For instance, neutral hydrogen extends further out than other disk tracers, making HI observations crucial for mapping the outskirts of galaxies and determining their density profiles and sizes [12]. These detailed studies of spatial structure within galaxies are often done through targeted observations of

individual sources. Experiments designed to conduct galaxy surveys do not typically have enough angular resolution to spatially resolve a large number of galaxies (see Chapter 3). The focus of these observations shifts away from the morphology of the galaxies and towards determining their redshift and other physical properties such as their mass and internal kinematics.

Radio telescopes perform measurements on several frequency channels and, given that we know the rest frequency of the 21 cm emission line, the observed frequency becomes a direct probe for the cosmological redshift  $z$  of the source through

$$\nu_{\text{obs}} = \frac{\nu_{21}}{1+z}. \quad (2.1)$$

Here  $\nu_{\text{obs}}$  is the observed frequency and  $\nu_{21}$  is the rest frequency of the 21 cm signal (1420 MHz).

In addition, emission from galaxies is not generally constrained to a single frequency channel and is instead spread over multiple channels forming a spectrum or profile. The spread in the HI profile is caused by the additional red-shifting or blue-shifting of light as the hydrogen gas in the galaxy travels towards or away from the observer relative to its bulk recessional velocity. There are two main categories for HI galaxy profiles. The first is a double-horn profile often observed in edge-on and spiral galaxies where the rotation of the gas is coherent and along our line-of-sight. This causes an increase in the signal observed at frequencies corresponding to the characteristic velocities of the HI gas following the relationship

$$\nu_{\text{obs}} = \nu_{21} \left( 1 - \frac{v}{c} \right). \quad (2.2)$$

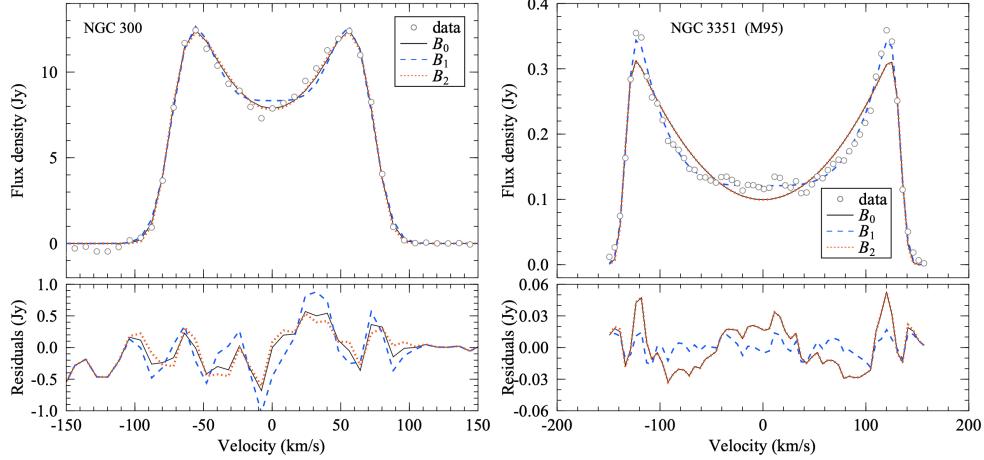
Where  $v/c$  is the fraction of the speed of the gas in the direction of the observer's line-of-sight relative to the speed of light. A fast rotating galaxy, typically also a massive galaxy, will therefore have a wider HI profile when compared to a slower rotating, less massive galaxy.

The double-horn shape is also dependent on the inclination of the galaxy as if the rotation happens mostly on a plane perpendicular to the line-of-sight, no additional red-shifting will occur and these effects will not be observed in the same way.

The second type of profile is a single-horn or Gaussian-like profile. These are observed in elliptical galaxies and face-on galaxies, where most of the motion along the line-of-sight is caused by random thermal motions and dispersion. The virial theorem establishes a connection between a galaxy's velocity dispersion and its total mass, again resulting in wider profiles corresponding to more massive sources. Due to the direct relationship between redshifted frequencies and the internal kinematics of the galaxies, HI profiles can be depicted both as a function of frequency and as a function of velocity given the conversion in Equation 2.2. Figure 2.2 shows two examples of HI profiles for galaxies NGC 300 and NGC 3351 from the HI Parkes All Sky Survey (HIPASS) [1]. These measurements have been centred at zero by accounting for the cosmological redshift, or bulk recessional velocity of the galaxies.

As described above, HI profiles give us insights into the physical properties and kinematics of galaxies as well as give us information about less intrinsic properties, such as the inclination angle of the observation. Integrating the area under the profile results in an estimate of the total HI mass of the galaxy ( $M_{\text{HI}}$ ) so that, in general, broader and taller profiles correspond to galaxies with higher HI contents compared to thinner, shorter ones. Similarly, the width of the profile containing 50% of the emission ( $W_{50}$ ) roughly corresponds to the radial velocity range of the gas in the galaxy. HI mass measurements can be used to infer other properties such as morphological type, HI size, and gas accretion history. Some of these properties have tight and well understood correlations [15], others require more careful modeling and analysis to establish. Nelson (2008) [16] performed an in-depth study to extract crucial parameters from 21 cm line profiles including the HI gas distribution, characteristics of the rotation curve, and asymmetries. They show that by combining two Gaussian curves with

independent widths  $\sigma$  and heights, a variety of profiles can be fit and a classification scheme for physical properties can be created based on these parameters.



**Figure 2.2:** Busy function fit to HI observed data of two galaxies, NGC 300 and NGC 3351 from HIPASS by the Parker Telescope. The data is shown in white circles while the colored lines are fits of different versions of the Busy function. Note that  $B_1$  (blue), the generalised busy function used for this project, presents comparable fits to the other two forms  $B_0$ ,  $B_2$  for both galaxy profiles. Figure taken from [1].

### 2.1.1 Busy function

Parametrizing HI profiles is essential if we want to be able to categorize galaxies and extract information in a coherent and reliable manner. A common and slightly more sophisticated model than the one used in Nelson (2018) is the Busy function, an analytic continuous function with enough flexibility to fit a multitude of HI profiles to good accuracy [1]. The Busy function is a simple combination of two error functions and a polynomial with the general form

$$B_1(x) = \frac{a}{4} \times (\text{erf}[b_1\{w + x - x_e\}] + 1) \times (\text{erf}[b_2\{w - x + x_e\}] + 1) \times (c|x - x_p|^n + 1). \quad (2.3)$$

Here  $x$  is the radial velocity or spectral axis. A description of each of the eight free parameters is given below.

- $a$ : the overall amplitude scaling of the profile
- $w$ : half-width of the profile, equivalent to  $W_{50}$
- $b_1, b_2$ : steepness of each of the 2 error functions
- $c$ : emphasis on parabola, determines the amplitude of the central trough
- $x_e$ : offset of the error functions
- $x_p$ : offset of the parabola to account for centering
- $n$ : degree of the polynomial

This generalized version of the Busy function is more flexible than its simplest version, providing better fits for asymmetrical profiles with different peak heights and shapes in exchange for very little added complexity. Although negative values for these parameters are mathematically allowed, the resulting profiles are not good fits to any galaxy profiles and so are typically not considered. Figure 2.2 shows the Busy function fit to two HIPASS galaxies, in particular, the blue line shows the fit by the generalized Busy function in Equation 2.3 with  $n = 4$ . Generally, the fit is good and follows the shape reasonably well for both galaxies.

### 2.1.2 Stacking

In addition to direct detections of individual HI galaxies, a technique called spectral stacking can be used to increase the signal-to-noise ratio (SNR) of galaxies whose emissions fall below the significance cut-off. In this method, profiles from multiple non-detection sources within a chosen redshift range are aligned and a weighted mean is taken on each frequency channel [17]. Stacking is useful for boosting the SNR of detections, especially at higher redshifts where statistically significant direct detections become more challenging. Fabello (2012) [18]

showed that the SNR increases by  $\sqrt{N}$  for  $N$  galaxies only up to a certain point ( $N \sim 300$  in their case). After that point, adding additional galaxies to the stack has decreasing returns due to two main causes: non-Gaussian noise dominating when averaging large samples, and confusion added by surrounding sources at high redshifts when the beam size is large [19].

While information about individual galaxies is lost in the process of spectral stacking, the average properties can still be estimated. For example, Chowdhury et al. (2020) [20] measured the average HI mass of star-forming galaxies at an average redshift of 1 by stacking  $\sim 7,600$  galaxies observed by the upgraded Giant Metrewave Radio Telescope (uGMRT) within  $0.74 < z < 1.45$ . They achieved a  $4.5\sigma$  detection and estimated an average  $M_{HI} = 1.19 \pm 0.26 \times 10^{10} M_\odot$ . With their results, they reach interesting conclusions about the evolution of SFR in galaxies, stating that the HI content found is only enough to sustain star formation for 1-2 billion years and so star formation for galaxies at  $z < 1$  must be fueled by new accreted neutral hydrogen gas. Delhaize et al. (2013) [21] performed a similar study with the purpose of probing the evolution (or lack thereof) of the cosmic HI density  $\Omega_{HI}$ . They split over 3,000 galaxies observed by the Parkes radio telescope and 15,000 HIPASS galaxies into two redshift bins and then stacked them separately to estimate  $M_{HI}$  and therefore  $\Omega_{HI}$  for these two bins. They found no significant evolution of this quantity over the period probed.

## 2.2 Astrophysics and cosmology applications

As illustrated thus far, observations of HI galaxies have numerous astrophysical and cosmological applications. The study of large populations or samples of HI galaxies also provides valuable insights into larger scale processes and structures and give us a window into evolutionary properties of galaxies. For instance, in the late 1970s, Tully and Fisher

found a strong correlation between the HI profile width  $W_{50}$  of spiral galaxies and their absolute magnitude  $M$  [22] given by

$$M = -a \log(W_{50}) - b, \quad (2.4)$$

where  $a, b$  are constants.

The absolute magnitude is directly related to the distance to the galaxy provided we know the apparent magnitude through other observations. The usefulness of this relation comes from the fact that  $W_{50}$  is a distance-independent measurement and so the Tully-Fisher relation is used as an unbiased extragalactic distance tool. In the 1980s and 1990s, the relation became widely used for estimations of the Hubble constant and for mapping out peculiar velocities of galaxies to better understand the expansion rate of the universe [12].

Another interesting application pertains to the contents, distribution, and cosmic evolution of different components of the total energy budget. The HI mass function (HIMF) catalogs the number density of HI galaxies in the universe as a function of their  $M_{\text{HI}}$  [23]. The HIMF can be used to estimate the cosmological mass density of neutral hydrogen  $\Omega_{\text{HI}}$ . At intermediate redshifts,  $3 < z < 1.6$ , Ly- $\alpha$  systems are used to trace  $\Omega_{\text{HI}}$  but at  $z < 1.6$  their abundance is significantly lower and their emission falls in the ultraviolet regime, making it difficult to detect with ground-base telescopes. We require HI galaxy surveys with large samples in the nearby universe to perform accurate measurements of the local HIMF [24]. The HIMF is often parameterized with a Schechter function: a power law increasing towards lower  $M_{\text{HI}}$  with a faint-end slope  $\alpha$ , a turning-point knee-mass  $M_*$  and an overall normalization factor  $\phi_*$  with an exponential decline at high  $M_{\text{HI}}$  [25]. The functional form of this parameterization is

$$\phi(M_{\text{HI}}) = \ln 10 \phi_* \left( \frac{M_{\text{HI}}}{M_*} \right)^{\alpha+1} e^{-\left( \frac{M_{\text{HI}}}{M_*} \right)}. \quad (2.5)$$

The HIMF, particularly its faint-end, is of great interest to galaxy formation models as it provides a measure of the abundance of low-mass objects. This is crucial for constraining models of hierarchical structure formation, which predict that low-mass objects are formed first and eventually merge or clump with each other giving rise to more massive objects. The models require very specific fractions of the total matter in the universe to be found within low-mass and high-mass systems as the universe evolves [26]. A robust measurement of the HIMF across the whole mass range is therefore a direct way to observationally verify or question the assumptions of these models [24]. Measurements from HI surveys like the Arecibo Legacy Fast ALFA survey (ALFALFA) and HIPASS (see Chapter 3.1) have significantly decreased the uncertainties in the faint-end slope  $\alpha$  of the HIMF [27]. With new HI surveys on the horizon, it is expected that our knowledge of this quantity and thus of  $\Omega_{\text{HI}}$  will improve, leading to tighter constraints on our structure formation models.

HI galaxy surveys also provide a clean method for probing fundamental properties of dark energy. With enough galaxies in a large cosmological volume, a galaxy power spectrum can be estimated and the baryonic acoustic oscillations (BAOs) can be measured. The BAOs can act as a standard ruler, serving as a reliable way to estimate distances and sizes in the universe [28, 29, 30]. Then, the measured BAO scales can be compared with those predicted by our standard models to constrain our understanding of the expansion rate of the universe, and thus of dark energy. Abdalla and Rawlings [31] estimated that for an instrument like the Square Kilometre Array (SKA) with large bandwidth and a wide field of view, the constraints on the dark energy parameter  $w = p/\rho$  would be of order  $\delta w \approx 0.01$ .

# Chapter 3

## Status of HI galaxy surveys

Over the years, HI galaxy surveys have transitioned from targeting a few individual HI rich galaxies to amounting large samples of sources with varying masses and properties over a range of redshifts. With new technologies being designed and built for radio telescopes and interferometers, HI galaxy surveys continue to improve and provide more detail into fundamental astrophysical and cosmological questions. Table 3.1 gives an overview of some of the important HI surveys that have existed in the past, that are currently underway, and the ones that are planned for the future.

### 3.1 Current and past HI surveys

There have been multiple projects and experiments aimed at observing objects emitting at a 21 cm wavelength. In the past, there were a series of targeted HI galaxy observations, particularly for nearby  $z = 0$  galaxies. The Faint Irregular Galaxies GMRT Survey (FIGGs) used the Giant Metrewave Radio Telescope to observe very faint nearby dwarf galaxies that are often undetected by optical surveys [32]. Similarly, the Westerbork HI survey of spiral and

Survey	$N_{\text{gal}}$ ( $N_{\text{gal}}^{\text{predicted}}$ )	$z_{\text{max}}$	$A_{\text{survey}}$ (deg $^2$ )	Operation dates
HIPASS	5,317	0.06	29,300	1997-2001
ALFALFA	31,500	0.06	7,000	2005-2011
FAST	41,741 (100,000)	0.09	21,700	2020+
CHORD	(1,000,000)	0.40	20,700	2024+
MIGHTEE	276 (2,750)	0.58	32	2018+
WALLABY	210,000 (600)	0.26	30,900	2024+
DINGO	(60,000)	0.43	120	2022+
SKA	(1,000,000,000)	2.00	30,900	future

**Table 3.1:** Details of select past, current, and future HI galaxy surveys as labeled in column 1. Column 2 ( $N_{\text{galaxies}}$ ) gives the number of HI galaxies released in catalogs at the time of writing this thesis. The numbers in parenthesis represent forecasts for future surveys. Column 3 ( $z_{\text{max}}$ ) is the maximum redshift out to which galaxies have been (are predicted to be) detected. Column 4 ( $A_{\text{survey}}$ ) gives the survey areas in square degrees and Column 5 details operation dates, with a + signaling ongoing and future surveys. More information on each survey is given throughout this chapter and in Chapter 4.

irregular galaxies (WHISP) and the Survey of HI in Extremely Low-mass Dwarfs (SHIELD) focus on a small sample of carefully selected galaxies with the purpose of patching gaps in optical and infrared surveys [33, 34].

Blind HI surveys, on the other hand, are untargeted and observe large volumes of the sky in search of HI galaxies within a range of HI masses, sizes, and redshifts. The Arecibo HI Strip Survey (AHISS) and the Arecibo Dual Beam Survey (ADBS) were two of the first truly blind HI surveys [23, 35]. Combined, these blind surveys detected a total of 326 HI galaxies out to  $z \sim 0.025$ . From this sample, a significant fraction were extremely low surface brightness dwarf galaxies, once again providing support for the need of blind HI surveys. In

the following sections I will provide brief descriptions of the critical blind HI galaxy surveys that followed AHISS and ADBS as well as some of the upcoming surveys.

### 3.1.1 HIPASS

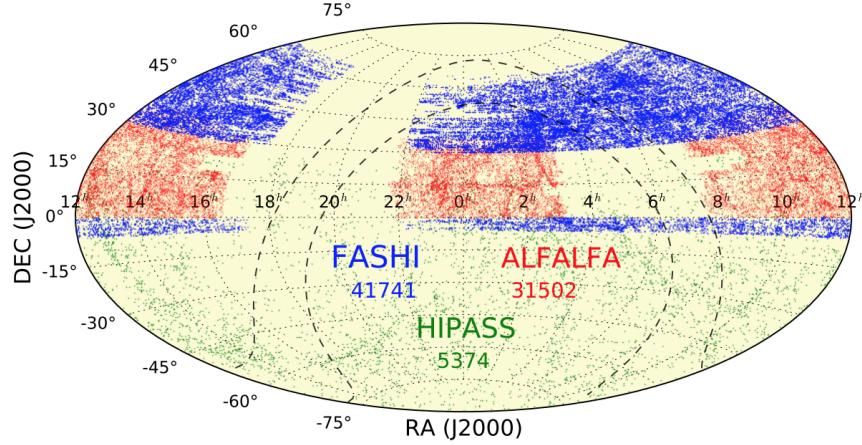
The HI Parkes All Sky Survey (HIPASS) started in 1997 using the Parkes 64 m telescope in Australia. It was the first blind HI survey to observe the entire southern sky at Dec  $< +2^\circ$  [36, 37]. In addition to the southern catalog (HICAT), HIPASS also surveyed a portion of the northern sky from  $+2^\circ < \text{Dec} < +25.5^\circ$  detecting 1,002 HI galaxies for the northern HIPASS catalog (NHICAT) [38]. In total, HIPASS covered about 71% of the sky and detected a total of 5,317 galaxies. Zwaan et al. (2018) [24] estimated the HIMF using the 1,000 brightest galaxies from HICAT and found that the faint-end slope was dependent on morphological type. They also calculated the mass density of neutral hydrogen  $\Omega_{\text{HI}} = (3.8 \pm 0.6) \times 10^{-4} h^{-1}$  and present this as a robust measurement given that the most uncertain measurements (low-mass end) have a small effect on this value.

### 3.1.2 ALFALFA

The current state-of-the-art HI galaxy survey is the Arecibo Legacy Fast ALFA survey (ALFALFA) that was performed by the Arecibo 305 m telescope from 2005 to 2011 [39, 40]. With over 4,400 hours of observation in this time period, this survey is responsible for detecting approximately 31,500 HI sources out to a redshift of  $z = 0.06$ . The detection methods employed by the ALFALFA team have led to 25,437 high significance detections with  $\text{SNR} > 6.5$  [25] but also to some detections with a lower SNR that have been matched to optical counterparts in their pipeline.

ALFALFA observed approximately 7,000 square degrees of the sky in two main strips.

Both strips covered a Declination of  $0^\circ < \text{Dec} < 36^\circ$  but were split into two Right Ascension regimes. One going from  $07h30m < \text{R.A.} < 16h30m$  and the other  $22h < \text{R.A.} < 03h$  [41]. The distribution of the detected sources from these observations are shown in Figure 3.1. The ALFALFA bandwidth was 100 MHz across, centered at 1385 MHz and spanning from 1335 – 1435 MHz. With a total of 4096 individual and equally spaced frequency channels, they each have a width of 24.4 kHz.



**Figure 3.1:** Overlapping source detections from ALFALFA, HIPASS, and FASHI shown as a function of their position on the sky. Figure taken from [2].

One of the main goals of ALFALFA was to improve our measurements of the faint end of the HIMF. With a huge increase in sensitivity compared to past surveys, finer angular and spectral resolution, and a larger sky coverage, ALFALFA has substantially helped make progress in this regards. As an example, a galaxy with  $M_{\text{HI}} = 10^9 M_\odot$  and a velocity width of  $W_{50} = 100 \text{ km/s}$  could be detected by HIPASS out to about  $D = 35 \text{ Mpc}$ . A similar galaxy could be detected by ALFALFA out to  $D \sim 80 \text{ Mpc}$  [25]. The results of these improvements are measurements of the HIMF as low as  $M_{\text{HI}} = 10^6 M_\odot$ . Furthermore, due to the size of the ALFALFA catalog, studies comparing the effects of environment on the measurements of the HIMF are now possible. Jones et al. (2020) [27] compared the global HIMF (entire ALFALFA

catalog) with the HIMF of galaxies within specific groups. They found significant differences in the low-mass slope and “knee” mass, highlighting the importance of large representative and well understood samples for HI galaxy surveys.

### 3.1.3 FAST

The Five-hundred-meter Aperture Spherical radio Telescope (FAST) is a powerful single dish telescope. Starting in 2020, the FAST All Sky HI survey (FASHI) has been observing the northern sky in a blind search for HI galaxies. In the first catalog release, FASHI has covered 7,600 square degrees, 35% of the total projected survey area, and detected 41,741 sources at  $z < 0.09$  [2]. Figure 3.1 shows the footprint of FASHI’s first data release compared to ALFALFA and HIPASS. The full survey will scan the sky at  $-14.3^\circ < \text{Dec} < +66.7^\circ$  with a bandwidth of 1050-1450 MHz and 7.6 kHz resolution [42]. As the survey continues over time, FASHI is expected to detect up to 100,000 HI sources based on estimates from this pilot catalog release.

## 3.2 Future HI surveys: The SKA

Now we turn to future HI galaxy surveys. In Chapter 4 we provide a detailed description of the CHORD observational parameters and instrumental properties but for now we focus on other surveys.

A lot of future HI galaxy surveys come from pathfinder or precursors to the giant radio telescope Square Kilometer Array (SKA) in western Australia and South Africa. The SKA will perform the largest spectroscopic galaxy survey once it becomes operational in the upcoming decade. It is predicted that phase 2 (SKA-2) will detect up to 1 billion galaxies over 3/4 of the sky up to  $z \sim 2$ . Phase 1 will cover over 5,000 square degrees and with a

tenth of the sensitivity of SKA-2 it will detect up to  $5 \times 10^6$  galaxies to  $z = 1.19$  [43]. The following surveys are precursors to the SKA that have either began operations or are set up to begin in the near future.

### 3.2.1 MIGHTEE

The MeerKAT International GigaHertz Tiered Extragalactic Exploration survey (MIGHTEE) is a new survey that has released early science results. The survey is carried out by the South African Meer-Karoo Array Telescope (MeerKAT), an SKA precursor facility with an array of  $64 \times 13.5\text{m}$  dishes observing at 900-1670 MHz (up to  $z = 0.58$ ). Currently, MeerKAT is the most sensitive centimeter wavelength interferometer in the southern hemisphere, only to be outperformed by the SKA [44]. Even though MIGHTEE will have a relatively small survey area (32 square degrees), its frequency range means it will observe a volume of  $0.033 \text{ Gpc}^3$ , 3 times the total HIPASS volume. The flux limit of MIGHTEE is 10 times deeper than ALFALFA on average, meaning that a  $M_{HI} = 10^9 M_\odot$  galaxy with  $W_{50} = 100 \text{ km/s}$  that is observed by ALFALFA out to  $D \sim 80 \text{ Mpc}$  could be detected all the way to  $D \sim 420 \text{ Mpc}$  by MIGHTEE. With this increase in-depth, the survey is predicted to detect up to 2,750 HI galaxies.

Between 2018-2019, early science observations were conducted, from which 276 detections of HI galaxies were made. This sample has been used to estimate the HIMF [45] which is in agreement with results from ALFALFA and other surveys. The early science data cubes have also been used for other applications. Sinigaglia et al. (2022) [46] stacked over 9,000 undetected galaxies with  $0.23 < z < 0.49$  and derived relationships between HI gas content, star formation processes, and galaxy evolution.

### 3.2.2 ASKAP: WALLABY and DINGO

Another SKA precursor is the Australian Square Kilometre Array Pathfinder (ASKAP), a 36 x 12m dish array in a radio-quiet zone in western Australia. ASKAP will undertake a multitude of neutral hydrogen surveys, in particular the Widefield ASKAP L-band Legacy All-sky Blind surveY (WALLABY) and the Deep Investigation of Neutral Gas Origins (DINGO) [47]. In its two phases – deep and ultradeep – DINGO is predicted to detect more than 60,000 HI galaxies at relatively high redshifts of  $0.1 < z < 0.43$  with 2,500 hours of integration time per 60 square degree field.

WALLABY is planned to cover 3/4 of the entire sky, with 100% of the southern sky and a portion of the northern sky up to Dec  $< +30^\circ$ . With 5 years of planned surveying, it is predicted to detect up to 210,000 HI galaxies out to  $z = 0.26$  [48]. A pilot survey has been completed in  $3 \times 60$  square degree fields out to  $z = 0.08$  and  $\sim 600$  detections have been made. Over 100 of those detections were spatially resolved by the  $30''$  ASKAP beam, making it possible to model the kinematics of the gas within them. As a high-resolution, wide-field survey, WALLABY will be able to sufficiently resolve thousands of HI galaxies and increase the sample of those available for more detailed studies in the future.

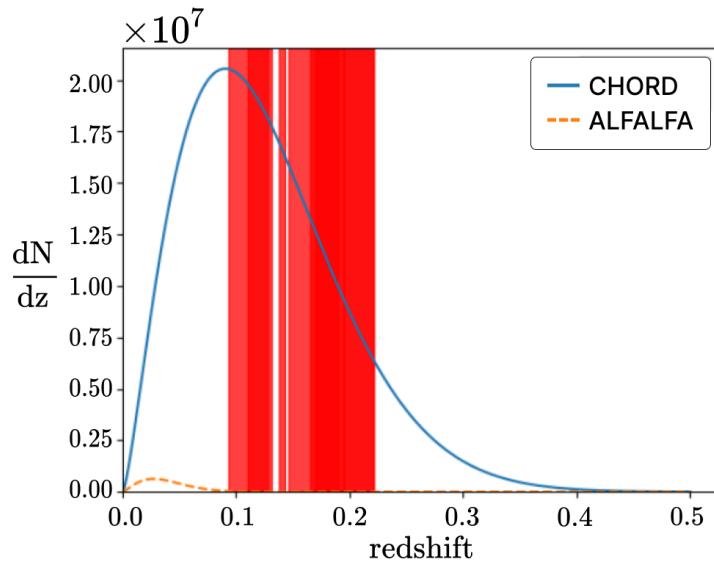
## Chapter 4

# The Canadian Hydrogen Observatory and Radio Transient Detector

The Canadian Hydrogen Observatory and Radio Transient Detector (CHORD) is a next generation radio interferometer currently being built at the Dominion Radio Astrophysical Observatory. CHORD's design stems from the tremendous success of the Canadian Hydrogen Intensity Mapping Experiment (CHIME) while incorporating cutting edge Canadian technologies. The construction of CHORD is driven by three main science goals, but its innovative technology and specifications open it to a wide range of explorations. The three main objectives are: detecting and localizing Fast Radio Bursts (FRBs), mapping the matter distribution in cosmological scales, and probing fundamental physics parameters. The scope of this thesis lies solely on the second point. For further information on other applications of CHORD, see Vanderlinde et al. (2019) [3].

CHORD's approach to the cosmology science case is two-fold. In the nearby universe, CHORD will have enough angular resolution and sensitivity to detect individual HI galaxies. Further out, observed through the lower end of CHORD's frequency coverage, the

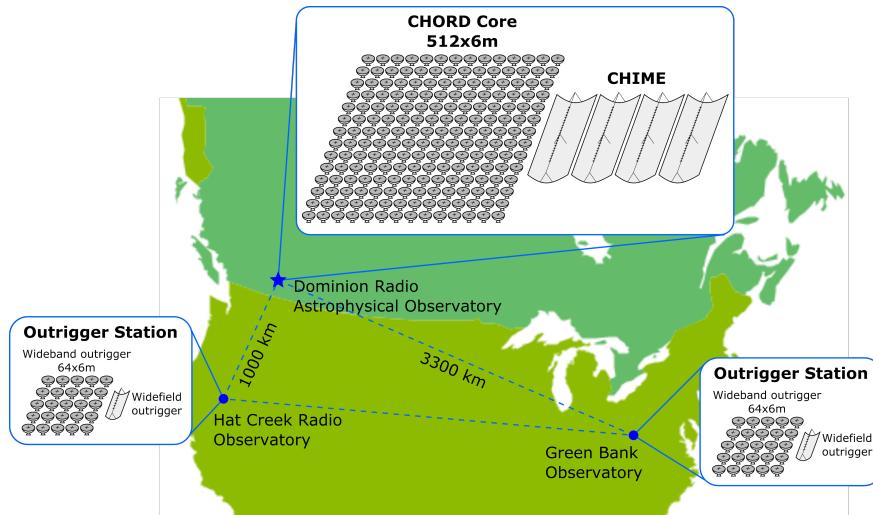
instrument will no longer be used for detecting single galaxies but will measure the integrated emission of unresolved sources. This technique of Intensity Mapping (IM) promises to improve cosmological constraints in a large range of epochs [49]. There is a third regime in which CHORD will still be able to pick out individual galaxies, say the ones with highest HI mass, while observing other sources as an integrated field. In combination, CHORD is forecast to produce the largest 3D map of the universe to date, covering a huge volume from  $z = 0$  to  $z \approx 3$ .



**Figure 4.1:** The distribution of detected galaxies as a function of redshift. The blue solid line shows the prediction for CHORD while the orange dashed line shows all the ALFALFA detections. CHORD shows a significant increase in source counts at all redshifts and peaks at a higher redshift. The red regions frequencies with challenging RFI (radio frequency interference) contamination. Adapted from [3].

The research done for this thesis is tuned to the lower redshift universe. CHORD's large spectral window and spatial coverage, and fine angular resolution give it the potential to

detect up to 10 million individual unresolved HI galaxies<sup>1</sup>. CHORD’s instantaneous FoV ranges from  $5 - 130$  deg $^2$  through its bandwidth and is intended to observe most of the northern sky. In addition, CHORD’s increased sensitivity means that it will be able to detect HI galaxies further down the HI mass function. As described in Section 3.1.2, the current state-of-the-art survey ALFALFA has detected 30,000 HI galaxies, meaning CHORD will provide a 300-fold improvement. Figure 4.1 is a visual comparison of both ALFALFA and CHORD detection counts as a function of redshift. The red regions mark regions with challenging RFI environments. Though the precise value of the maximum redshift where CHORD will detect galaxies will strongly depend on the RFI and other factors, it is safe to say it will be higher than those accessed by ALFALFA. The major expansion of the volume surveyed and improved observation depth combine to forecast the huge contribution of CHORD to HI galaxy detection efforts.



**Figure 4.2:** CHORD system overview. Image credit: Juan Mena.

---

<sup>1</sup>The term “unresolved” often takes on different meanings in cosmology and in astronomy. Cosmologists use the term unresolved to refer to measurements that cannot separate individual sources from one another. In astronomy, an unresolved observation refers to one where the internal structure of the source is not spatially resolved. Throughout this thesis, we assume the latter definition: galaxies are detected individually but their morphology is not resolved.

The full CHORD instrument is expected to be operational in the summer of 2025. At this stage, CHORD will have 2 outrigger stations each containing a CHIME-like 30 m cylindrical reflector and an array of 64 individual 6 m dishes. In addition, CHORD will have a core array of  $512 \times 6$  m composite dishes in a regular rectangular configuration with a 6.3 m distance between East-West dishes and 8.5 m for North-South ones. In combination, the instrument totals a huge collecting area of  $14,400 \text{ m}^2$ . CHORD is an ultra-wide band radio interferometer, covering frequencies in the range of 300 – 1500 MHz with a coarse frequency resolution of 0.586 MHz. MacKay et al (2023) [50] performed simulations with the CHORD feeds and found that the system temperature was  $T_{\text{sys}} < 30 \text{ K}$  for most of the frequency band. Figure 4.2 situates CHORD on the map and gives an overview of the instrument’s set-up. A summary of the instrument parameters is given in Table 4.1. A year prior to the full completion of CHORD, a subset of 66 dishes will be set up as the instrument’s pathfinder and operations will begin.

Parameter	Pathfinder	Full CHORD
Core dishes	66	512
Dish diameter	6 m	6 m
Spectral range	300-1500 MHz	300-1500 MHz
Spectral resolution	586 kHz	586 kHz
Outrigger dishes		64/station
Cylinders		1/station
Collecting area	$1,800 \text{ m}^2$	$14,400 \text{ m}^2$
Longest baseline	$\sim 80 \text{ m}$	$\sim 3300 \text{ km}$

**Table 4.1:** Overview of CHORD instrument parameters.

# Chapter 5

## Radio Interferometry

### 5.1 Mathematical Description

As radio astronomy evolves and new instrumentation is developed, two of the key parameters we try to optimize are the sensitivity of observations and the angular resolution. The sensitivity of a radio telescope depends on both observational and instrumental parameters. For a single-dish telescope, these dependencies are described by the radiometer equation written as

$$\sigma_{\text{rms}} = \frac{k_B T_{\text{sys}}}{A_{\text{eff}} \sqrt{\Delta\nu \Delta t}}. \quad (5.1)$$

Here  $\sigma_{\text{rms}}$  represents the root mean squared (rms) noise of the particular observation and  $k_B$  is the Boltzmann constant. This equation frames the sensitivity dependence on various factors as follows.

- $T_{\text{sys}}$ : The system temperature, with a sky contribution and a receiver contribution.  
This is a set property of the instrument, its electronics, and the sky.
- $A_{\text{eff}}$ : The effective collecting area of the telescope, also a set property of the instrument.

- $\Delta\nu$ : The spectral width of the observation, this could potentially be modified per observation depending on the instrument set up.
- $\Delta t$ : The integration time spent on the source, this could either be determined by the time spent pointing at the source or by the interval over which the source is crossing the FoV of the instrument.

Larger and larger single-dish telescopes have been built in order to achieve both a larger  $A_{\text{eff}}$ , leading to increased sensitivity, and a better angular resolution. The largest single-dish radio telescope to have been built is the Five-hundred-meter Aperture Spherical radio Telescope (FAST) in Guizhou, China [4]. We approximate the angular resolution of this telescope with the known relation assuming a circular aperture

$$\theta = 1.22 \frac{\lambda}{D}, \quad (5.2)$$

where  $D$  is the diameter of the dish and  $\lambda$  is the observing wavelength. For a 21 cm observation of neutral hydrogen in the local universe, the resolution of FAST would be  $\theta \approx 1.8'$ . If we attempted to observe sources at higher redshifts (with longer wavelengths), this resolution would only get worse. There are considerable engineering challenges that the community would have to face to continue building larger functioning single dish radio telescopes. To name a few, the financial costs of materials strong enough to withstand the stress caused by such large structures would soar, and the possibility of steering the dish to precisely track a source would quickly become a mammoth task.

As an alternative, radio interferometry consists of creating a synthesized telescope out of multiple dishes in predetermined configurations. In order for these dishes to function as a large single dish telescope, the positions of the antennas must be known precisely and the systematics should be well understood. Instead of the dish diameter, the relevant sizes for

interferometry are given by the separations between different pairs of antennas, called the baselines and labelled  $\vec{b}$ .

Before giving a more in-depth mathematical description of interferometry, I establish parallels with Equations 5.1 and 5.2 and provide some intuition for the scales of interferometric measurements. As mentioned, one of the main cases for interferometers is that a large number of dishes can increase the effective area  $A_{\text{eff}}$  of the instrument. The rms noise is also further suppressed by the number of baselines, as each of them performs their independent measurement. The radiometer equation equivalent for an interferometer containing  $N$  dishes is

$$\sigma_{\text{rms}} = \frac{k_B T_{\text{sys}}}{A_{\text{eff}} \sqrt{N(N-1)\Delta\nu\Delta t}}, \quad (5.3)$$

where  $N(N-1)$  gives the total number of baselines.

In terms of angular resolution Equation 5.2 is replaced by the approximation

$$\theta \approx \frac{\lambda}{b_{\max}}, \quad (5.4)$$

where  $b_{\max}$  is the length of the longest baseline. With this, it is clear that with interferometers we can achieve better angular resolutions than we ever could with a single dish instrument. The largest angular scale accessible to the interferometer (the instantaneous FoV) is set by the dish size  $D$  with  $\theta \approx \lambda/D$ . We can relate both the small and large angular scales to physical sizes perpendicular to the observer's line-of-sight through

$$r_{\perp} = D_c(z)\theta, \quad D_c(z) = \frac{c}{H_0} \int_0^z \frac{dz}{E(z)}. \quad (5.5)$$

Here  $D_c(z)$  is the comoving distance to the source at a given redshift  $z$ . In this equation,  $E(z) = \sqrt{(\Omega_m(1+z)^3 + \Omega_\Lambda)}$ , where  $\Omega_m$  is the present day fraction of the total energy density budget accounted for by matter (baryonic and dark matter) and  $\Omega_\Lambda$  is the equivalent for

dark energy. Their values depend on the chosen cosmology [51].

Along the line-of-sight, the particular configuration of the array does not impact the sensitivity of the instrument to different scales. The largest scale accessible is dictated by the total bandwidth of the instrument while the smallest is dictated by its spectral resolution  $\Delta\nu$ . The relationship is defined by

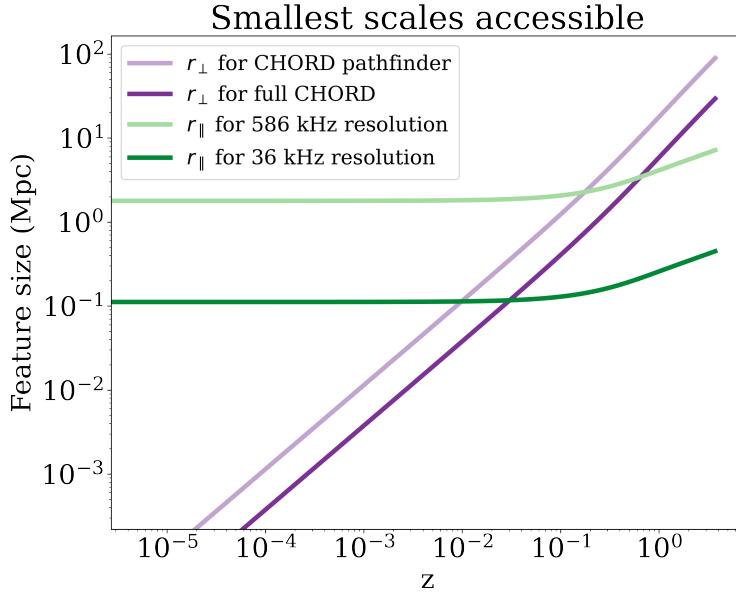
$$\Delta r_{\parallel} = \frac{c}{H_0 \nu_{21}} \frac{(1+z)^2}{E(z)} \Delta\nu, \quad (5.6)$$

where  $H_0$  is the Hubble constant. Calculations for the CHORD pathfinder and full CHORD core array are shown in Figure 5.1 where we use  $H_0 = 69.32 \text{ km Mpc}^{-1} \text{ s}^{-1}$ ,  $\Omega_m = 0.2865$ ,  $\Omega_\Lambda = 0.7134$  [52]. The purple lines represent the  $r_{\perp}$  feature sizes resolved by the instrument which are smaller by a constant ratio for the full CHORD core compared to the pathfinder for the entire redshift range. This makes sense given the difference in longest baselines of  $b_{\max} \sim 80\text{m}$  for the pathfinder and  $b_{\max} \sim 250\text{m}$  for the full CHORD core. The green lines represent the  $r_{\parallel}$  feature sizes in both the coarse CHORD spectral resolution of  $\Delta\nu = 586\text{kHz}$  and a resolution about 16 times finer  $\Delta\nu \sim 36 \text{ kHz}$ . The difference in these is very apparent and will be further discussed in Section 5.2.

While the intuition for an interferometer can be built with connections to the single dish counterparts, these two approaches to radio astronomy have fundamental differences. A significant difference is exactly what the interferometer measures, the visibilities. The analytical expression of the instantaneous visibility measured for a wavelength  $\lambda$  by a pair of antennas  $i, j$  under the flat sky approximation of some temperature field  $T(\vec{\theta}, \lambda)$  is

$$V_{ij}(\lambda) = \int d^2\theta T(\vec{\theta}, \lambda) A_i(\vec{\theta}, \lambda) A_j^*(\vec{\theta}, \lambda) e^{-2\pi i \vec{u}_{ij} \cdot \vec{\theta}}. \quad (5.7)$$

Here  $A_i(\vec{\theta}, \lambda)$  and  $A_j^*(\vec{\theta}, \lambda)$  are the primary beams of the  $i$ -th and  $j$ -th dishes, respectively, and describe the sensitivity of the antennas to different angles on the sky  $\vec{\theta}$ . This equation



**Figure 5.1:** Scales accessible to the CHORD instrument. The purple lines represent the  $r_{\perp}$  feature sizes resolved by the pathfinder (light purple) and full-CHORD (dark purple). The green lines represent the  $r_{\parallel}$  feature sizes in both the coarse CHORD spectral resolution of 586 kHz (light green) and a resolution about 16 times finer  $\sim 36$  kHz (dark green).

has the basic form of a Fourier transform that roughly samples a mode  $\vec{u}_{ij} = (u, v)_{ij}$ . This uv-mode depends on the baseline  $\vec{b}_{ij}$  formed by the antennas through the relationship

$$\vec{u}_{ij} = \frac{\vec{b}_{ij}}{\lambda}. \quad (5.8)$$

The extent to which a particular interferometer is able to sample the Fourier modes (its uv coverage) is therefore entirely determined by the baselines formed by its dishes. More points on the uv plane are naturally sampled by the instrument due to the Earth's rotation, i.e. we get slightly different baseline vectors as a function of time. The uv coverage is then also a function of observing time and is more complete as we observe for a full sidereal day. Coverage can also be improved with the addition of new dishes if those form new

baseline vectors and not duplicates of existing baseline vectors. Not all scales are sampled as we don't have an infinite number of dishes forming a smooth, continuous coverage of all possible baselines. Holes in the uv plane conceptually signify the loss of information on certain Fourier scales. Practically, incomplete uv coverage poses challenges for map-making (see Chapter 6) which may be a necessary data product for some science cases.

To express the whole visibility data set measurement, we must include the time evolution. We make the assumption that the true sky  $T(\vec{\theta}, \lambda)$  does not change over time. The primary beams could have a temporal evolution and vary for each dish, though we might start with the assumption that they do not and model any discrepancy as a small perturbation. Finally, we include the time evolution of  $\vec{u}_{ij}$  so we can write

$$V_{ij}(\lambda; t) = \int d^2\theta A_i(\vec{\theta}, \lambda; t) A_j^*(\vec{\theta}, \lambda; t) T(\vec{\theta}, \lambda) e^{-2\pi i \vec{u}_{ij}(t) \cdot \vec{\theta}} + n_{ij}(t). \quad (5.9)$$

Here  $n_{ij}(t)$  is the time-dependent noise injected by the pair of antennas  $i, j$  at the time of observation. This view of interferometry serves as a good foundation for conceptual understanding and for the simulations undertaken in this thesis. The following section explains a more detailed aspect of interferometry, focusing on the spectral properties in what is a key element of the pipeline developed for this project.

## 5.2 Up-channelization

As described in Section 5.1, the fundamental output of a radio interferometer is a visibility matrix which describes the cross-correlation of measurements from pairs of antennas. This section will give an overview of the process of increasing the spectral resolution of an interferometric measurement through an up-channelization algorithm. This mathematical exposition closely follows a note written by Kendrick Smith (private memo, CHORD

collaboration).

In a radio interferometer, each antenna measures an electric field time-stream  $E_0(t_0)$  with a short sampling rate  $t_0$ . For CHORD, the sampling rate is  $t_0 = 0.417$  ns. In order to obtain spectral information from this raw measurement, the data is split into chunks of  $N$  samples and Fourier transformed, resulting in a 2-dimensional complex quantity  $E_1(c, t_1)$ . Here,  $c$  indexes the frequency channel and  $t_1 = N \times t_0$  is the slower sampling rate. At the core of this process is a conversion between temporal and spectral resolution that conserves the amount of information recorded by the instrument. The caveat of this very simplistic version of channelization is a large amount of spectral leakage where each channel would contain some complicated combination of signals from a wide range of frequencies.

A more sophisticated approach to channelization is to use a Polyphase Filterbank (PFB) algorithm, which in its Fourier form can be written as

$$E_1(c, t_1) = \int_{-N/2}^{N/2} d\nu \widetilde{W}(c - \nu) E_0(\nu) e^{2\pi i \nu t_1}. \quad (5.10)$$

Here  $E_0(\nu)$  is the Fourier transform of  $E_0(t_0)$ . For CHORD, we have  $N = 4096$  samples, so  $t_1 = 4096 \times 0.417$  ns = 1.71  $\mu$ s and  $c = 0, 1, \dots, 2047$ . This results in the figure cited for CHORD's coarse channel resolution of  $\Delta\nu = (1500\text{MHz} - 300\text{MHz})/2048 \approx 586\text{kHz}$ , as specified in Chapter 4. The weighting term  $\widetilde{W}(c - \nu)$  represents the response of channel  $c$  to frequency  $\nu$  and is chosen to be a sinc-Hanning window function

$$W(s) = \cos\left(\frac{\pi(s - MN/2)}{MN}\right)^2 \operatorname{sinc}\left(\frac{s - MN/2}{N}\right),$$

where  $M$  is the number of taps in the PFB (see [53] for an in-depth description, for our purposes we fix  $M = 4$ ). This window function is used because its Fourier transform  $\widetilde{W}(s)$  approximates a step function, meaning that the resulting channelized data suffers from less severe spectral leakages.

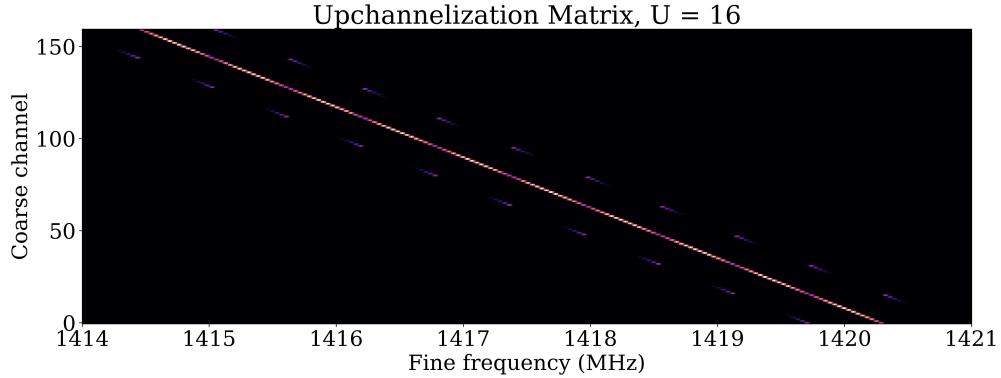
Particular science cases, such as FRB search and the HI galaxy survey, will benefit from a finer frequency resolution than the coarse  $\Delta\nu = 586$  kHz channel width. For this purpose, a two-stage PFB algorithm will be used by CHORD to up-channelize the data and obtain a frequency resolution a factor of  $U$  higher so that  $\Delta\nu_{\text{upchan}} = 586\text{kHz}/U$ . The algorithm will convert  $E_1(c, t_1) \rightarrow E_2(c, u, t_2)$  where  $c$  still indexes the coarse channels,  $u = 0, 1, \dots, U - 1$  indexes the fine channels within each coarse channel, and  $t_2$  is an even coarser time sampling rate of  $U \times t_1$ . The Fourier form of the second round of PFB, the up-channelization step, is

$$E_2(c, u, t_2) = \int d\nu R(c, u; \nu) E_0(f) e^{2\pi i \nu U t_2}. \quad (5.11)$$

Here  $R(c, u; \nu)$  is the up-channelization response matrix that can be written as

$$R(c, u; \nu) = \widetilde{W}(c - \nu) \widetilde{C}(u, -\nu), \quad (5.12)$$

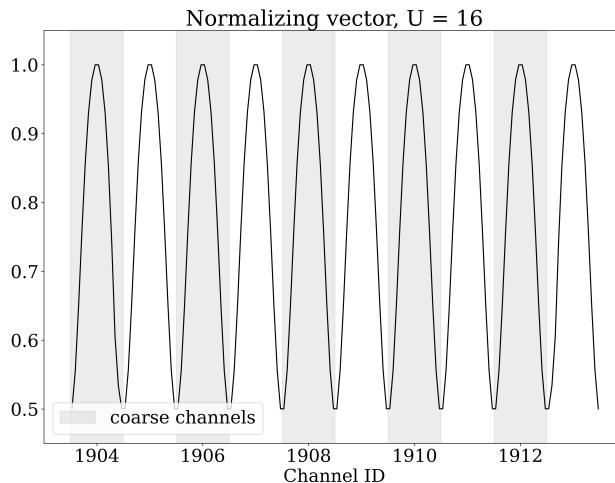
containing the Fourier transform of both the sinc-Hanning function from the first round of PFB and of the up-channelization kernel  $C(u, s) = W(s) e^{\pi i s(U-1)/U} e^{-2\pi i u s/U}$ .



**Figure 5.2:** Up-channelization matrix  $R(c, u; \nu)$  for a frequency range 1414–1421 MHz with  $U = 16$ . There is a bright diagonal region illustrating the maximum sensitivities of channels of certain frequencies. There are also features depicting spectral leakage and modulation (see text for explanation).

Figure 5.2 shows an example of the response matrix  $R(c, u; \nu)$  as described by Equation 5.12 with  $U = 16$  over the frequency range  $\nu \in [1414, 1421]$ . The bright diagonal line down the middle of the plot illustrates the fact that this algorithm performs well in terms of spectral leakage, i.e. each fine channel corresponds mostly to a narrow frequency range. However, the two-round PFB algorithm is not perfect and to either side of the diagonal in a periodic pattern we see a small amount of leakage, subsequently denoted as spectral aliasing. Another interesting feature is that throughout the diagonal, there is some modulation of the response, it is not uniformly bright across. The effect of this modulation is that it imprints predictable spectral features onto the data which are, at least in principle, removable. Figure 5.3 shows that the length of modulation corresponds to the size of each coarse channel, with a peak in sensitivity in the centers of coarse channels and as low as 50% of that sensitivity at their edges. This normalizing vector, computed by applying  $R(c, u; \nu)$  to a constant (frequency independent) signal, can be divided out of any up-channelized data to recover, in principle, the non-modulated signal.

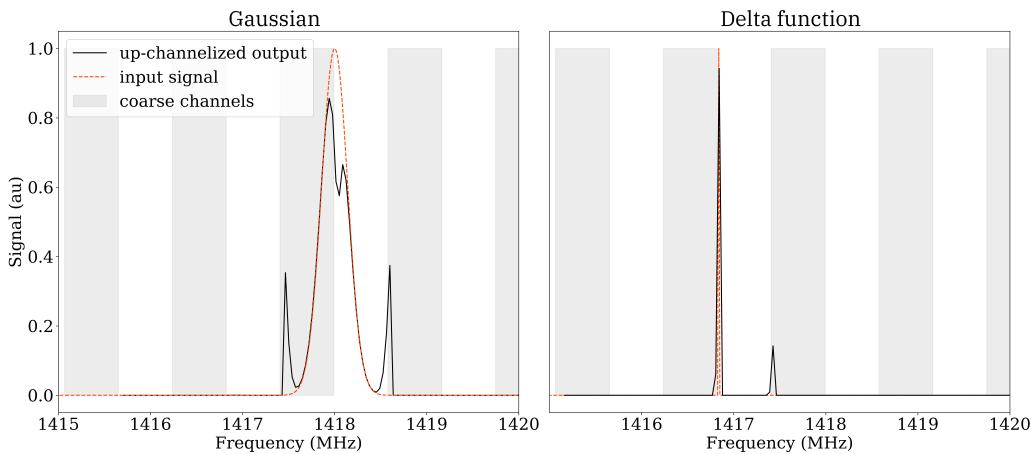
Figure 5.4 shows the outcome of up-channelizing a signal in practice. The right panel contains a Gaussian centered between two coarse channels (shown in red). The up-channelized signal is shown in black. Peaks to either side correspond to spectral aliasing and are almost symmetrical, with the difference stemming from a small shift in the alignment of the Gaussian towards one of the channels. Noticeably, even though the modulation has been divided out of the up-channelized output, a significant dip in the signal at the edge of the channel is still evident around 1418 MHz. On the right panel, an up-channelized delta function is shown. The input signal was centered at the center of a fine channel near the edge of a coarse channel. The response in this case is highly asymmetric. Both of these signals were up-channelized to the same  $U = 16$  factor. Figure 5.4 illustrates that  $U$ , the shape of the input signal, and its precise location relative to the



**Figure 5.3:** Normalization vector for a frequency range  $1414 - 1421$  with  $U = 16$ . The coarse channels are marked by the alternating gray and white background. There is an envelope around each coarse channel, with peak sensitivity in the centers of coarse channels and a significant drop in sensitivity around their edges.

coarse and fine channels all have considerable impacts in the output of the two-stage PFB algorithm described in this section.

The optimal value of  $U$  will depend on the science case. For HI galaxies, the spectral extent of the profiles is the major determining factor and can be quantified by their half-width  $W_{50}$ . Using the full ALFALFA catalog, Oman 2021 [54] determined the number density of HI sources as a function of their  $W_{50}$ , the HI velocity width function (HIWF). From this distribution, likely values of  $W_{50}$  can be extracted to help define representative spectral extent values and inform the choice of the up-channelization factor  $U$ . At the higher end of their HIWF, Oman 2021 predicts [54] a small but significant fraction of galaxies with  $W_{50} \sim 500$  km/s. For 21 cm, this corresponds to  $\Delta\nu \sim 2.5$  MHz. With CHORD's coarse resolution, a galaxy spanning this range would be observed on approximately 4 channels. This observation would likely still not have enough resolution to identify the profile shape



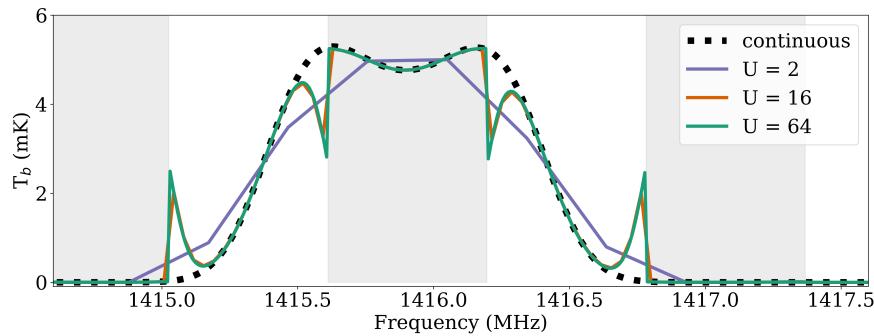
**Figure 5.4:** Two examples illustrating the effects of up-channelization on a Gaussian function and a delta function. The coarse channels are marked by the alternating gray and white background. We show how the placement of the original signal produces different aliasing responses and symmetries in the output signal.

and extract physical parameters from the source. At the lower end, we have  $W_{50} \sim 10$  km/s, corresponding to  $\Delta\nu \sim 50$  kHz. We would require a boost of  $U = 8$  to 16 to observe these galaxies on more than one frequency channel and as high as  $U = 32$  to 64 to start resolving interesting features such as double-horns and possible asymmetries.

Different science cases will have different requirements and so the flexibility of the two-stage PFB algorithm for up-channelization is a desirable feature. For both the HI galaxy survey and the FRB search, the need for high resolution does not necessarily span the entire CHORD band. For HI galaxies, for example, the known dominance of RFI at certain channels means we could skip up-channelization of such channels and instead redirect resources to achieve a better resolution at higher frequencies near  $z = 0$ . It is therefore possible to apply different up-channelization factors across the CHORD band depending on the desired outputs and intentions for the data.

It is good to note that higher  $U$  values are not necessarily always better than lower ones

after some minimum threshold has been reached. The resolution will continue to increase but at the cost of more computational power and coarser temporal resolution. Figure 5.5 shows a continuous simulated HI galaxy profile (in black) up-channelized by different factors  $U$ . The coarse channels are marked by the alternating gray and white background. With CHORD's coarse resolution, all of the detail would be lost with the profile spanning only 3 channels. With  $U = 2$ , there is only a slight improvement (shown in purple) but the spectral features, namely the dip in the center, are not identified. At  $U = 16$  (shown in orange), the double-horn nature of the profile becomes apparent and this of course continues for  $U = 64$  (shown in green). However, for  $U \geq 16$ , there is significant spectral aliasing around 1415 MHz and 1417 MHz and very clear dips at the edges of the coarse channels due to modulation. These extra imprinted features could result in a more complicated detection process and eventually in the extraction of incorrect parameters and physical properties of the source from the observed profile. Finally, there seems to be little improvement in going from  $U = 16 \rightarrow U = 64$  as the features were already appropriately traced with the lower up-channelization factor. All of this combined calls for a careful analysis to determine the optimal up-channelization scheme for each science case.



**Figure 5.5:** Similar to Figure 5.4 but with a simulated HI galaxy profile up-channelized with factors  $U = 2, 16, 64$  (respectively in purple, orange, and green). The dotted black line shows the continuous profile.

# Chapter 6

## Map-making

### 6.1 Mathematical Description

In Chapter 5, we described the output of a radio interferometer observation as a set of time-ordered complex numbers, the visibilities. These contain all of the information measured for a given baseline of the array and while a lot can be done with this data product, it is often useful to transform it into a map. A map is an image-based representation of the data observed by the instrument. The process of creating a map from visibility data will be described mathematically and conceptually in this chapter. Contrary to other formalisms common in radio astronomy, in which most of the analysis takes place in Fourier space, maps represent data in real space. As simple as the definition is, map-making is not a straightforward process, requiring a multitude of mathematical manipulations and theoretical assumptions to produce the end results. The properties of one's instrument, as well as specific analysis choices, will result in maps containing artifacts that affect our ability to identify HI galaxies. We address these effects in Sections 7.3 and 7.6 and throughout Chapter 8.

First, we reformulate the mathematical description of the visibility measurement as a

matrix equation

$$\vec{v} = \mathbf{B}\vec{s} + \vec{n}, \quad (6.1)$$

where  $\vec{v}$  is a vector containing the discrete visibility data for each pair of antennas and  $\vec{s}$  is a vector containing the true sky intensity.  $\mathbf{B}$  is a matrix encoding the instrument's response to the sky and by comparison with Equation 5.9 is written as

$$\mathbf{B}_{ij}(\vec{\theta}, \lambda; t) = A_i(\vec{\theta}, \lambda; t) A_j^*(\vec{\theta}, \lambda; t) e^{-2\pi i \vec{u}_{ij}(t) \cdot \vec{\theta}}. \quad (6.2)$$

This matrix has  $m$  columns and  $n$  rows, where  $m$  is the number of pixels on the sky and  $n = 1, 2, \dots, (N_{\text{baselines}}) \cdot (N_{\text{time}})$ . Finally,  $\vec{n}$  is the noise vector added by the instrument at the time of the measurement. If we were to expand the matrices and vectors we would have

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1m} \\ B_{21} & B_{22} & \cdots & B_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{nm} \end{bmatrix} \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_n \end{pmatrix}. \quad (6.3)$$

From this mathematical description, the output of an observation  $\vec{v}$  can be interpreted as a particular linear combination of the real sky  $\vec{s}$  due to the instrument's response  $\mathbf{B}$  plus some added noise  $\vec{n}$ .

Map-making is conceptually straightforward, it is the act of recovering  $\vec{s}$  from  $\vec{v}$  given some previous knowledge of the instrument and its response to the sky. In the framework above, map-making simply involves inverting a linear operation which would be trivial under special circumstances. Practically, however, this inversion can present some challenges, some of which are the impact of the noise on the data and the exact mathematical nature of  $\mathbf{B}$  for a given observation. For instance, holes in the uv-coverage of the instrument make this process

formally uninvertible. For the remainder of this chapter, I will describe the mathematical process of map-making, but the reader is encouraged to keep this clear conceptual picture in mind.

## 6.2 Maximum Likelihood

To perform the statistical inversion required to make a map, we can use the classical approach of maximizing a likelihood function for the sky. In this case, the likelihood function is

$$L(\vec{s}) = p(\vec{v}|\vec{s}) \propto e^{-\chi^2}, \quad (6.4)$$

where  $p(\vec{v}|\vec{s})$  is the probability of measuring a set of visibilities  $\vec{v}$  given some model of the sky  $\vec{s}$ . Now, assuming the noise follows a zero-mean Gaussian distribution with known covariance  $\mathbf{N} = \langle \vec{n}\vec{n}^\dagger \rangle$ , we can write

$$\chi^2 = (\vec{v} - \mathbf{B}\vec{s})^\dagger \mathbf{N}^{-1} (\vec{v} - \mathbf{B}\vec{s}). \quad (6.5)$$

Given  $\mathbf{N}^{-1} = (\mathbf{N}^{-\frac{1}{2}})^\dagger (\mathbf{N}^{-\frac{1}{2}})$ , we can rewrite and simplify this expression

$$\begin{aligned} \chi^2 &= ((\vec{v} - \mathbf{B}\vec{s})\mathbf{N}^{-\frac{1}{2}})^\dagger (\mathbf{N}^{-\frac{1}{2}}(\vec{v} - \mathbf{B}\vec{s})) \\ &= (\mathbf{N}^{-\frac{1}{2}}(\vec{v} - \mathbf{B}\vec{s}))^* (\mathbf{N}^{-\frac{1}{2}}(\vec{v} - \mathbf{B}\vec{s})) \\ &= |\mathbf{N}^{-\frac{1}{2}}(\vec{v} - \mathbf{B}\vec{s})|^2. \end{aligned}$$

We want to minimize  $\chi^2$  with respect to  $\vec{s}$  so we write

$$\begin{aligned} \frac{d\chi^2}{d\vec{s}} &= 2|\mathbf{N}^{-\frac{1}{2}}(\vec{v} - \mathbf{B}\vec{s})|(-\mathbf{N}^{-\frac{1}{2}}\mathbf{B}) \\ 0 &= -(\mathbf{N}^{-\frac{1}{2}}\vec{v})(\mathbf{N}^{-\frac{1}{2}}\mathbf{B}) + (\mathbf{N}^{-\frac{1}{2}}\mathbf{B}\vec{s})(\mathbf{N}^{-\frac{1}{2}}\mathbf{B}). \end{aligned}$$

Now we rearrange to find the estimated sky vector  $\hat{s}$  that maximizes the likelihood of having

observed  $\vec{v}$  with our telescope as follows

$$\begin{aligned} (\mathbf{N}^{-\frac{1}{2}} \mathbf{B} \vec{s}) (\mathbf{N}^{-\frac{1}{2}} \mathbf{B}) &= (\mathbf{N}^{-\frac{1}{2}} \vec{v}) (\mathbf{N}^{-\frac{1}{2}} \mathbf{B}) \\ \hat{s} &= (\mathbf{N}^{-\frac{1}{2}} \mathbf{B})^\dagger (\mathbf{N}^{-\frac{1}{2}} \vec{v}) \\ \hat{s}_{dirty} &= \mathbf{B}^\dagger \mathbf{N}^{-1} \vec{v}. \end{aligned} \quad (6.6)$$

where  $\mathbf{B}^\dagger$  denotes the conjugate transpose of  $\mathbf{B}$ . This procedure results in the simplest possible map made from the visibility data, the dirty map (Equation 6.6).

As implied earlier, depending on the particulars of the experimental set-up and instrument model,  $\mathbf{B}$  will take a different form. In section 6.3, I introduce the *m*-mode formalism which is a logical and useful approach to map-making for a particular but substantial subset of radio interferometers.

### 6.3 The *m*-mode formalism

The *m*-mode formalism is a powerful mathematical basis over which we can express quantities in a convenient way for the specific case of a transit telescope with a fixed pointing. As a driftscan telescope, CHORD will output measurements consistent with the use of this formalism and so this is a relevant tool-set to explore in this thesis.

The particular characteristic that sets driftscan and tracking telescopes apart and allow for this clever mathematical manipulation is the direct linear relationship between the angular and temporal variables. An instrument with a fixed pointing will sweep out the entire  $360^\circ$  of a stripe of the sky over a sidereal day. This observation can thus be parameterized equally by reference to the azimuthal angle of pointing or to the time of observation. This allows us to perform some significant and very useful simplifications to the base visibility equation by replacing the time-dependence in Equation 5.9 for an

azimuthal dependence and write

$$V_{ij}(\lambda; \phi) = \int d^2\hat{n} \mathbf{B}_{ij}(\hat{n}, \lambda; \phi) T(\hat{n}, \lambda) + n_{ij}(\phi). \quad (6.7)$$

We take the Fourier transform of the visibilities along the periodic  $\phi$  direction. For the following mathematical manipulations, I will drop the wavelength  $\lambda$  for clarity but the dependence remains. Let the Fourier transform of  $V_{ij}(\phi) = V_m^{ij}$ , where  $m$  represents the  $m$ -th Fourier mode so that

$$\begin{aligned} V_m^{ij} &= \int \frac{d\phi}{2\pi} V_{ij}(\phi) e^{-im\phi} \\ &= \int \frac{d\phi}{2\pi} \left[ \int d^2\hat{n} \mathbf{B}_{ij}(\hat{n}, \lambda; \phi) T(\hat{n}, \lambda) + n_{ij}(\phi) \right] e^{-im\phi}. \end{aligned} \quad (6.8)$$

Now, we use the spherical harmonic expansion of the sky  $T(\hat{n})$  and the matrix  $\mathbf{B}$

$$T(\hat{n}) = \sum_{lm} a_{lm} Y_{lm}(\hat{n}), \quad \mathbf{B}_{ij}(\hat{n}; \phi) = \sum_{lm} B_{lm}^{ij}(\phi) Y_{lm}^*(\hat{n}),$$

where  $l, m$  are the spherical harmonics degree and order,  $a_{lm}$  are the coefficients of the spherical harmonic expansion of the sky and  $Y_{lm}(\hat{n})$  are the spherical harmonics. In the second expansion,  $B_{lm}^{ij}(\phi)$  is the representation of  $\mathbf{B}$  from Equation 6.2 in spherical harmonic space, which we call the beam transfer matrix. Now we can use these expansions in our Fourier transform to write

$$\begin{aligned} V_m^{ij} &= \int \frac{d\phi}{2\pi} \left[ \int d^2\hat{n} \sum_{lm'} B_{lm'}^{ij}(\phi) Y_{lm'}^*(\hat{n}) \sum_{lm'} a_{lm'} Y_{lm'}(\hat{n}) + n_{ij}(\phi) \right] e^{-im\phi} \\ &= \sum_{lm'} \int \frac{d\phi}{2\pi} \left[ B_{lm'}^{ij}(\phi) a_{lm'} \int d^2\hat{n} Y_{lm'}^*(\hat{n}) Y_{lm'}(\hat{n}) + n_{ij}(\phi) \right] e^{-im\phi} \\ &= \sum_{lm'} \int \frac{d\phi}{2\pi} B_{lm'}^{ij}(\phi) a_{lm'} e^{-im\phi} + n_{ij}(\phi) e^{-im\phi}. \end{aligned} \quad (6.9)$$

To get to Equation 6.9 we used the orthonormal properties of the spherical harmonics,

namely

$$\int d^2\hat{n} Y_{lm'}^*(\hat{n})Y_{lm'}(\hat{n}) = \int d^2\hat{n} |Y_{lm'}(\hat{n})|^2 = 1.$$

Now, we use the fact that the beam transfer matrix at any  $\phi$  is just a rotation of some arbitrarily set initial beam transfer matrix  $B_{lm'}^{ij}(\phi) = B_{lm'}^{ij}(\phi = 0)e^{im\phi} = B_{lm'}^{ij}e^{im\phi}$ . Using this simplification, we write

$$\begin{aligned} V_m^{ij} &= \sum_{lm'} \int \frac{d\phi}{2\pi} B_{lm'}^{ij} e^{im'\phi} e^{-im\phi} a_{lm'} + n_{ij}(\phi) e^{-im\phi} \\ &= \sum_{lm'} B_{lm'}^{ij} a_{lm'} \int \frac{d\phi}{2\pi} e^{im'\phi} e^{-im\phi} + \int \frac{d\phi}{2\pi} n_{ij}(\phi) e^{-im\phi}. \end{aligned}$$

The first integral gives

$$\int \frac{d\phi}{2\pi} e^{im'\phi} e^{-im\phi} = \delta_{mm'},$$

where the Kronecker delta now sets  $m' = m$ ; the  $m$ th Fourier mode corresponds to the spherical harmonic order  $m$  in this scenario. The second integral is simply the Fourier transform of the noise. With this, we reach a very useful representation of Equation 6.1,

$$V_m^{ij} = \sum_l B_{lm}^{ij} a_{lm} + n_m^{ij}. \quad (6.10)$$

The denotation of this mathematical description as the  $m$ -mode formalism becomes clear, as we now have defined a basis over which all relevant quantities: the visibilities, the beam transfer matrices, the spherical harmonic coefficients, and the noise, can be expressed and calculated for each  $m$ -mode and frequency  $\nu$  independently. Telescopes set limitations on the range of  $m$ -modes needed for this calculation, as not all  $m$ -modes are accessible. This is a huge computational advantage that makes this mathematics feasible on a practical level. Generally, these computations would require one to sum over all  $l$  and  $m$ , but with this formalism we have decoupled equations for each  $m$ -mode which only need to be evaluated for the relevant values of  $m$ .

With this set up, if one were to follow the manipulations shown in Section 6.2, one would arrive at the following expression for computing the spherical harmonic coefficients of the sky from a set of visibilities:

$$\hat{a}_{dirty} = \mathbf{B}^+ \mathbf{N}^{-1} \vec{v}, \quad (6.11)$$

where  $\hat{a}_{dirty}$  could then be used to compute the sky dirty map  $\hat{s}_{dirty}$  with tools like HEALPY.SPHTFUNC.ALM2MAP.<sup>1</sup>

Going one step further, we can perform a simple manipulation of Equation 6.11 by substituting Equation 6.10 for  $\vec{v}$  resulting in

$$\begin{aligned} \hat{a}_{dirty} &= \mathbf{B}^+ \mathbf{N}^{-1} (\mathbf{B} \vec{a} + \vec{n}) \\ \langle \hat{a}_{dirty} \rangle &= \mathbf{B}^+ \mathbf{N}^{-1} \mathbf{B} \vec{a}. \end{aligned} \quad (6.12)$$

To get to Equation 6.12, we used the fact that  $\langle \vec{n} \rangle = 0$ . From this, we can see that  $\hat{a}_{dirty}$ , from which we can calculate  $\hat{s}_{dirty}$ , does not give a true representation of the sky, but instead represents a convolution of that sky with the instrumental response, or as it is often called, the dirty beam. Despite this, the dirty map is still a useful quantity to compute. It preserves all the information contained in the visibilities, meaning that the statistics for parameters quantified from the dirty maps are equivalent to those resulting from the analysis of visibilities directly [55]. The dirty map is a convenient visualization and diagnostic tool for complex data analysis pipelines and, as long as we can quantify the dirty beam, it is a sufficient product.

---

<sup>1</sup><https://healpy.readthedocs.io/en/latest/generated/healpy.sphtfunc.alm2map.html>

## 6.4 CLEAN

Once a dirty map is made, there are two possible approaches. One is to try to deconvolve the map to retrieve a guess of what the true sky  $\vec{s}$  looks like, and the other is to work directly with the dirty map. Both approaches are equally valid and can lead to high quality science results given that we have a precise quantification of the dirty beam. In this section I will focus on the first approach by describing the CLEAN algorithm. CLEAN is a non-linear iterative algorithm commonly used for deconvolving images up to a desired threshold. The following description of the algorithm assumes that a visibility data set has already been measured or simulated.

1. Make a dirty map  $\hat{\mathbf{s}}_{\text{dirty}}$  following Equation 6.11. Also initiate a clean map  $\hat{\mathbf{s}}$  where the detected sources can be stored.
2. Find the brightest pixel in the dirty map with location  $\vec{x}_*$  and flux  $b$ .
3. Create a map  $\mathbf{m}$  so that  $\mathbf{m}(\vec{x}) = 0$  everywhere except at  $\mathbf{m}(\vec{x}_*) = b$ . This is an intermediate product and is overwritten for each iteration.
4. To the clean map  $\hat{\mathbf{s}}$ , add a point source at position  $\vec{x}_*$  with flux  $b$  so that  $\hat{\mathbf{s}}(\vec{x}_*) = b$ .
5. Convolve the point source in map  $\mathbf{m}$  with the instrument's beam  $\mathbf{B}$  so that we have an estimate of the instrument's response to that point source.
6. Subtract this response from the initial dirty map,  $\hat{\mathbf{s}}_{\text{dirty}} = \hat{\mathbf{s}}_{\text{dirty}} - \mathbf{B} * \mathbf{m}$ .
7. Repeat steps 2-6 until the residual dirty map has a maximum amplitude smaller than some pre-determined value.

The pre-determined threshold could, for example, be chosen based on one's knowledge of the noise in the map. Once the threshold is reached, one is left with a *clean*, or partially

deconvolved sky map given by  $\hat{\mathbf{s}}$ , and a residual dirty map  $\hat{\mathbf{s}}_{\text{dirty}}$  containing sources, noise, and artifacts that were below the set threshold. This is a simple algorithm conceptually; however, it poses some practical challenges. The first is that CLEAN is a nonlinear algorithm which makes the process of quantifying one's errors very challenging [56]. The second limitation is that this algorithm is unreliable for images containing extended sources. We assume we have point sources in Steps 2 to 4 and extended emission is never explicitly accounted for. The best outcome we can expect is that after the bright pixels are removed, the remaining emission is detected as a separate source. It is also worth noting that any deviation in our beam model from the true beam will introduce structure into the dirty map that may be picked up as a source if its brightness is above the threshold value. In Chapter 8, these limitations are explored in the context of CHORD simulations.

# Chapter 7

## The Pipeline

With all the machinery set up, I will now describe the simulation and analysis pipeline used for this thesis. The end-to-end simulation is performed with a combination of `RADIOCOSMOLOGY`<sup>1</sup>, an existing CHORD/CHIME code, and newly developed code. In the following sections I will provide specifics on each steps of this pipeline. A schematic of the full pipeline is shown at the end of this chapter (Figure 7.14).

### 7.1 Sky map simulations

The first step is creating a simulated input map, a data cube containing the information of the “true” sky that we will observe and attempt to recover through the entire process. This input map contains multiple components. The ones described in this section are simulated with `CORA`<sup>2</sup>. The HI galaxies that we add from mock catalogs are presented in Section 7.2.

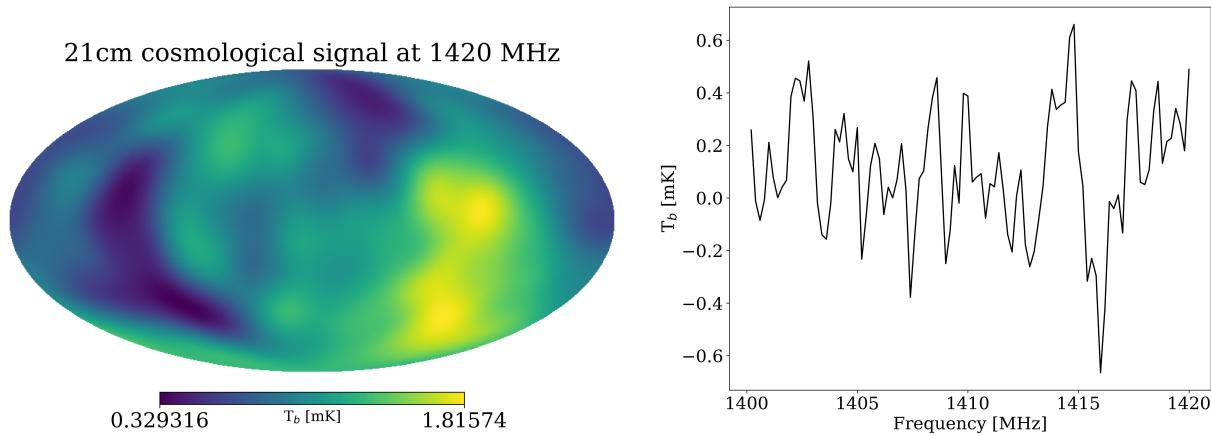
---

<sup>1</sup><https://github.com/radiocosmology/>

<sup>2</sup><https://github.com/radiocosmology/cora>

### 7.1.1 21cm cosmological unresolved background

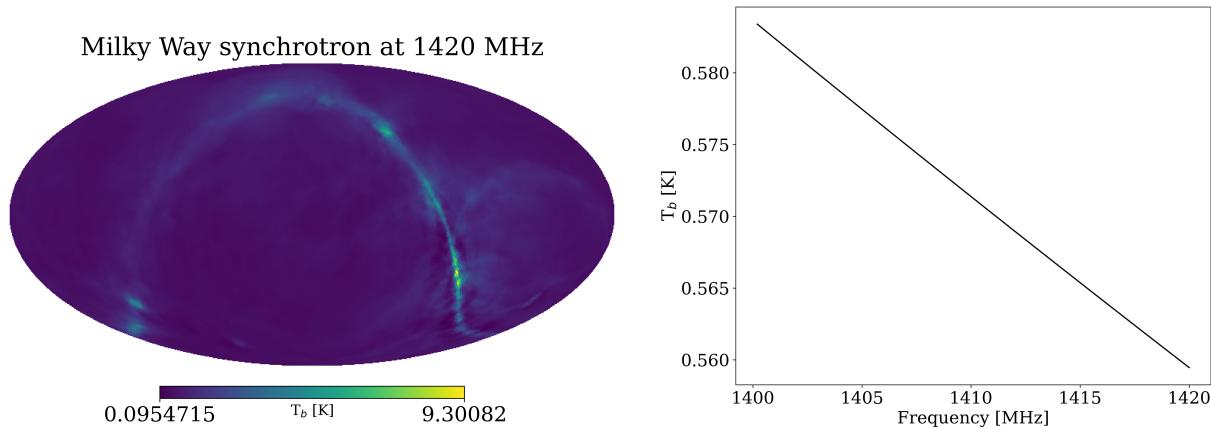
The 21cm cosmological unresolved background is composed of dim HI galaxies that are too faint to be detected individually. It is simulated based on a provided correlation function of cosmological HI brightness temperature fluctuations. One realisation of this field is shown in Figure 7.1, where the left panel shows a Mollweide projection of the spatial brightness temperature fluctuations at 1420 MHz and the right panel shows the frequency dependence of the signal at an arbitrarily selected pixel near the top of the CHORD frequency band. There is large scale structure visible spatially. It is worth noting that the amplitude of this component is in the order of  $\sim 1\text{mK}$ , much smaller than other components and more importantly, fainter than the individual HI galaxies which we are interested in.



**Figure 7.1:** Unresolved 21cm cosmological emission. (Left) Spatial features at 1420 MHz. (Right) Spectral features on the higher end of the CHORD band for an arbitrary pixel. Unlike in Figures 7.2-7.3, the brightness temperature  $T_b$  is given in mK.

### 7.1.2 Synchrotron emission from the Milky Way

Synchrotron radiation is a significant source of radio emission from our galaxy and is caused by relativistic electrons accelerated by magnetic fields. In this pipeline we employ a semi-realistic simulation of the full sky galactic synchrotron emission from the Milky Way. Synchrotron radiation has a smooth power-law spectrum with spectral indices varying spatially [57].



**Figure 7.2:** Synchrotron emission from the Milky Way. (Left) Spatial features at 1420 MHz. (Right) Spectral features on the higher end of the CHORD band for an arbitrary pixel. We see the characteristic galactic plane spatial structure coupled with a smooth frequency evolution.

The simulation uses the 408 MHz Haslam map [58] to constrain the all-sky galactic synchrotron emission at higher frequencies in agreement with a determined spectral evolution of the emission. The map shown in Figure 7.2 uses the spectral indices derived in Miville-Deschenes et al. (2008) [59] with WMAP polarization data. From this process, we end up with a realistic spatial distribution of the Milky Way's synchrotron emission and a simple, smooth frequency dependence at each pixel.

### 7.1.3 Extra-galactic point sources

Extra-galactic point sources include all sources small enough or too far away to be resolved by the CHORD beam, namely each individual source has no spatial extent but contains spectral information. Nonetheless, the population of point sources in the sky does contain spatial information relating to their distribution as observed by surveys. The method used in this pipeline combines data from survey catalogs for bright sources and simulated populations for fainter sources.

The brightest sources with flux densities of  $S_{max} = 10$  Jy and higher at 151MHz are obtained from the NRAO VLA Sky Survey (NVSS) [60] and the VLA Low-frequency Sky Survey (VLSS) [61]. This adds a realistic spatial distribution for the point sources in the simulated true sky map.

Sources with a flux density between  $10 \text{ Jy} < S_{mid} < 0.1 \text{ Jy}$  are simulated using the Di Matteo scalings with a power-law source count function [62], as well as a power law spectral index drawn from a Gaussian distribution for each source. The sources are then placed on pixels drawn from a uniform distribution across the sky. The flux distribution of point sources is given by

$$\frac{dN}{dS} \propto S^{1-\beta}, \quad (7.1)$$

where  $N$  is the number of sources per  $S$  flux bin in mJy and  $\beta$  is the source count index.

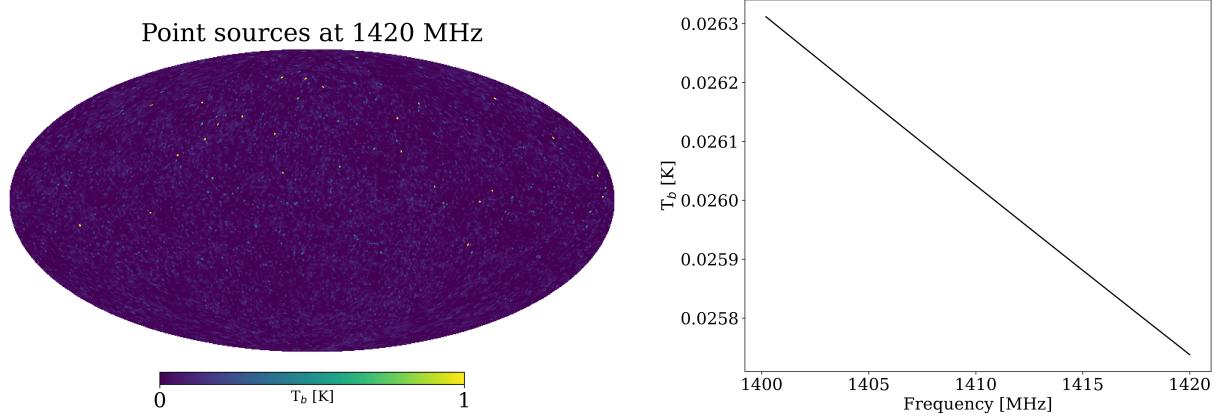
The spectral distribution is described by

$$\frac{dS}{d\nu} = S_{\text{pivot}} \left( \frac{\nu}{\nu_{\text{pivot}}} \right)^{\alpha}, \quad (7.2)$$

where  $dS/d\nu$  is the flux value in mJy for a given spectral  $\nu$  bin. This is proportional to the frequency  $\nu$  normalized by the frequency of the pivot in the power law model  $\nu_{\text{pivot}}$ . The quantity is normalized by  $S_{\text{pivot}}$ , which is the flux computed at the pivot frequency and  $\alpha$  is

the spectral index.

Finally, sources with flux densities smaller than  $S_{min} < 0.1$  Jy are generated using Gaussian correlation functions between frequency slices and drawing from an angular power spectrum to determine their spatial distribution as in [63].



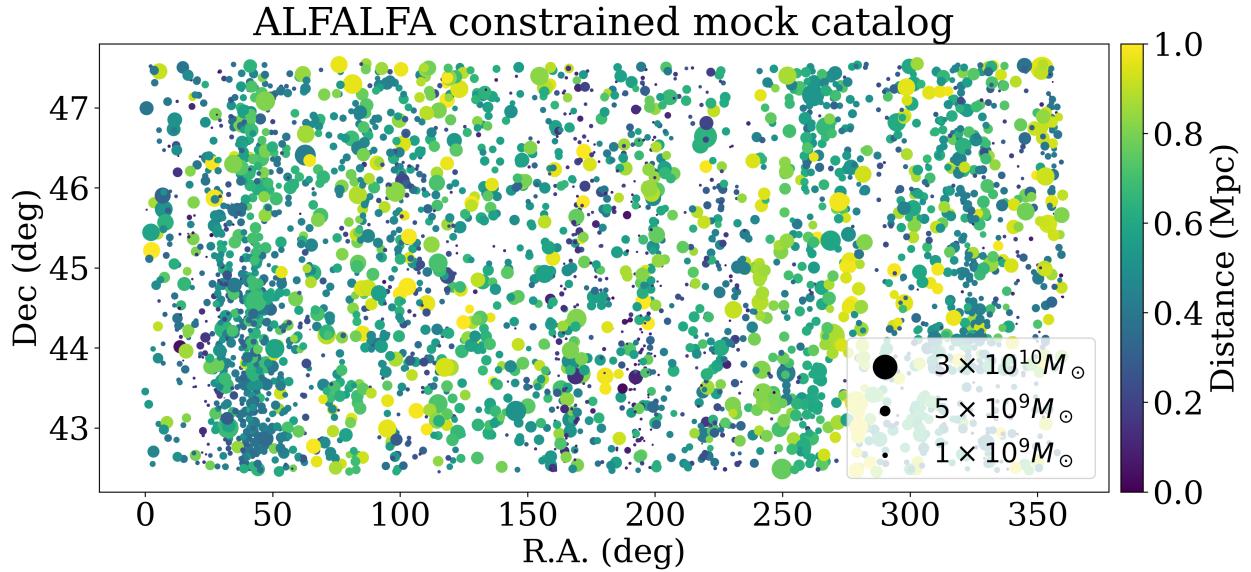
**Figure 7.3:** Extra-galactic point sources signal contribution in the sky. (Left) Spatial features at 1420 MHz. (Right) Spectral features on the higher end of the CHORD band for an arbitrary pixel. The color bar for the brightness temperature  $T_b$  has been modified to show a maximum of 1K for visualization.

## 7.2 Mock HI galaxy catalogs

The main component of the input maps is the HI galaxy population derived from mock catalogs. Unlike the Synchrotron emission and the 21 cm cosmological signal, the galaxies added at this stage will have no spatial extent and will be treated as point sources. It will be a reasonable future addition to this pipeline to include information about the morphology and physical scale of each individual galaxy based on constrained simulations for an additional layer of realism. For the time being, each galaxy consists of a simulated HI profile with spectral features as described in Section 2 that is injected at an individual pixel in the input

map. This step of the pipeline was written specifically to ensure good consistency with CORA and the remainder of the RADIOCOSMOLOGY code infrastructure.

The mock galaxy catalog used for the results of this thesis is an ALFALFA-constrained catalog containing  $N_{\text{gals}} = 3,532$  galaxies centered around Dec = 45°. The stripe occupied by these galaxies has a width of  $\sim 5^{\circ}$ , ensuring that it extends the entirety of the CHORD observing stripe whose width is determined by the primary beam and is  $\theta = 0.21\text{m}/6\text{m} \sim 2^{\circ}$ . We have developed a system to simulate these catalogs that allows for some flexibility in the set-up and properties of the galaxies.



**Figure 7.4:** Spatial distribution of sources from the ALFALFA constrained mock galaxy catalog used in this thesis. The color bar corresponds to the source's distance away from the observer in Mpc. The sizes of the points correspond to the  $M_{\text{HI}}$  of the galaxies. The legend shows representative masses, however, the entire sample does contain a few more massive sources and multiple less massive sources than the ones referenced in the legend.

Firstly, one can choose between the ALFALFA-constrained, flux-limited or volume-limited catalogs. The ALFALFA-constrained catalog follows the results from

Brooks et al. (2023) [64], where they found systematic differences in the HIWF estimated from the fall and spring skies. By comparing the observations with a simulation performed with the Sibelius-DARK N-body simulation and GALFORM, they found that these differences were mostly driven by under- and over-densities in the fall and spring skies, respectively, that were unaccounted for in their original analysis [27]. The catalog used for this thesis is sampled from the HIMF and HIWF produced in this work. This will provide not only good agreements in terms of the HI masses and velocity widths but also a more realistic realization of the spatial distribution of HI sources in the nearby universe due to large scale structure. Figure 7.4 shows the positions of sources in this mock catalog, where the source sizes represent their  $M_{\text{HI}}$  and their colors represent their distance away from the observer in units of Mpc. The legend shows three representative  $M_{\text{HI}}$ , however, there are more and less massive sources in the catalog.

Another important component of the mock catalogs is that they can be centered at a chosen declination stripe and the source population is adjusted in properties and positioning to account for different survey areas or sensitivity differences. Table 7.1 contains a small sample of the mock catalog. The full catalog is composed of a list of galaxies labeled ID = 0, 1, ...,  $N_{\text{gals}} - 1$  with corresponding values for the following properties.

- $M_{\text{HI}}$ : HI mass in units of  $M_{\odot}$ .
- $V_{\text{HI}}$ : maximum HI velocity in km/s.
- $i$ : inclination of the galaxy with respect to the observer in radians.  $i = 0$  corresponds to a face-on galaxy and  $i = \pi/2$  corresponds to an edge-on galaxy.
- $D$ : distance to the galaxy in Mpc.
- $W_{50}$ : half-width of the HI profile in km/s.

- $z$ : redshift
- ra: Right Ascension of the source in degrees
- dec: Declination of the source in degrees

ID	$M_{\text{HI}}$ ( $M_{\odot}$ )	$V_{\text{HI}}$ (km/s)	$i$ (deg)	$D$ (Mpc)	$W_{50}$ (km/s)	z	R.A. (deg)	Dec (deg)
0	$1.9 \times 10^9$	46.3	67.5	77.4	100.2	0.018	50.9	47.2
1	$3.0 \times 10^9$	51.5	46.8	53.2	141.2	0.012	12.2	46.5
2	$2.7 \times 10^9$	53.7	48.3	91.9	143.8	0.021	70.1	44.5
3	$7.0 \times 10^8$	39.0	69.3	51.2	83.4	0.012	86.6	44.4
4	$5.6 \times 10^9$	6.8	16.7	174.1	47.2	0.039	3.4	45.8
5	$2.2 \times 10^9$	81.3	73.2	79.5	169.8	0.018	3.7	47.3
:	:	:	:	:	:	:	:	:

**Table 7.1:** Sample of mock HI galaxy catalog. The properties shown are explained in the text and their values have been rounded for clarity. The full catalog from which these galaxies were taken contains  $N_{\text{gals}} = 3,532$  galaxies and was simulated to cover a stripe centered at Dec = 45° of approximately 5° in width.

Frequency spectra, the observables of HI galaxies, can be simulated from this empirically-constrained catalog. This process involves randomly sampling parameters  $b_1, b_2$  from [0, 1) and  $c$  from [0, 3] and fixing  $n = 2$  to construct a Busy function as described in Section 2.1.1. These parameters control the overall shape of the Busy function, the heights of the peaks and the dominance of the parabola component. For this catalog,  $x_e$  and  $x_p$  are both set to zero. The centering of the profiles is instead made by determining the redshift of the sources given their distance away from the observers in agreement with Brooks et al. (2023) [64]. Both parameters  $w$  and  $a$  are estimated based on the ALFALFA HIMF and HIWF. We first sample  $V_{\text{HI}}$  from the HIWF and  $i$  from a uniform distribution between  $[0, \pi/2]$  and use

$V_{\text{HI}} \approx w/2 \sin(i)$  to estimate  $w$ . Then, provided the overall shape and width of the spectra,  $a$  is selected so that the integral of the profile  $S_\nu$  is in agreement with the necessary  $M_{\text{HI}}$  sampled from the HIMF given

$$M_{\text{HI}} = 2.356 \times 10^5 \frac{D^2}{1+z} \int S_\nu \, dV. \quad (7.3)$$

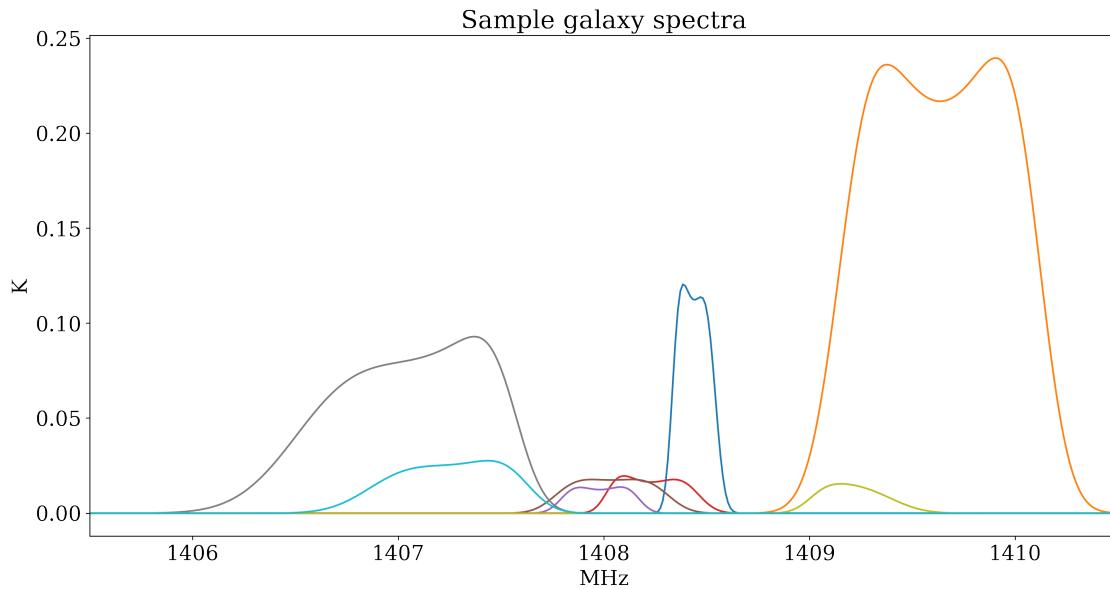
Furthermore, observations have established that a variety of physical mechanisms can inject kinetic energy into galaxies leading to a minimum characteristic velocity dispersion of  $10 \pm 2 \text{ km/s}$  [65, 66, 67]. For star forming galaxies (which are well represented in HI samples), this effectively serves as a lower limit for the width of the HI profile. This is implemented by convolving the spectra with a 10 km/s full width half maximum Gaussian.

The spectra simulated in this way have units of Jy for each HI velocity in km/s. Alternatively, we can represent this as brightness temperatures in K as a function of observed frequency in MHz. The conversion from these sets of units is achieved through two relations. Equation 2.2 is applied to the velocity axis when it is centered at 0 km/s to retrieve a frequency axis and the shift to the central frequency corresponding to the redshift  $z$  is done afterwards. We then use the Rayleigh-Jeans approximation (valid for low frequency observations) to convert the signal to units of temperature (K) depending on the wavelength of the observation  $\lambda$  and the angular resolution of the instrument  $\theta \approx \lambda/b_{\max}$  with

$$T_b = \frac{S_\nu \lambda^2}{2k_B \theta^2}. \quad (7.4)$$

Figure 7.5 shows a sample of simulated HI galaxy spectra over a section of CHORD's frequency band. The variety in shapes and amplitudes correspond to the range of properties described above, from inclination  $i$ ,  $M_{\text{HI}}$ ,  $z$  and  $V_{\text{HI}}$ . These profiles are then injected into input sky maps by slicing through individual pixels along the frequency direction. They can either be combined with the other sky components (see Section 7.1) or be left on their own

before performing the observation of the sky with the desired instrument set up.



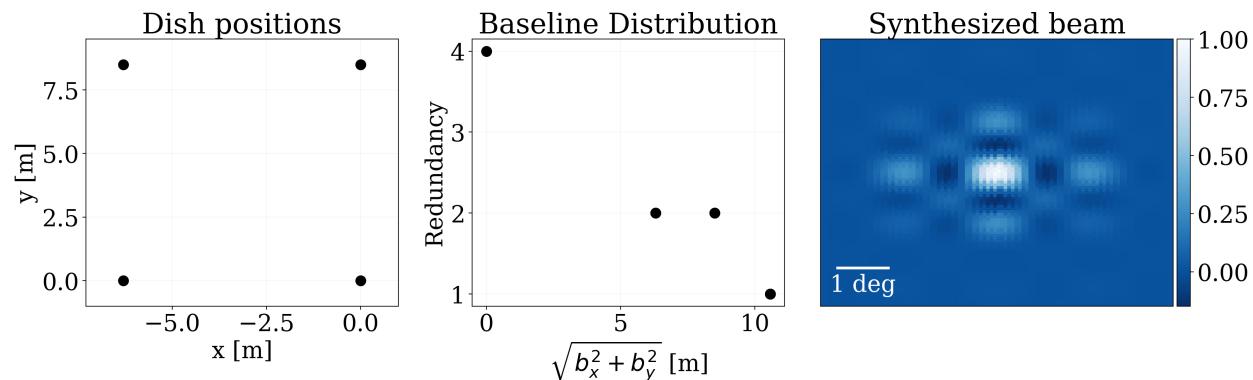
**Figure 7.5:** A sample of simulated HI galaxy spectra shown over a section of CHORD’s frequency band. The variety of shapes and amplitudes correspond to the differences in physical and observational parameters outlined throughout this section and in section 2.1.1.

### 7.3 Instrument set-up: beams

Once the input maps have been created, including diffuse emission, extragalactic point sources, and the HI galaxies, the next step is to observe this sky. In order to do this, a telescope object is first set-up containing information about the number and exact configuration of the dishes as well as their sizes. We also provide frequency specifications: the total bandwidth and frequency resolution. Even though CHORD is a driftscan interferometer, a pointing still needs to be specified. This will correspond to the Declination on which CHORD will be centered as it observes across the entire sky (all Right Ascensions) to create the observing stripe. With all of these determined, a telescope

object is created with DRIFTSCAN<sup>3</sup>.

The left panel of Figure 7.6 shows an example configuration for a  $2 \times 2$  array. The middle panel shows the distribution of the baselines, first defined in Section 5.1. Namely, it shows the redundancy of each baseline length for the given array configuration. In this instance there are 4 length-0 baselines (auto-correlations for each antenna), four total mid-length baselines and a single baseline longer than 10 m. The redundancy of each baseline provides a boost to the sensitivity of the instrument on the scale that it probes. Intuitively, if an instrument was composed of a mostly compact array with one or two very long baselines, it would theoretically be able to access very small scales but practically its sensitivity to those scales would be very poor.



**Figure 7.6:** The left panel shows the distribution of dishes for a  $2 \times 2$  array. The middle panel shows the number of baselines of each given length. The right panel shows the simulated synthesized beam at 1418 MHz for this array. The color bar shows the relative sensitivity of the beam as a fraction of the main lobe sensitivity. A scale is shown to provide a sense of the size of the features.

With all of this combined, we can simulate the response of the instrument to the sky, the beam transfer matrices. Applying this transfer function to a single point source with

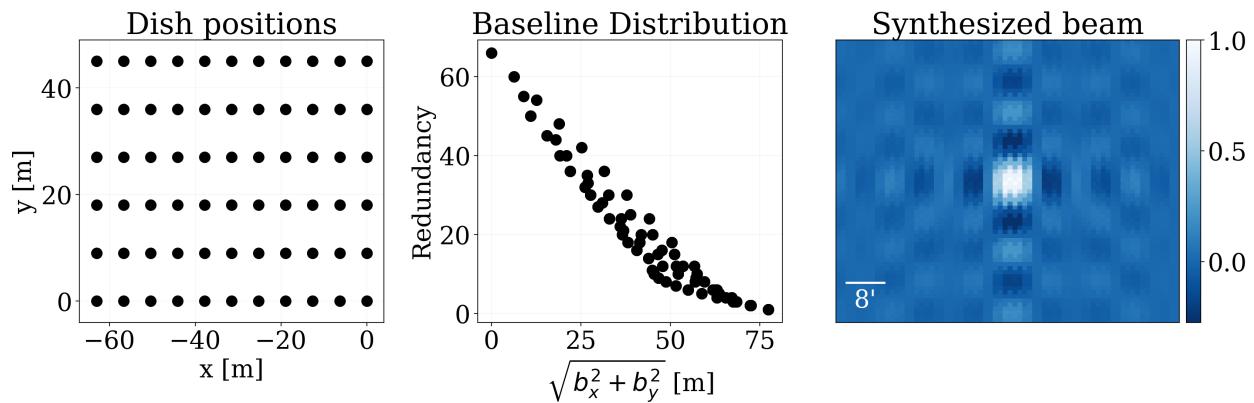
---

<sup>3</sup><https://github.com/radiocosmology/driftscan>

DRIFTSCAN we obtain the synthesized beam. The right panel of Figure 7.6 shows the synthesized beam for a  $2 \times 2$  array at 1418 MHz. At this frequency, the bright center or main lobe has an angular extent of  $\theta_{\min} \sim 1^\circ$ . We calculate that this is indeed representative of the longest baseline, as  $\theta_{\min} \sim 21\text{cm}/10\text{m} \sim 1.2^\circ$ , confirming the approximation postulated in Section 5.1. Objects smaller than this extent will not be resolved by the beam and will be observed as point sources. In the peripherals of the main lobe, there are side lobes with a considerable fraction of the sensitivity of the main lobe (up to 30%). The effects of this will become apparent in Section 7.6.

Figure 7.7 shows the configuration of a CHORD pathfinder-like instrument. There is a  $6 \times 11$  grid with 6 dishes with a 6.3 m separation in the EW direction and 11 dishes with 8.5 m separation in the NS direction. The middle panel shows the redundancy of various baseline lengths and unlike for the  $2 \times 2$  array, there is a more continuous coverage of baseline lengths with the largest one being  $b_{\max} \sim 77\text{m}$ . The right panel shows the synthesized beam for this set up with the major difference being the much smaller scale of the main lobe with extent  $\theta_{\min} \sim 8'$  compared to  $\theta_{\min} \sim 1^\circ$  for the  $2 \times 2$  array. In a similar order of magnitude calculation, we estimate  $\theta_{\min} \sim 21\text{cm}/77\text{m} \sim 9'$  which is again in agreement with the apparent size of the main lobe shown. A more subtle difference is that the next brightest side lobe is only up to 20% as sensitive as the main lobe compared to 30% for the smaller array, with an even larger difference in the EW direction.

Similarly, going from the CHORD pathfinder to the full-CHORD, we will have approximately a  $21 \times 24$  array (in the core) so the longest baseline will be  $b_{\max} = \sqrt{(21 \times 6.3\text{m})^2 + (24 \times 8.5\text{m})^2} \approx 243\text{m}$  and so  $\theta_{\min} \sim 21\text{cm}/b_{\max} \sim 0.05^\circ$  or  $3'$ . If we take the outriggers into consideration, the longest baseline is significantly larger, at  $b_{\max} \sim 3,300$  km. The angular resolution would be much better in this case, going to  $\theta_{\min} \sim 21\text{cm}/b_{\max} \sim 0.01''$ . For the rest of the results shown in this thesis, the CHORD



**Figure 7.7:** Same as Figure 7.6 but for the CHORD pathfinder configuration. There are 6 dishes with a 6.3 m separation in the EW direction and 11 dishes with 8.5 m separation in the NS direction.

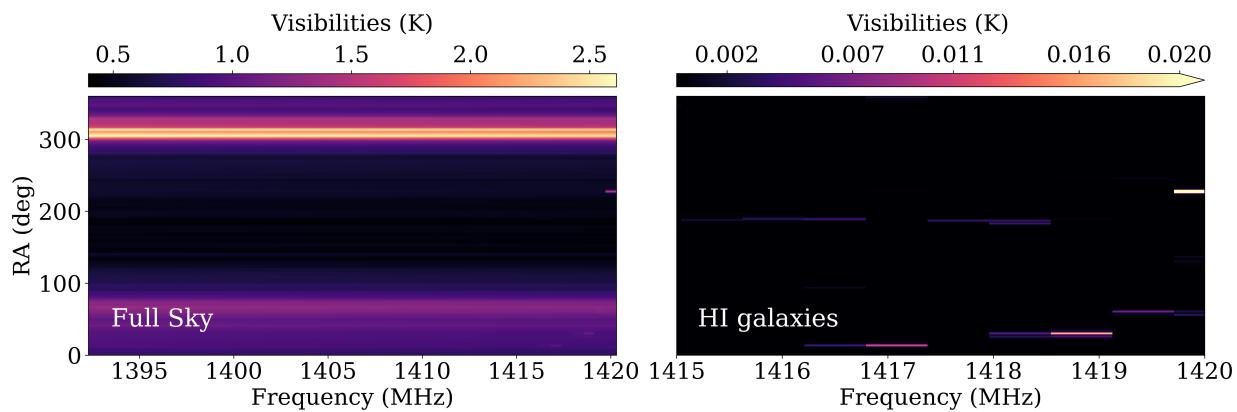
pathfinder configuration is assumed.

The synthesized beam, referred to as dirty beam in Chapter 6, will add complex structure to the observations performed by this instrument. In a map of the observed sky, any given pixel will then not only include information from sources located at that pixel but also from sources located at pixels that lie within side lobes. This will lead to spatial aliasing that can complicate the process of detection as well as interfere with proper localization and determination of source properties from the output maps. A deep understanding of the instrument and its beam is therefore of utmost importance to reading and analyzing maps created by radio interferometers. A depiction of the effects of incorrectly determining the synthesized beam for detection is given in Section 8.2.

## 7.4 Visibility Simulation

With the beam transfer matrices and the pristine input sky maps set up, the next step in the pipeline is to simulate CHORD's observation of the sky. As described in Chapter 5, for

a radio interferometer, this involves simulating the visibilities. With DRACO<sup>4</sup>, the m-mode formalism is used to compute the visibilities of the input sky map using Equation 6.10. First, in a process similar to that used to compute Figure 5.1, the scales that the instrument is sensitive to ( $m$  and  $l$ ) are estimated. Then, the coefficients of the spherical harmonic expansion of the sky  $a_{lm}$  are calculated and the beam transfer matrices are applied to them, resulting in simulated visibilities. For the time being, these are noiseless, though thermal noise is added at a later step in the pipeline (Section 7.5).



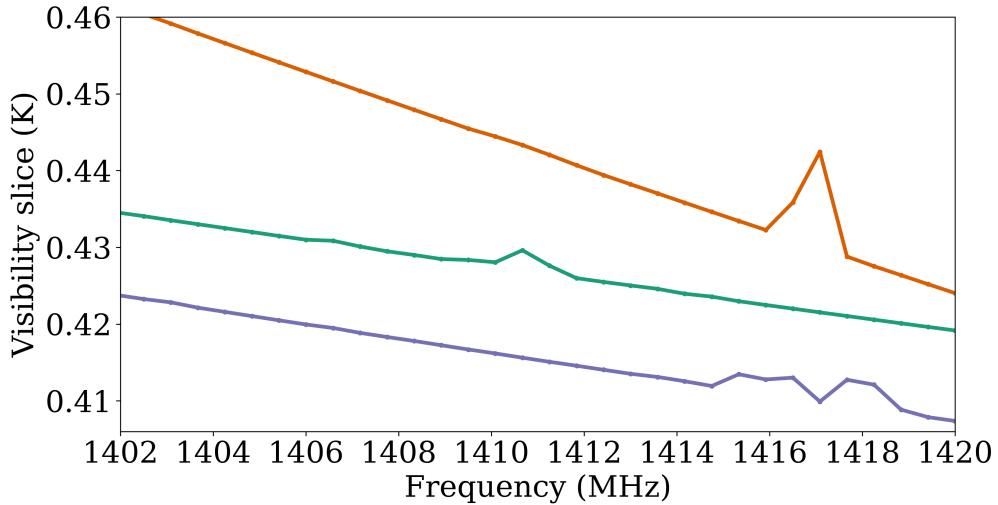
**Figure 7.8:** Visibilities of a full sky map (left panel) and an HI galaxies only map (right panel) for a single baseline as observed by a pathfinder-like array. The contribution from HI galaxies has been boosted by a factor of 10 in the full sky map for illustration purposes.

The data products at this step are complex arrays of shape  $(N_{\text{freqs}}, N_{\text{baselines}}, N_{\text{time}})$ . Here, the last dimension  $N_{\text{time}}$  corresponds to the number of time samples which for a driftscan telescope represents both the passage of time and the subsequent measurements of different patches of the sky along the observing stripe. The left panel of Figure 7.8 shows the visibilities for a sky containing HI galaxies, the 21 cm cosmological signal, synchrotron emission and other extra-galactic point sources for a single baseline. We see that as the telescope scans

---

<sup>4</sup><https://github.com/radiocosmology/draco>

across the sky, it observes two brighter regions which correspond to the galactic plane. Overall, the frequency evolution (shown in the x-axis) is fairly smooth for all times or RA values (shown in the y-axis) as expected, given that the dominant sources of emission have slow frequency dependencies. Due to the much fainter nature of the HI galaxies, their emission has been boosted by a factor of 10 for illustration purposes. At the higher end of the frequency band, signatures from these individual HI galaxies become apparent as compact spectral regions of higher emission (for example, see RA  $\sim 220^\circ$  near 1420 MHz). The right panel shows the isolated visibilities for the HI galaxies limited to the higher end of the frequency range and the emission has been returned to its original intensity. Their higher variability in intensity as a function of both RA and frequency is very clear when compared to the other components.



**Figure 7.9:** Slices along 3 arbitrary RA values from the full sky visibilities. The emission from HI galaxies in the map has not been boosted. Amongst the smooth spectral dependencies of the dominant signal, the spectral shapes of the galaxies are evident even for faint sources.

In Figure 7.8, it may seem that the HI galaxies are getting buried by the more dominant

emission from synchrotron radiation or extragalactic sources. Although the overall amplitude of these sources is much smaller, their characteristic spectral structure, in contrast to the smooth dependency of the other components, is a means of disentangling them. This is made clear in Figure 7.9, which shows three slices along arbitrary RA values of the full sky visibilities. The smooth frequency dependence is seen as an overall trend and the position dependent spectral index described in Section 7.1 is visible through the varying slopes of the three slices. Most importantly, the presence of HI galaxies is evident in all three cases, including a very faint second source with a peak at around  $\nu \approx 1406$  MHz on the slice shown by the green line. The emission from HI galaxies was not artificially boosted for Figure 7.9. Note that CHORD’s coarse spectral resolution was used for these particular observations and so more detailed spectral features are lost.

## 7.5 Noise and Calibration Systematics

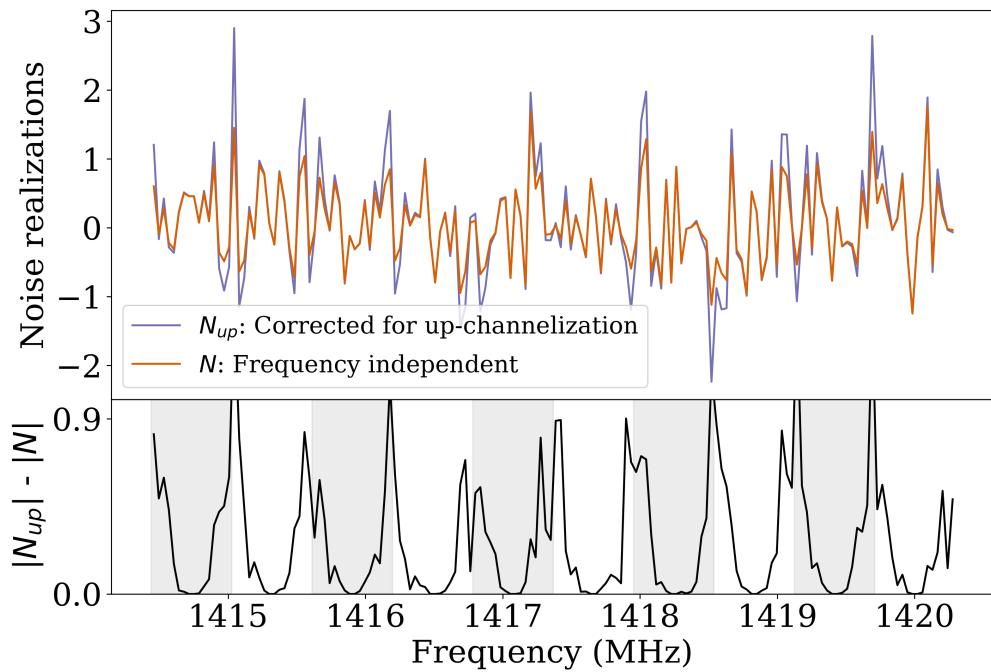
The pipeline constructed and presented in this thesis is flexible so that simulated observations can contain chosen levels of realism. The simplest observations will only contain systematics due to the beam structure with a uniform weighting of the instruments baselines but the weighting can be changed and noise can be added. The selected up-channelization scheme (Section 5.2) will also result in additional spectral aliasing. In this section, we explore the addition of thermal noise and its interplay with up-channelization as well the effects of potential calibration errors. Both of these add an additional layer of realism to the pipeline simulations.

### 7.5.1 Thermal noise

Visibility measurements will intrinsically contain thermal noise with a standard deviation given by  $\sigma_{\text{rms}}$  in Equation 5.3. The statistics of the noise are independent from frequency channel to frequency channel as they each get an uncorrelated draw of the noise. However, in Section 5.2 the process of up-channelization is introduced and the necessity for normalizing the data to remove the modulation at the edges of coarse channels is highlighted. As shown in Figure 5.3, the renormalization can lead to a significant boost in the signal at certain channels. For realistic noisy visibilities, this will not only boost the signal but will ultimately also increase the noise variance. The dataset is therefore left with a thermal noise that has frequency dependent variance.

In the pipeline, the first step is to create Gaussian thermal noise with  $\sigma_{\text{rms}}$  as in Equation 5.3. The user is free to modify the value of  $\Delta t$  by setting both the number of days of the observation and the sampling along the sky plane. For  $\Delta\nu$ , the value is established once the up-channelization scheme is selected and the updated spectral resolution is determined. The system temperature can also be modified but the default is set to  $T_{\text{sys}} = 30\text{K}$ . The baseline component  $N(N - 1)$  is also fixed once the array set-up is chosen for the beam transfer matrix computations. The simulated noise at this point has shape  $(N_{\text{freq}}, N_{\text{baselines}}, N_{\text{time}})$ . The second and final step to having appropriately estimated noise variance is to apply the normalization vector along the frequency dimension.

Figure 7.10 shows the resulting frequency evolution for the noise variance of a 30 day observation of a pathfinder-like array with a spectral resolution of  $\Delta\nu = 36.59 \text{ kHz}$ . This corresponds to an up-channelization factor of  $U = 16$ . The orange line shows a noise realization that does not account for the effects of the up-channelization envelope. The purple line uses the same random seed to simulate the realization, but this time the variances are normalized by the vector shown in Figure 5.3. The bottom panel shows the



**Figure 7.10:** Frequency evolution of noise variance. The orange line shows thermal noise free of up-channelization effects while the purple line accounts for the modulation correction. The bottom panel shows the different in variance for the two cases with the white and gray alternating to represent the coarse channels.

difference between their variance as a function of frequency. Notice that the thermal noise that has been corrected for the effects of up-channelization is always larger or equal to the frequency independent one. The alternating gray and white regions mark coarse channel boundaries. As expected, the largest differences are found at the edges of the channels where the modulation effects we correct for were most significant.

### 7.5.2 Calibration errors

Another systematic that affects interferometric measurements are calibration errors. The instrument measures electric fields and in the process of converting the induced voltage into

physical units of Kelvin, the normalization factor (or gain) is applied to the raw visibilities. For a pair of antennas  $i, j$ , the measured visibilities can be mathematically written as

$$V_{ij}^{\text{meas}} = g_i g_j V_{ij}^{\text{true}} + n_{ij}. \quad (7.5)$$

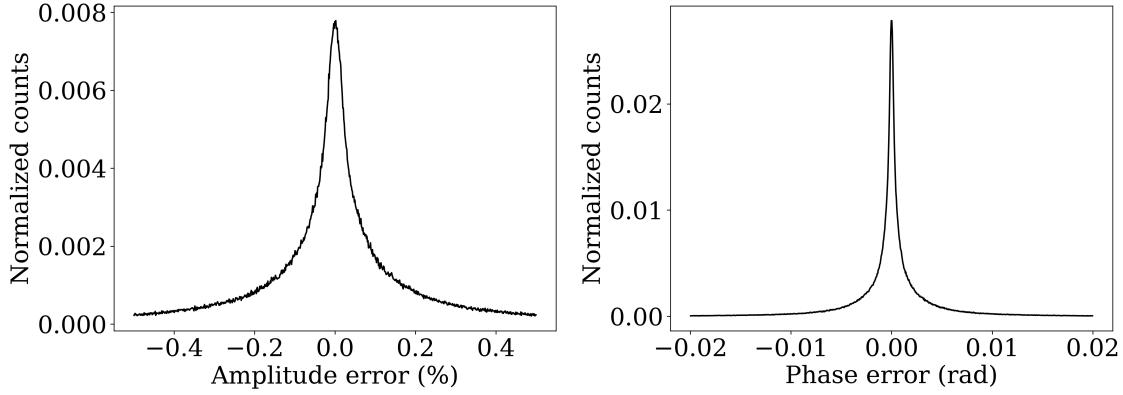
Here,  $g_i, g_j$  is a complex number with amplitude  $A$  and phase  $\theta$  described by

$$g_i g_j = A e^{i\theta}. \quad (7.6)$$

The process of calibration involves estimating  $g_i g_j$  to recover the true visibilities. We have made the simplifying assumption that the noise  $n_{ij}$  is injected into the measurement after the instrumental effects that make it necessary to calibrate take place. In reality, most stages of the measurement will add some level of thermal noise.

The estimation of these gains  $g_i g_j$  heavily depend on the knowledge of both instrumental parameters and the observed sky. Limitations to that knowledge give rise to calibration errors [68]. For example, for redundant calibration, antennas are assumed to produce identical primary beams and their exact location with respect to one another is assumed to be fully known [69]. Any deviations from the ideal redundant case will produce sub-optimal calibration solutions. We estimated the probability distribution of phase and amplitude calibration errors for CHORD assuming the use of CorrCal [70]. The distributions are shown in Figure 7.11 and notably, the errors are non-Gaussian, containing long tails to either side.

In the pipeline, calibration errors are implemented through first sampling phase and amplitude errors ( $A_{ij}^{\text{err}}, \theta_{ij}^{\text{err}}$ ) from these distributions. The visibilities output from the pipeline are originally assumed to be perfectly calibrated. The sampled errors are then applied to



**Figure 7.11:** CHORD estimated calibration error distribution. The left panel shows the amplitude percent error distribution and the right panel shows the phase error distribution in radians. Notably both are non-Gaussian with extended tails.

those visibilities, resulting in measured visibilities given by

$$V_{ij}^{\text{meas}} = G_{ij} V_{ij}^{\text{true}}, \quad (7.7)$$

where

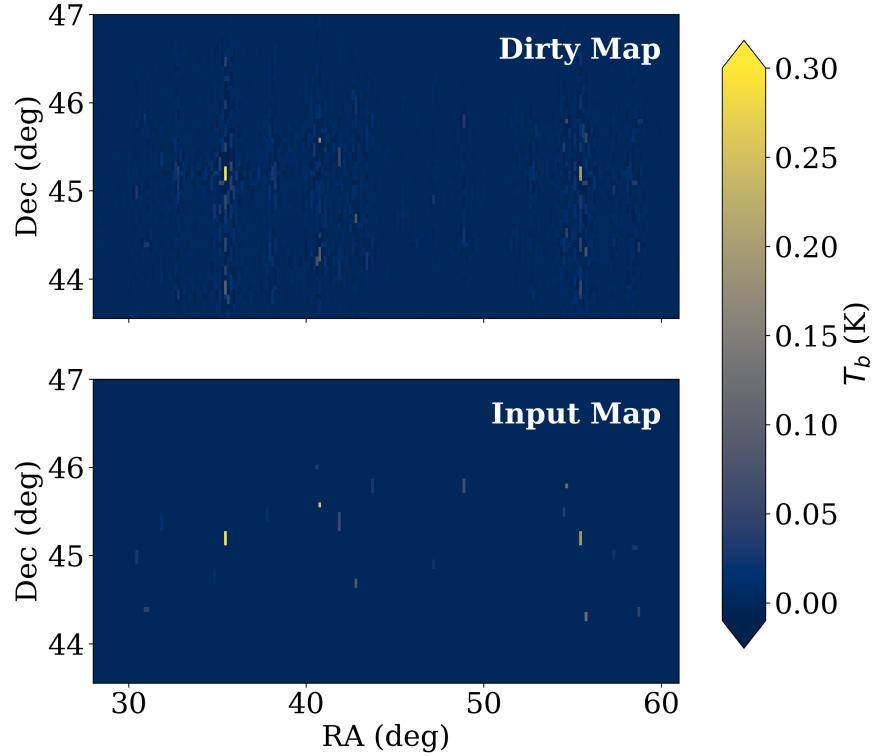
$$G_{ij} = \left(1 + \frac{A_{ij}^{\text{err}}}{100}\right) \times \exp(i\theta_{ij}^{\text{err}}). \quad (7.8)$$

We do this prior to injecting noise such that if one were to subtract the calibrated visibilities from the true ones, the distributions from Figure 7.11 would be recovered.

## 7.6 Dirty maps

For the science case of HI galaxy search, an image space representation for the data products is useful as a lot of the detection methods are applied in this basis. Therefore, as described in Chapter 6, dirty maps are computed from the visibilities using m-modes as shown in Equation 6.12. Now, this equation requires a particular choice for the noise covariance matrix  $\mathbf{N}^{-1}$ . This matrix encodes information about the sensitivity of the instrument for

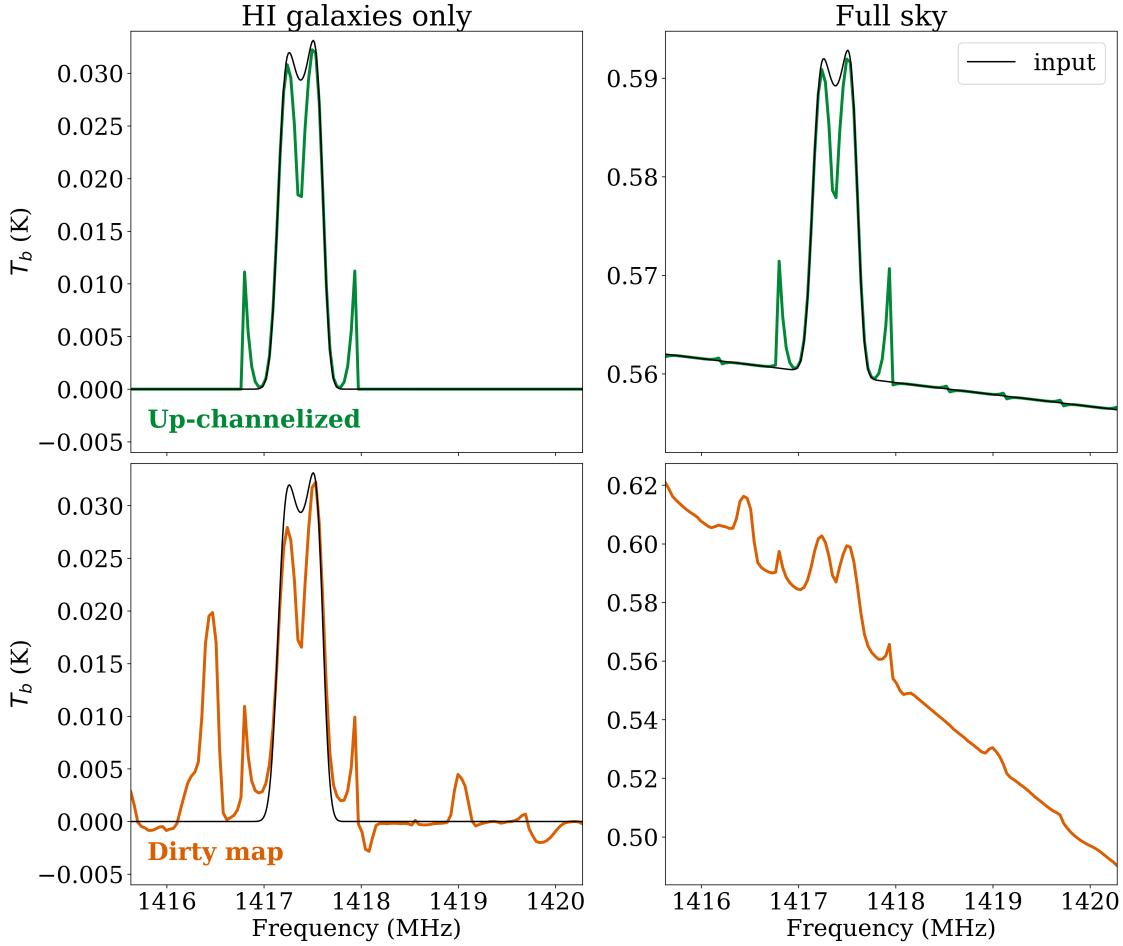
a particular observation and is used to weight the data appropriately. For instance, in redundant arrays,  $\mathbf{N}^{-1}$  is such that baselines with higher redundancies are weighed higher due to the better sensitivity to the scales they probe and those with a low redundancy are down-weighted. Regardless of the specific set up, an  $\mathbf{N}^{-1}$  weighting provides maps with optimal noise properties.



**Figure 7.12:** The top panel shows the dirty map computed from a simulated observation of an input sky containing only HI galaxies at 1418 MHz. Various features arise in the dirty map due to the synthesized beam structure.

Figure 7.12 shows a cut-out of a dirty map simulated with DRACO and an input map containing only HI galaxies at  $\nu = 1418.0$  MHz. In the bottom panel, the true sources are seen distributed across the patch of sky, containing brightness temperatures ranging up to  $T_b \sim 0.4$  K. The top panel shows the corresponding dirty map, where the positions of bright

sources are still in apparent agreement by visual inspection, however, there is an entire new layer of spatial features not present in the input map. In particular, some of this spatial aliasing is even brighter than true sources. For example, at around (RA, Dec) = (35, 44) degrees, the aliasing from the brightest galaxy located just above that is significantly brighter than multiple of the fainter sources in this map.



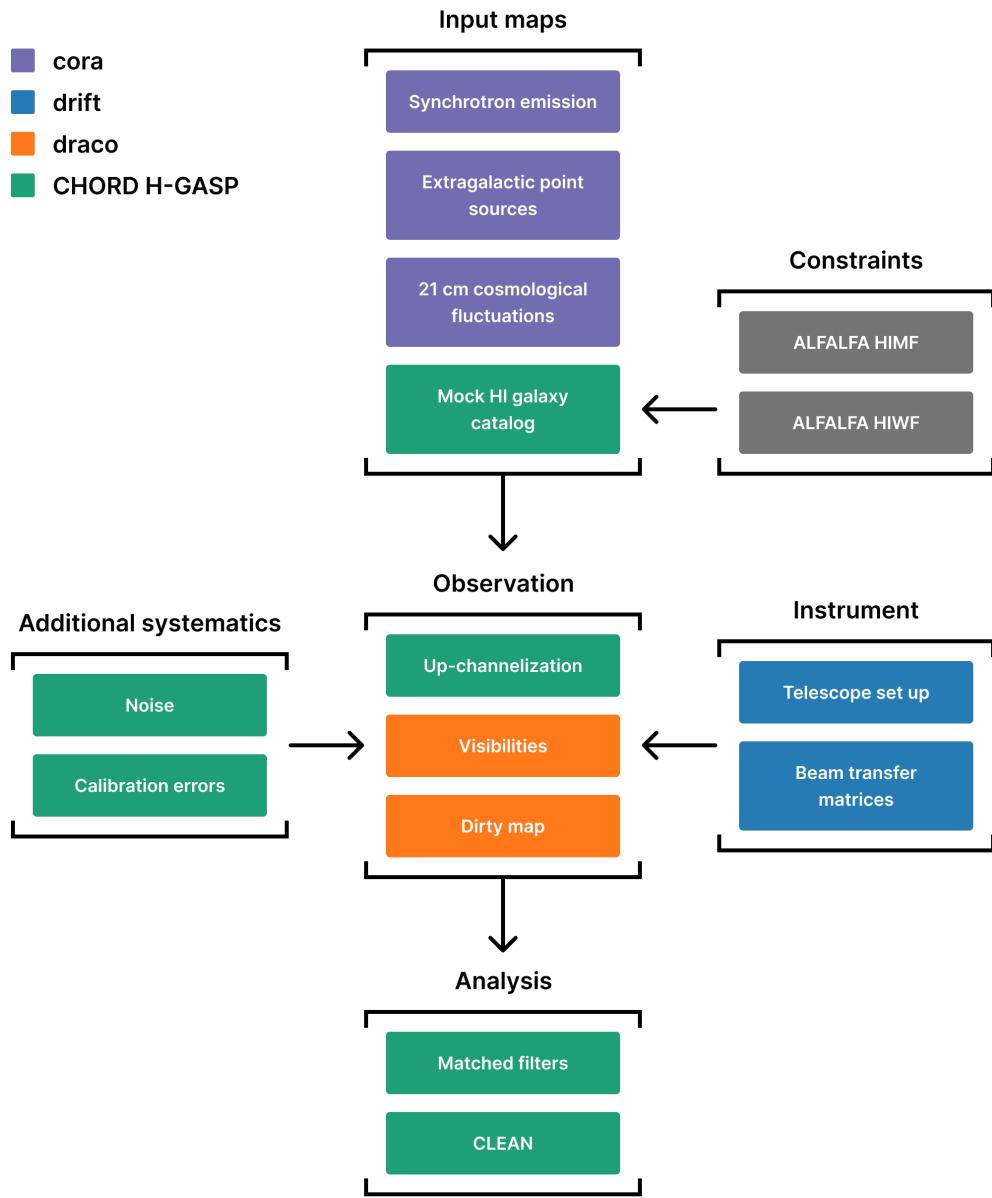
**Figure 7.13:** HI profiles resulting from the pipeline. The top panels show an HI galaxy and a full sky map after up-channelization. The lower panels show the same profiles after the full observation process and dirty map computation. Characteristic features of spectral and spatial aliasing are further discussed in the text.

A different view into the effects of spatial aliasing is seen by slicing through one pixel of the dirty map across all frequencies and analyzing the resulting profile. That is precisely what we do in Figure 7.13. The left column contains maps of HI galaxies only, whereas the right column includes synchrotron emission, extragalactic point sources, and the 21 cm cosmological emission as well as HI galaxies. The difference between the top and the bottom rows is that the top row contains maps that have been up-channelized with a factor of  $U = 16$  resulting in spectral aliasing but have not been put through the CHORD beam and so contain no spatial aliasing. The bottom row contains both effects as the profiles have been extracted from dirty maps of up-channelized observations. For panels A, B, and C, the black line delineates the pristine galaxy profile found at this pixel prior to any systematics.

In panels A and B of Figure 7.13, the characteristic features of up-channelization are apparent, with sharp, un-natural peaks surrounding the central source at equal intervals to either side. Similar to Figure 7.9, in panel B, the sky components other than the HI galaxies provide an overall boost in the signal and a power-law trend to the profile but the HI galaxy features are still very much distinguishable. Panel C shows the slice through a dirty map containing up-channelized HI galaxies only. In addition to the features observed in panel A, there seems to be excess emission at frequencies where there were originally no sources. This is the signature of spatial aliasing – emission from bright sources leaking into nearby pixels due to complex beam structures. There is now also negative emissions and this is to be expected given the negative regions surrounding the main lobe in the synthesized beam as shown in Figures 7.6 and 7.7. Finally, panel D demonstrates the case for a map containing all of the sky components with a combination of spatial and spectral aliasing. The increase in the slope can be understood with reference to the process of map-making. The pixel we cut through in the dirty map to make this plot contains in fact a linear combination of the emission from all nearby pixels, where how much each pixel contributes to the total

is determined by the precise shape of the beam. Most of the pixels will be brighter at lower frequencies than at higher frequencies due to the power law dependence of synchrotron emission. Summing up many bright pixels at the lower frequency compared to many fainter pixels at higher frequencies results in an increase in the slope.

Given the intricacies of these simulations, it is notable that the overall shape of the original HI profile is visible even in panel D. All of the systematics described thus far are often seen as obstacles in the pursuit of detecting and correctly categorizing HI galaxies for the CHORD pathfinder catalog. Spectral aliasing produces sharp features and suppresses the signal at the edges of coarse channels. In an attempt to recover some of that signal, the noise injected into the observations by the instrument is imprinted with additional spectral structure. Spatial aliasing causes significant leaking amongst pixels that can disguise itself as additional HI sources or confuse the detection of real HI galaxies. Crucially, these effects are all predictable. Up-channelization is a deterministic process that can be analytically described, and the features from dirty maps can be dealt with given one has a good understanding of the instrument’s synthesized beam. We could therefore leverage these in the detection process to recover as much information as possible from these simulations and eventually from CHORD data. In the next chapter, some tools for HI recovery are explored in the context of these systematics.



**Figure 7.14:** Schematic of the complete pipeline. The colors denote which packages are used for each step. CORA, DRIFT, and CAPUT are from the existing CHIME/CHORD infrastructure. CHORD H-GASP: HI Galaxy Simulation Pipeline is the package developed for this thesis. Different empirical constraints and systematics are included for more realistic outputs. The results are analysed using two main algorithms for source detection.

# Chapter 8

## HI galaxy recovery

### 8.1 Spatial matched filter

In this Chapter we use different tools to detect HI galaxies from the outputs of the pipeline and explore their performance based on the systematics we know to be present in the data. For the first section, we focus on the search for sources in the spatial domain using a matched filter. Matched filters are used to extract signals from noisy data in a variety of physics applications [71, 72, 73]. This approach requires pre-existing knowledge of the signal’s overall shape and relies on there being differences between the properties of the signal and the noise in the data [74]. As demonstrated in Chapter 7.6, dirty maps have characteristic features due to the instrument’s synthesized beam that are deterministic in nature. We can reliably predict the shapes of these features given the particular configuration of the interferometer. Matched filters are therefore a suitable approach to source detection in dirty maps.

Matched filters look for signals in data using a template that follows the predicted shape of the signal. Given that parameters like the amplitude and position of the signal are unknown, these parameters are not fixed *a priori* in the template, but rather are characterized after the

potential detection is made [75]. Conceptually, the template slides across the data recording how well it describes the signal (or lack thereof) at each position. If the template is a poor fit for the data at a particular position, the output will have a small amplitude. If there is good alignment between the template and the data, the output will have a large amplitude signifying a high-significance detection of the predicted signal at that position. In this way, matched filtering acts as a way to track the likelihood of finding a signal of some characteristic shape at each point in the data vector.

Say we have some data described by  $d(\vec{x}) = s(\vec{x}) + n(\vec{x})$  with some signal  $s$  and a noise component  $n$ , both a function of the position vector  $\vec{x}$ . Then, mathematically, the matched filter output  $y(\vec{x})$  is expressed as a convolution between  $d(\vec{x})$  and the template  $t(\vec{x})$  written as

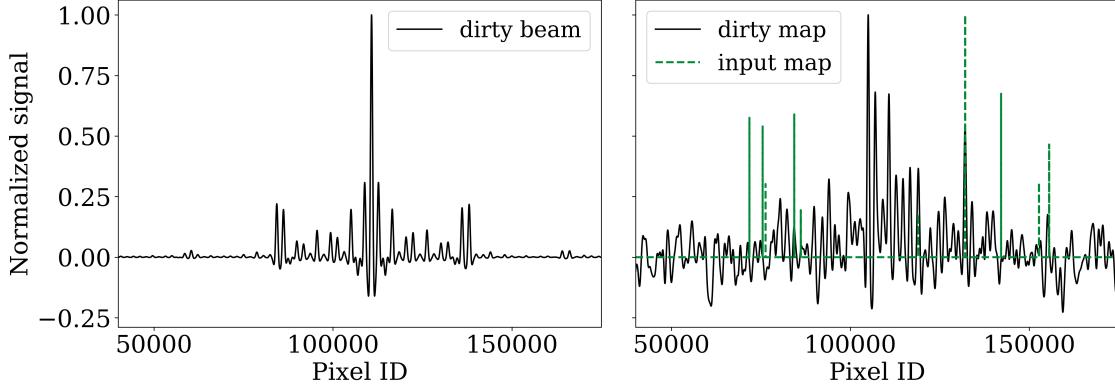
$$y(\vec{x}) = d(\vec{x}) * t(\vec{x}). \quad (8.1)$$

For large datasets, this operation can be very computationally expensive. Invoking the convolution theorem, which dictates that a convolution in position space is equivalent to a simple multiplication in Fourier space, this can be rewritten as

$$\tilde{Y}(\vec{k}) = \tilde{D}(\vec{k})\tilde{T}(\vec{k}), \quad (8.2)$$

where  $\tilde{D}(\vec{k})$ ,  $\tilde{T}(\vec{k})$  are the Fourier transforms of the data and template respectively and  $\vec{k}$  is the Fourier dual for position. Then, to recover the matched filter output in position space, an inverse Fourier transform is performed on  $\tilde{Y}(\vec{k})$ . This is more computationally efficient and therefore viable for large data sets. For the remainder of this section, we describe the use of this simple form of the matched filter for detecting HI galaxies from simulated dirty maps. With more careful consideration of the inputs and the exact mode in which the operation takes place, the matched filter can be further optimized, but the current set up already produces results worth exploring. More in-depth discussions on the mathematics of matched

filters can be found in Turin (1960) [74], Bazdresch (2018) [76], and Saintonge (2007) [77]. Zubeldia et al. (2021) describe of its use as a maximum likelihood estimator [78].



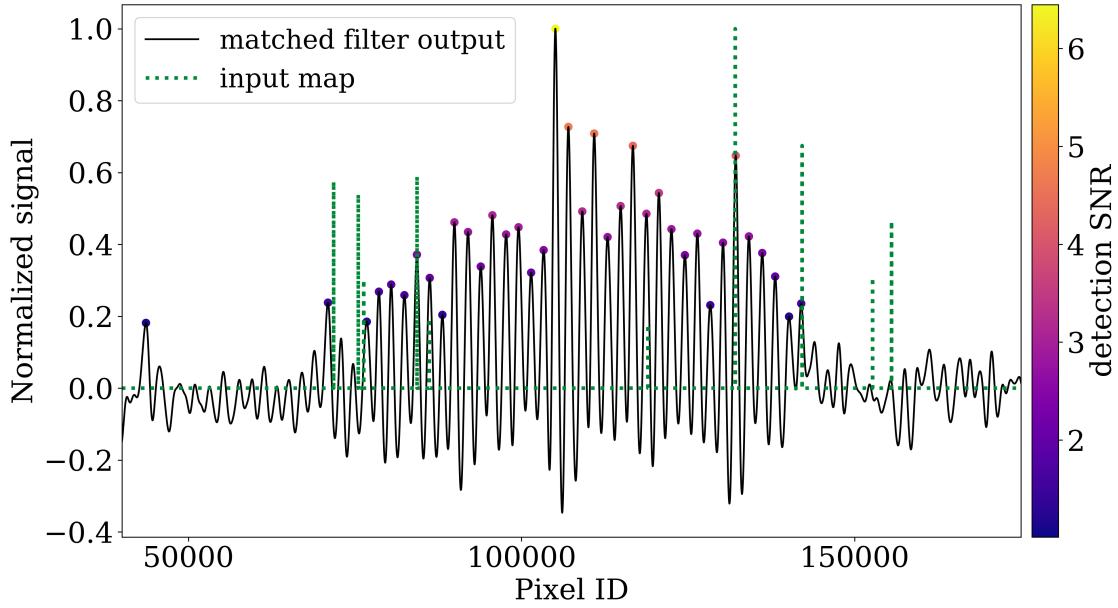
**Figure 8.1:** One-dimensional representations of the dirty beam template  $t(x)$  (left) and data  $d(x)$  (right) used for the matched filter. The dirty beam was simulated for a pathfinder configuration. The dirty map contains 10 injected galaxies shown in green.

Up to this point, we have been visualizing dirty maps in two angular dimensions (RA, Dec). Nevertheless, the dirty maps simulated by the pipeline are healpix maps, meaning they exist as a one-dimensional vector, where each pixel corresponds to a particular point in the two-dimensional sky. This representation of the dirty map allows for the use of a one-dimensional matched filter that contains the same amount of information than a two-dimensional matched filter would with the added benefits of computational simplicity. To illustrate this, the left panel of Figure 8.1 shows the one-dimensional representation of a dirty beam, or synthesized beam. The symmetry in the dirty beam is apparent, with side lobes emerging at equal distances to either side of the peak sensitivity. A two-dimensional projection of the left panel would resemble Figure 7.7. On the right panel, we show the one-dimensional representation of a dirty map containing 10 injected HI galaxies (shown in green) with a noise level corresponding to 200 days of observation. The presence of the complex spatial response of the dirty beam is clear given the much more intricate signal

shown in black compared to the simple localized emission of the true galaxies.

The dirty beam is used as the template  $t(x)$  and the dirty map is the data  $d(x)$  from which the signal is to be extracted. The matched filter output  $y(x)$  is shown in black in Figure 8.2. Given that  $y(x)$  contains a value for every pixel in the dirty map, some threshold needs to be established to quantify which pixels contain detections and which do not. A simple way to go about this is to estimate the variance of  $y(x)$  in a region believed to contain no signal. Then a minimum signal-to-noise ratio (SNR) can be set, under which peaks in  $y(x)$  are taken to be noise fluctuations and not detections. Using this set up, in Figure 8.2, we show the detections with  $\text{SNR} > 1$  colored corresponding to their SNR as given by the color bar. For comparison, the injected HI galaxies are still shown by the green dotted lines. It is immediately evident that there are both false-positive detections (red points in pixels that don't contain HI galaxies) and false-negatives (pixels with HI galaxies that are not flagged as detections). Interestingly, this classification seems to be affected not only by the brightness of the sources but also by their position in the sky. For instance, we see a relatively faint source near pixel  $\sim 120,000$  being correctly detected and a much brighter source closer to  $\sim 70,000$  being missed.

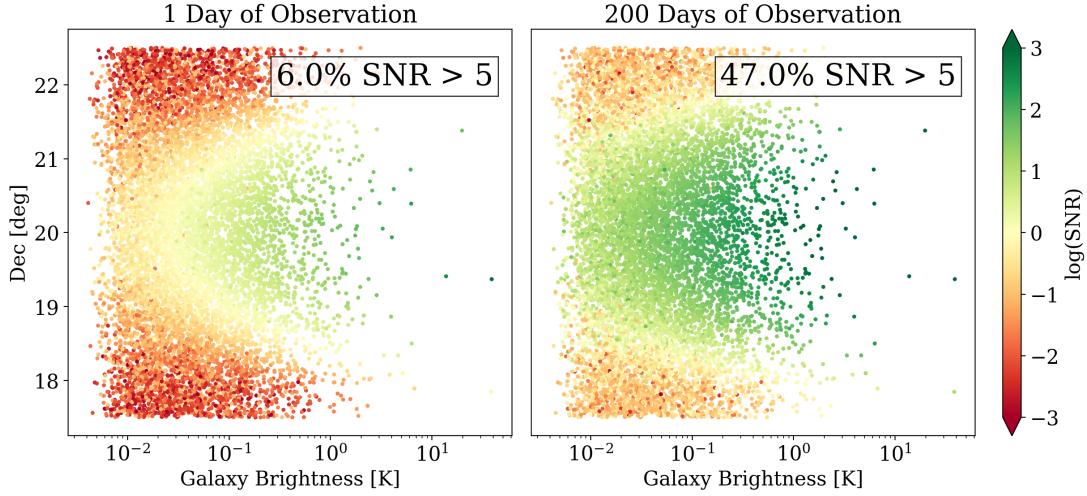
To further explore this effect in the recovery of HI galaxies, we ran a matched filter on a dirty map with a large sample of galaxies ( $N_{\text{gals}} = 10,000$ ). In pursuit of understanding the effect of the position of the sources on their detection, we track the SNR of all real detections and discard the false detections. Figure 8.3 shows the SNR of the detections made as a function of both the brightness of the injected galaxies and their Declination for two noise levels. For this observation, the CHORD pathfinder had an observing stripe centered at  $\text{Dec} = 20^\circ$ . The approximate size of the curve separating detections with  $\log(\text{SNR}) > 1$  from ones with  $\log(\text{SNR}) < 1$  is in agreement with CHORD's primary beam size  $\theta \sim \lambda/D \sim 0.21\text{m}/6\text{m} \sim 2^\circ$ . From this we conclude that both bright and faint sources near the edges



**Figure 8.2:** Matched filter output  $y(x)$  for the dirty map. The red points locate pixels where a detection with  $\text{SNR} > 1$  was made while the true pixels of the injected galaxies are shown in green. We see both false-positive and false-negative detections.

of this  $\sim 2^\circ$  stripe are likely to go undetected by the matched filter due to the suppression of their signal by the primary beam. The shape of the curve stays consistent as a function of the observing time, but there is a significant increase in detections above  $5\sigma$  from 6% to 47% as the observing time increases from one to 200 days.

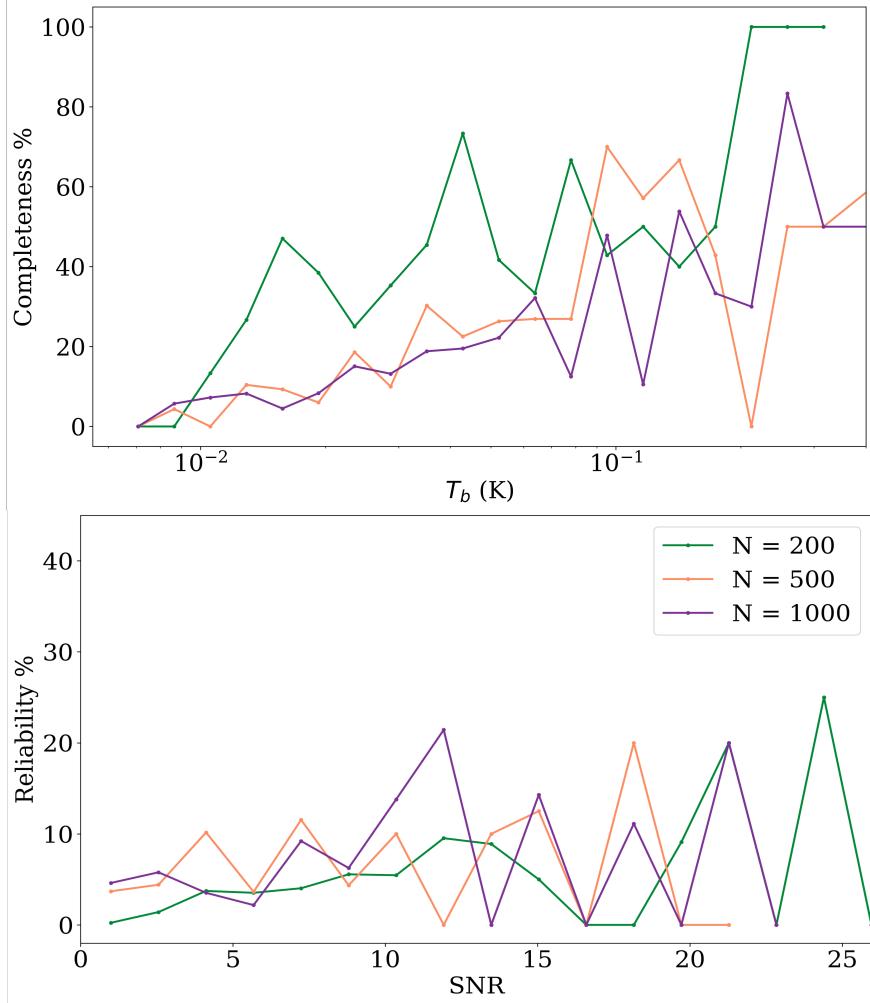
We have established that HI galaxy recovery has a dependence on both the galaxy brightness and its distance from the center of the observing stripe. Now, we quantify how the matched filter performs as a function of the number of galaxies in the map. In the limit of a map containing a single source, one could imagine that the template (if chosen correctly) would perfectly localize the source in the map. Then, as more sources are added and the combination of their spatial features becomes increasingly complicated, we expect the matched filter to have worse success. We simulate observations for maps with



**Figure 8.3:** Detection SNR as a function of the galaxy brightness and the Declination of the source. The left and right panels show results for noise levels equivalent to one and 200 day of observations respectively. The labels in the figure show the percentage of sources detected with a  $5\sigma$  or higher significance.

$N_{\text{gals}} = 200, 500, 1000$  and take  $1\sigma$  to be the detection limit for the remainder of this section. The set up and templates used for matched filtering these maps are the same as the ones described with regards to Figure 8.1.

In the top panel of Figure 8.4, we show the completeness (percentage of galaxies that were correctly detected) as a function of the galaxy brightness. Overall, the detected sample is more complete for brighter sources in all three sample sizes, reaching up to 100% for the  $N_{\text{gals}} = 200$  map. There is indication of an increase in completeness for less crowded maps, though the results are comparable for  $N_{\text{gals}} = 500$  and 1000. We also look at the reliability of the detections (percentage of detections that are real galaxies and not false-positives) in the bottom plot. The reliability is relatively low for all  $N_{\text{gals}}$ , with a small increase up to  $\sim 25\%$  for  $\text{SNR} \sim 25$ , signifying that only a quarter of the detections made with the matched filter are in fact true galaxies in the map.



**Figure 8.4:** Summary statistics for the detections made with the matched filter in dirty maps with  $N_{\text{gals}} = 200, 500, 1000$ . The completeness (top panel) is higher for brighter sources and for less crowded maps. The reliability (bottom panel) remains under  $\sim 25\%$  for all dirty maps and values of SNR.

There are a few important caveats for the way matched filters were used in this thesis. Firstly, the template used was a synthesized beam simulated for a single pointing of the CHORD pathfinder. Given that the synthesized beam is a position dependent quantity, it is guaranteed that there will be discrepancies between the template and the data at all positions

other than that original pointing. An improvement would be to use multiple templates and estimate a combined SNR output from those or to use an average synthesized beam that may better represent the instrument's response at more sky positions. Secondly, a more robust characterisation of the detection threshold is needed for the matched filter to yield results with higher reliability. Finally, the approach taken in this thesis requires one to run the matched filter independently for each frequency slice of the dirty map and then get an integrated SNR per pixel by summing the outputs over the frequency axis. From this integrated statistic, one must then employ another algorithm recover individual HI galaxy profiles from pixels with high SNR. In Section 8.3 we explore the use of a frequency matched filter for this purpose. An optimal matched filter, which is currently in development, will account for both spectral and spatial features at once.

## 8.2 CLEAN

The profiles extracted from pixels with high SNR as described in Section 8.1 will contain excess signal due to spatial aliasing, as illustrated in Figure 7.13, and may be difficult to interpret. This is due to the fact that though the matched filter accounts for the complex spatial structures to give high-significance predictions of signal locations, there is no forward modelling involved and the features are not removed. A reasonable alternative to the spatial matched filter is CLEAN. As described in Chapter 6.4, this nonlinear algorithm forward-models the instruments response to detect and remove sources one-by-one from the map until reaching a desired threshold. This algorithm can help recover profiles more closely resembling pristine HI galaxy profiles without the features resulting from spatial aliasing. Two major caveats for using CLEAN were described in Section 6.4. The first is that the reconstruction of the noise properties in the resulting image is very challenging, as described in Section

**6.4.** The second is that CLEAN can be unreliable for images containing extended sources or for maps created with very complex features in the synthesized beam. In this context, both the matched filter and CLEAN have their benefits and limitations. The former is a computationally efficient and statistically robust method for source detection while the latter provides results that are easier to interpret but lack appropriate noise estimates.

Throughout this section, key properties of CLEAN will be further explored, including the algorithm's sensitivity to the beam model, the position of sources with respect to one another, and the dynamic range of galaxy brightness in the map. For this purpose, an input map was created containing five sources within two degrees of each other with carefully selected fluxes. The sources injected were only as bright as the brightest source's side lobes once observed by the instrument. This was chosen to illustrate that any misrepresentations of the instrument's response to the brightest source may directly result in difficulties when detecting the remaining sources. A dirty map was simulated for this test map with a noise level corresponding to 300 days of observation.

This specific configuration allowed us to explore the performance of CLEAN for three different simulated beams. The first is a self-consistent pathfinder synthesized beam that is simulated by injecting a single point source and doing a mock observation with the pipeline (we call this simulated beam for the remainder of this section). The second beam used is a Gaussian beam with the form

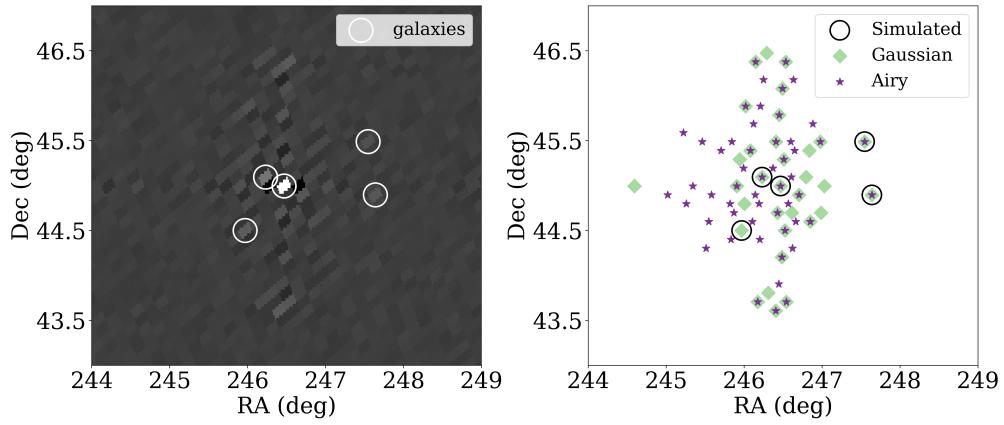
$$G(r) = e^{-\frac{r^2}{2\sigma^2}}, \quad \sigma = \frac{\text{FWHM}}{2\sqrt{2\log(2)}},$$

where  $\text{FWHM} = \lambda_{\text{obs}}/b_{\text{max}}$  is the full-width-half-maximum of the Gaussian estimated at the observed wavelength  $\lambda_{\text{obs}}$  and for the maximum baselines length  $b_{\text{max}}$ . For this particular dirty map, the values used were  $b_{\text{max}} = 86\text{m}$  and  $\lambda_{\text{obs}} = 0.2114\text{m}$ . The two-dimensional variable  $r$  defines the extent along RA and Dec in degrees up to which the Gaussian is

computed. The third beam used for this example is an Airy beam with the form

$$A(r) = \frac{2J_1(x)}{x}, \quad x = \frac{\pi b_{\max} \sin(r)}{\lambda_{\text{obs}}},$$

where  $J_1(r)$  is the Bessel function of first order and  $r$  is the same as in the Gaussian beam. It is important to note that while Gaussian and Airy beams are often used to model primary beams in radio interferometers, we use them here as synthesized beams by inputting the largest baseline length instead of the dish diameter as the characteristic size.



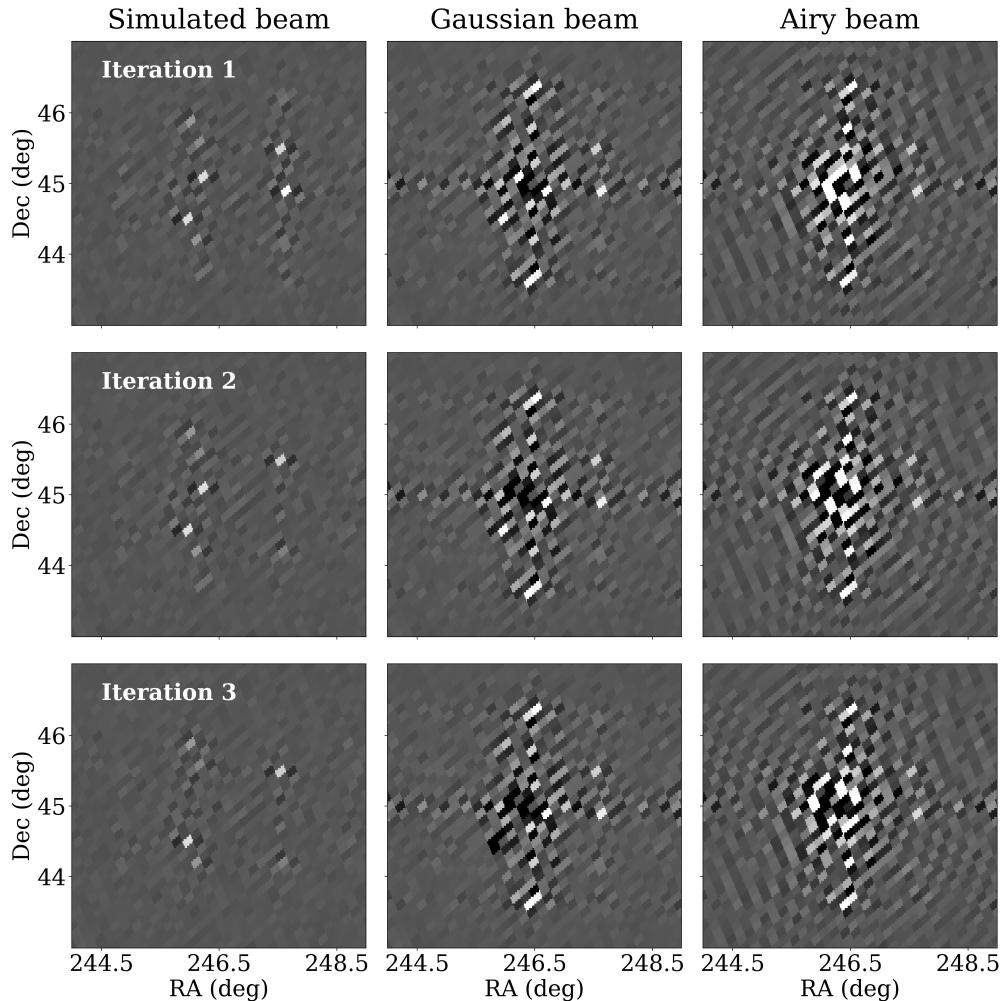
**Figure 8.5:** Performance of CLEAN using three synthesized beams. The left panel shows the dirty map with the galaxies circled in white. The right panel shows the detections made with the simulated synthesized beam (black), the Gaussian beam (green), and the Airy beam (purple).

Figure 8.5 illustrates the performance of each of these beams in detecting the sources. The left panel shows the dirty map with the input galaxies circled in white to differentiate them from the other spatial features. The right panel shows the detections made with each of the three beams. Firstly, CLEAN had a perfect performance when using the simulated synthesized beam (shown as black circles). This is expected given this synthesized beam is self-consistent with the observation. All five sources were correctly localized and no residual

features remained above the noise level leading to no false detections. The detections made using the Gaussian beam (shown in green) and the Airy beam (shown in purple) were less successful. Four out of the five sources were correctly detected in all cases, while the fifth was detected only when using the Gaussian beam and the simulated synthesized beam. Notably, the use of both the Gaussian and Airy beams resulted in a large number of false positives in the regions surrounding the bright sources. With the Gaussian beam, a total of 73 iterations were completed before the noise threshold of  $1\sigma$  was reached, whereas with the Airy beam a break point had to be set at 102 iterations, after which the incorrectly modelled instrumental response caused an unbounded increase in the residuals.

To better understand why this was the case and also gain some intuition into some of the impacts of mischaracterizing the instrument beam on CLEAN’s performance, we look more closely at three consequent iterations of the algorithm in Figure 8.6. Each column shows the iterations for CLEAN using the named synthesized beam. Each iteration starts with the localization of the brightest source, the modeling of the instrument’s response to the source, and the subtraction of that model from the dirty map. The residuals of this subtraction are show in each row. Iteration 1 starts identically for the three cases, as the brightest source is localized correctly in the dirty map, but the residuals already show significant differences. While CLEAN seems to have subtracted the brightest source reasonably well with the simulated synthesized beam, the same cannot be said for the other two beams. In particular, the use of a Gaussian synthesized beam resulted in over-subtracting for pixels surrounding the source and under-subtracting for the side lobes present  $\sim 1^\circ$  to either direction of the source. This makes sense given that a Gaussian beam monotonically decreases radially away from the peak and contains no side lobes. For the Airy beam, we see similar behaviors around the neighborhood of the source but we also see very clear imprinted features extending radially outwards. These features alternate in over- and under-

subtracting, which leads to regions with larger signals in the residuals compared to the dirty map.



**Figure 8.6:** Three consequent iterations of CLEAN using the simulated synthesized beam, a Gaussian beam, and an Airy beam. Each row show the residuals after the instrument's response to the remaining brightest source is subtracted from the dirty map.

For the following two iterations, CLEAN continues to perform well with the simulated synthesized beam, accurately subtracting off the instrument's response to two additional sources. In the case of the Gaussian beam, the next two sources are also localized correctly

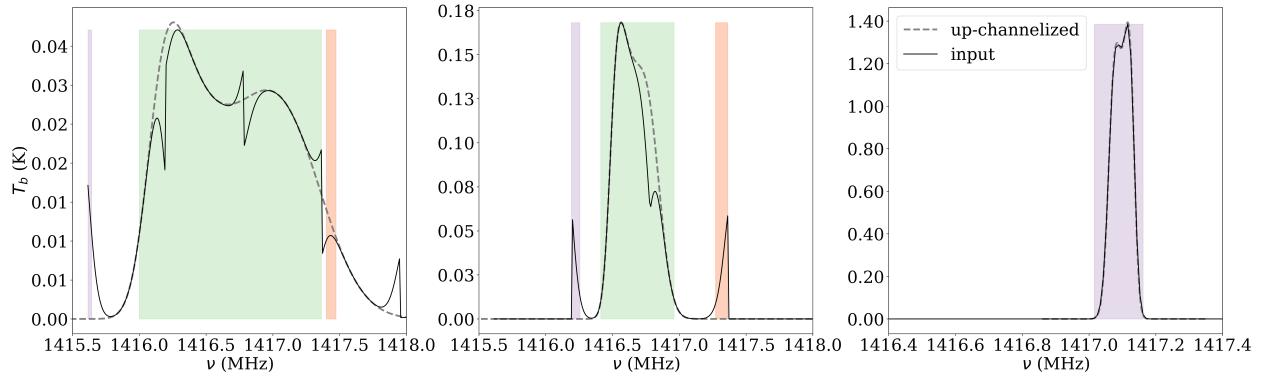
but their subtraction continues to suffer from the limitations of the previous iteration. It is clear from this process that the mischaracterisation of the synthesized beam’s side lobes can lead to multiple false detections due to improper subtraction of the instrument’s response away from the center of the source. Lastly, subsequent iterations of CLEAN with an Airy beam provide a clear illustration of the common pitfalls of this algorithm. The oscillating nature of the Airy beam (from positive to negative values) and its inadequacy in modeling the instruments response result in a positive feedback loop that amplifies the imprinted radial patterns and buries the true sources. The lesson learned from this exploration is first and foremost the relevance of a good understanding of the instrumental response for detecting sources from data containing complex spatial structure.

Assuming one manages to circumvent the issues typically present with CLEAN, the algorithm will output a pristine data cube from which galaxy profiles can be extracted and further examined. In the following section, we use a frequency matched filter to extract individual HI galaxies profiles from data presumed to have been perfectly CLEANed. The investigation done in this section makes it evident that this view is excessively optimistic as in reality the data cubes will contain some combination of remaining instrumental features and those imprinted onto the data during the process of CLEAN itself. However, this is still a helpful exercise as it provides insights into the complexities that spectral aliasing can add to the detection process.

### 8.3 Frequency matched filter

Following the mathematical set up provided in Section 8.1, we use a frequency matched filter to detect individual galaxy profiles that have been up-channelized by different factors  $U$ . We define a template to be of a Gaussian form with its width being a free parameter to be set

depending on the features one wants to pick out from the profiles. This is the first iteration of a more sophisticated matched filter that will fold in information about the up-channelization scheme into a more realistic set of templates. For now, the template used has no information about the spectral features imprinted by up-channelization, the dips and aliases, and also has no variation in shape to represent asymmetries or double-horns. Nonetheless, this first attempt is an opportunity to start building a complete understanding into how a matched filter performs on data with these specific spectral systematics.



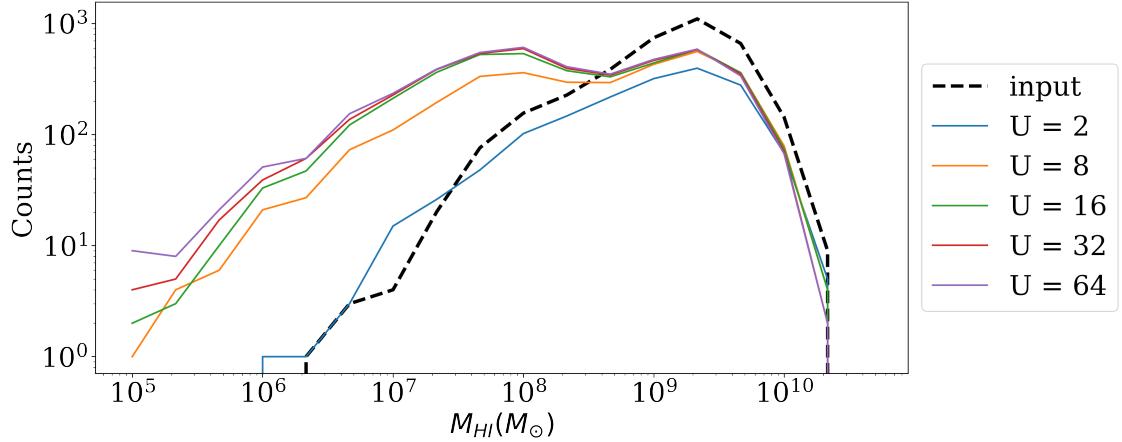
**Figure 8.7:** Examples of the matched filter outputs for three profiles that have been up-channelized with  $U = 64$ . The colored bands represent frequency ranges flagged by the algorithm as a detection. The original galaxy profiles are shown in the gray dashed lines.

Figure 8.7 shows three example outputs from the matched filter applied to galaxies up-channelized with  $U = 64$  (black solid lines). The colored bands represent frequency ranges flagged by the algorithm as a detection and the original HI profiles are shown behind (dashed gray lines). The left panel shows a particularly wide HI galaxy profile that occupies four coarse channels. The features imprinted by up-channelization are present at multiple frequencies, with multiple dips and two sharp aliases to either side. The matched filter detects three sources in this case – one in the lower frequency alias, one in the main body of the profile and a third one still in the main body but separated by a feature

imprinted by the coarse channel modulation. In this particular case, there is a false detection and a duplicate detection of the same galaxy, both of which would lead to an underestimate in the profile width. This is a common scenario for wide profiles. In the center panel there is a moderately wide galaxy profile spanning two coarse channels. Each of the two aliases are picked up and the galaxy profile seems to be detected in its full range. The right panel contains a thin and tall profile that lives entirely within a single coarse channel. In this case, the high up-channelization factor helps resolve the double-horn feature but does not add any aliasing and the galaxy is recovered perfectly.

We run the algorithm on the entire mock HI galaxy catalog for up-channelization factors of  $U = 2, 8, 16, 32, 64$  to examine their impacts on the number of detections made and any possible biases in the recovered galaxies. To quantify this, the HI mass of each detected galaxy is estimated from the profiles within each detection range using Equation 7.3. A distribution of the HI masses recovered for each  $U$  in comparison to the input is shown in Figure 8.8. We see that for all  $U$ , the number of detections is always smaller than the input; some sources are not detected. For  $U = 2$ , the number of galaxies at intermediate  $M_{\text{HI}} \sim 10^6 - 10^{7.5}$  is only slightly overestimated. For all other  $U > 2$ , the low-mass end is significantly overestimated when compared to the distribution of input galaxies. There is a clear increase in detections across the entirety of the mass spectrum as a function of  $U$ , i.e. more galaxies are detected throughout for higher up-channelization factors.

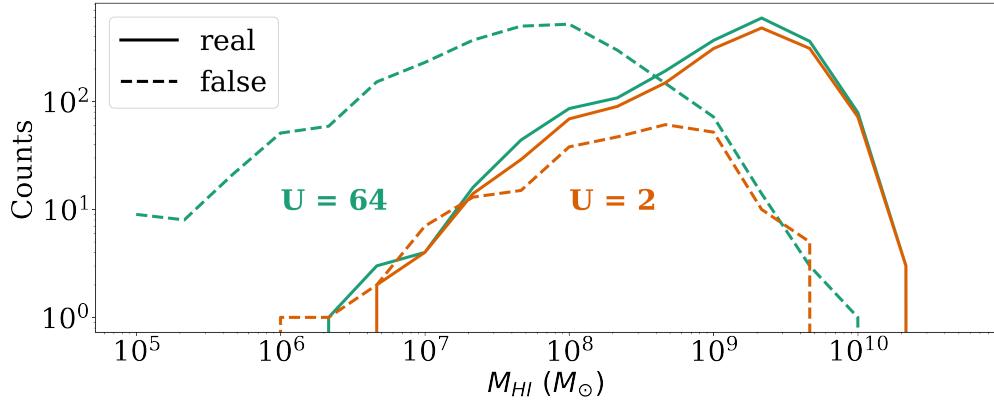
The shape of these recovered distributions, as well as the overestimation of low-mass galaxies needs to be well understood. In this pursuit, the detections were classified as real galaxies (the detection range falls within the range of an input galaxy) or false detections (it does not correspond to an input galaxy). The distribution of HI masses in these two categories is shown in Figure 8.9 for both  $U = 2$  (orange) and  $U = 64$  (green). The false detections (dashed lines) generally have lower HI masses than the real detections (solid



**Figure 8.8:** HI mass distribution of detected sources for  $U = 2, 8, 16, 32, 64$  shown in comparison to the input HI mass distribution (black). Interesting features are described in the text.

lines), though some have comparable masses. The  $M_{\text{HI}}$  distribution of real detections follows very similar shapes for both up-channelization factors, though  $U = 64$  does have higher counts everywhere. For false positive detections, on the other hand, the  $M_{\text{HI}}$  distribution is drastically different as a function of  $U$ . For  $U = 2$ , all detections are concentrated near the mid to high mass regime whereas for  $U = 64$ , there is a large number of false detections at low masses. The mean mass of real and false detections for this high up-channelization factor is strikingly different. This makes sense given the examples from Figure 8.7, where small and sharp aliases were incorrectly detected by the matched filter as galaxies. Low up-channelization factors are in a way immune to this as they don't have enough resolution to show these features, but for high up-channelization factors, these thin aliases make up the majority of the low-mass detections. The shapes of the overall counts shown in Figure 8.8, and in particular the dips in the distributions near  $M_{\text{HI}} \sim 10^{8.5} M_{\odot}$ , are not intrinsic features, but are derived from the interplay of the real and false detection populations.

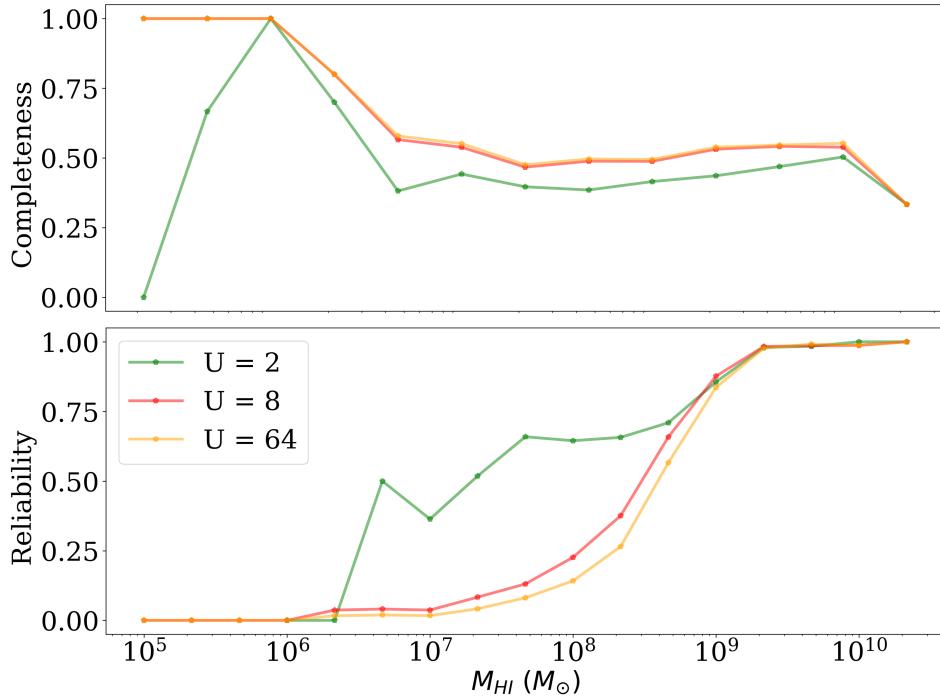
We look at some summary statistics for the detection of galaxies as a function of their



**Figure 8.9:** Distribution of HI masses for detections classified as real (solid lines) and false (dashed lines) for  $U = 2$  (orange) and  $U = 64$  (green). False detections generally have lower HI masses, though that effect is more present for higher  $U$  factors.

HI mass and the up-channelization factor  $U$ . Figure 8.10 illustrates the completeness and reliability of the detections as a function of the described parameters for  $U = 2, 8, 64$ . The least complete sample comes from  $U = 2$  throughout the entire HI mass range. There is a significant increase in completeness in going from  $U = 2$  to higher up-channelization factors but not a considerable difference between the samples detected with these higher values of  $U$ . The roles are inverted for reliability, with lower  $U$  values leading to more reliable detections across the HI mass range, in particular at intermediate  $M_{\text{HI}}$ . False detections make up a considerable fraction of the total detections for catalogs up-channelized with  $U > 2$ , with the reliability getting slightly worse as  $U$  increases. This illustrates the need to determine an optimal up-channelization factor  $U$  for any given science case. If one is after reliable detections, lower  $U$  values are preferred. However, if the science case requires detecting low-mass galaxies, as many that were described in Chapter 2.2 do, a higher  $U$  is preferred, even if it results in a larger fraction of false positives.

These results give a somewhat pessimistic view of the process of galaxy detection with



**Figure 8.10:** Summary statistics for the detections of galaxies as a function of HI mass and  $U$ . The top panel shows the completeness percentage, the bottom panel shows the reliability of detections. Samples corresponding to lower  $U$  seem to be less complete but more reliable for all  $M_{HI}$ .

the CHORD pathfinder. We do in fact predict a large number of false detections in the low-HI mass range for high  $U$  given the systematics present. However, the approach taken for this thesis does not account for a handful of methods and improvements that we can confidently expect will be put in place for the analysis of real CHORD data. We bring some of those to light now.

The simplest way to limit false positive detections, given what we learn from Figure 8.9, is to create a low-mass cut at some  $M_{HI}^{\min}$  so that all sources detected with  $M_{HI} < M_{HI}^{\min}$  are automatically disregarded. It is possible to estimate the limiting HI mass for a survey with

the following approximation [31],

$$M_{\text{HI}} = \frac{16\pi}{3} \frac{m_{\text{H}}}{A_{12}hc} \frac{D_L^2(z)}{1+z} f^{-1} \frac{V(z)}{\sqrt{V(z)/\Delta V}} S_N \sigma_{4h} \sqrt{\frac{4}{t}}. \quad (8.3)$$

Here,  $m_{\text{H}}$  is the hydrogen mass,  $A_{12} = 2.6 \times 10^{-15} \text{ s}^{-1}$  is the Einstein coefficient describing the probability per unit time of the HI hyperfine transition [11] and  $h$  and  $c$  are Planck's constants and the speed of light, respectively. Other variables that factor into this estimate are the luminosity distance to the source  $D_L(z)$  and the HI velocity at redshift  $z$ ,  $V(z)$ . A few instrumental parameters follow:  $f$  is the fraction of the sensitivity of one's telescope relative to the SKA,  $\Delta V$  is the channel width in units of km/s,  $\sigma_{4h}$  is the rms noise for a 4 hour observation of a single FoV and  $t$  is the integration time for that FoV in hours. Finally,  $S_N$  is the signal to noise ratio chosen to signify a statistically robust detection.

This low-mass cut approach will completely eliminate false detections (and inadvertently some real detections) but it will leave the false positives at higher HI masses intact. To tackle this, this simple cut could be paired with a more refined attempt at understanding the detection sample distributions. Given the deterministic nature of up-channelization, one could estimate the ratio between real and false detections, as was done for this thesis, and make informed decisions on the reliability of the sources. Even though this will not provide a definitive answer as to which galaxies are real and which are not, it can provide some statistics about our confidence in them. Going a step further, once potential detections are made with CHORD or other interferometers, follow-up observations can be done with other instruments that have simpler systematics, such as large single-dish telescopes to confirm them. More detailed studies of individual sources of interest is possible in this way after the bulk detection is done by the interferometer.

Finally, a more sophisticated matched filter algorithm that accounts for up-channelization features is in progress. The more realistic templates are expected to aid with detections in two

main ways. The first one is by decreasing the number of false detections. The templates will follow similar shapes to up-channelized HI galaxy profiles, with dips and evenly spaced sharp aliases to either side. Due to this, individual thin fine peaks will likely not be detected by the algorithm. The second improvement will come from our ability to detect sources with lower SNR than we would with a simple Gaussian matched filter. Faint galaxies could be buried beneath thermal Gaussian noise. However, their profiles will vary significantly from Gaussian forms as illustrated throughout this section and through Section 5.2. Appropriately shaped templates that are tuned to match these very specific and deterministic features in their profiles will be able to dig deeper into the noise to detect these sources. The quantification of the results of this more sophisticated matched filter is yet to be done but it is certain to be an improvement over the simplified model used for this thesis.

# Chapter 9

## Conclusion

In this thesis, we examined the recovery of HI galaxies with mock CHORD pathfinder observations. In Chapter 2, we introduced the science case for HI galaxies, discussing a variety of astrophysical and cosmological applications. We reviewed the current state of HI galaxy surveys in Chapter 3, highlighting contributions from HIPASS and ALFALFA that paved the way for up and coming blind HI surveys. The observational parameters and science goals of FAST and future SKA pathfinder instruments were also presented to lay the ground for a discussion of CHORD and the CHORD pathfinder in Chapter 4. In Chapters 5 and 6 we gave a conceptual and mathematical description of radio interferometry and map-making. We also provided details on the spectral properties of future CHORD observations and the up-channelization algorithm.

With all of the tools set up, we presented the end-to-end pipeline in Chapter 7. We create input sky maps that contain galaxies from an ALFALFA-constrained mock HI catalog. The spectra of the injected galaxies are Busy functions [1] with parameters sampled from predetermined distributions in agreement with the ALFALFA HI mass function (HIMF) and HI velocity width function (HIWF). In addition to the HI galaxies, we include other semi-

realistic sky components. Using the existing CHORD pipeline, we set up an instrument with a  $6 \times 11$  dish configuration, emulating the CHORD pathfinder array, and simulate the instrument’s response to the sky. We include additional systematics due to the up-channelization process in the mock observations. The resulting dirty maps are analyzed in the hopes of identifying key characteristics in the data. We highlight the effects of spatial aliasing due to the pathfinder’s complex synthesized beam, showing that it leads to excess emission in pixels near bright sources. In addition to the spatial effects, the HI galaxy profiles contain spectral features imprinted by the up-channelization scheme that add to the very particular set of systematics of these mock observations.

While the spatial distribution of HI sources is in agreement with the full ALFALFA catalog constraints, this work treats all galaxies as point sources. We expect that the angular size of nearby galaxies will extend across multiple pixels and thus their morphology may be resolved by the CHORD beam. A good extension to this pipeline would be to include this additional layer of realism. In addition, we currently use an up-channelization scheme that has a constant  $U$  factor for the entire CHORD bandwidth. As described in Section 5.2, this is not necessarily the optimal scheme as fine spectral resolution may not be needed at frequencies where we expect severe RFI contamination. In the future, a more flexible up-channelization scheme could be implemented in this pipeline to help inform the decisions for real CHORD and CHORD pathfinder observations. In this work, we focus on the systematics for the CHORD pathfinder array given that its first light is imminent, but similar explorations can be done for full CHORD, including the antennas in the outrigger sites. This could lead to some very interesting investigations into the possibility of detailed studies of nearby galaxies and their morphology given the finer angular resolution of full CHORD.

In Chapter 8 we explore the use of matched filters for the detection of HI galaxies from the simulated dirty maps. We use separate matched filters for the spatial search of sources in each

frequency slice and for the determination of galaxy profiles along the frequency direction. This simplified approach precedes a more sophisticated three dimensional matched filter and helps us start building an intuition for the effects of the spatial and spectral systematics on our ability to detect and categorize HI galaxies. Currently, the spatial matched filter uses a simulated synthesized beam as a template and results in a relatively complete ( $> 50\%$ ) sample of detections for maps that are not over-crowded with sources. It also results in a large number of false detections across all scenarios explored. A more realistic template and robust characterisation of the noise properties is needed for better results. For the frequency matched filter, we find that its performance is strongly correlated with the up-channelization factor  $U$ . More galaxies are detected for finer spectral resolutions (higher  $U$ ), especially at lower  $M_{\text{HI}}$ . These detections include both real detections that were made possible by resolving the spectral features of the galaxies, and false detections coming from the sharp aliases at the edges of coarse channels. A matched filter that contains information on the deterministic features of up-channelization will greatly improve the outcome of this search. Finally, we illustrated the importance of a good characterisation of the instrument's beam for accurate detections through an exploration using CLEAN.

Galaxy surveys are essential probes for mapping out the distribution of matter in the nearby universe. They provide crucial information about the processes of galaxy formation and evolution, as well as tighter constraints on cosmological model parameters. Technological and scientific advances for single-dish telescopes and radio interferometers are paving the way for a new generation of HI galaxy surveys with better sensitivities, larger sky coverage, and finer resolutions. In this rapidly evolving landscape, a new set of challenges presents itself in the form of complex instrumental systematics. The exploration done in this work provides insights into the prospects and complexities of interferometric HI galaxy surveys, all of which must be well understood in the hopes of answering fundamental questions about the universe.

# Bibliography

- [1] T. Westmeier, R. Jurek, D. Obreschkow, B. S. Koribalski, and L. Staveley-Smith, “The busy function: a new analytic function for describing the integrated 21-cm spectral profile of galaxies,” *Monthly Notices of the Royal Astronomical Society*, vol. 438, pp. 1176–1190, 2014.
- [2] C.-P. Zhang, M. Zhu, P. Jiang, C. Cheng, J. Wang, J. Wang, J.-L. Xu, X.-L. Liu, N.-P. Yu, L. Qian, H. Yu, M. Ai, Y. Jing, C. Xu, Z. Liu, X. Guan, C. Sun, Q. Yang, M. Huang, and Q. Hao, “The FAST all sky HI survey (FASHI): The first release of catalog,” *Science China Physics, Mechanics and Astronomy*, vol. 67, Dec. 2023.
- [3] K. Vanderlinde, A. Liu, B. Gaensler, D. Bond, G. Hinshaw, C. Ng, C. Chiang, I. Stairs, J.-A. Brown, J. Sievers, J. Mena, K. Smith, K. Bandura, K. Masui, K. Spekkens, L. Belostotski, M. Dobbs, N. Turok, P. Boyle, M. Rupen, T. Landecker, U.-L. Pen, and V. Kaspi, “The Canadian Hydrogen Observatory and Radio-transient Detector (CHORD),” in *Canadian Long Range Plan for Astronomy and Astrophysics White Papers*, vol. 2020, p. 28, Oct. 2019.
- [4] D. Li, J. M. Dickey, and S. Liu, “Preface: Planning the scientific applications of the Five-hundred-meter Aperture Spherical radio Telescope,” *Research in Astronomy and Astrophysics*, vol. 19, p. 016, Feb. 2019.

- [5] E. White, F. D. Ghigo, R. M. Prestage, D. T. Frayer, R. J. Maddalena, P. T. Wallace, J. J. Brandt, D. Egan, J. D. Nelson, and J. Ray, “Green Bank Telescope: Overview and analysis of metrology systems and pointing performance,” *Astronomy and Astrophysics*, vol. 659, p. A113, Mar. 2022.
- [6] J. A. Sobrin, A. J. Anderson, A. N. Bender, B. A. Benson, D. Dutcher, A. Foster, N. Goeckner-Wald, J. Montgomery, A. Nadolski, A. Rahlin, P. A. R. Ade, Z. Ahmed, E. Anderes, M. Archipley, J. E. Austermann, J. S. Avva, K. Aylor, L. Balkenhol, P. S. Barry, R. B. Thakur, K. Benabed, F. Bianchini, L. E. Bleem, F. R. Bouchet, L. Bryant, K. Byrum, J. E. Carlstrom, F. W. Carter, T. W. Cecil, C. L. Chang, P. Chaubal, G. Chen, H.-M. Cho, T.-L. Chou, J.-F. Cliche, T. M. Crawford, A. Cukierman, C. Daley, T. d. Haan, E. V. Denison, K. Dibert, J. Ding, M. A. Dobbs, W. Everett, C. Feng, K. R. Ferguson, J. Fu, S. Galli, A. E. Gambrel, R. W. Gardner, R. Gualtieri, S. Guns, N. Gupta, R. Guyser, N. W. Halverson, A. H. Harke-Hosemann, N. L. Harrington, J. W. Henning, G. C. Hilton, E. Hivon, G. P. Holder, W. L. Holzapfel, J. C. Hood, D. Howe, N. Huang, K. D. Irwin, O. B. Jeong, M. Jonas, A. Jones, T. S. Khaire, L. Knox, A. M. Kofman, M. Korman, D. L. Kubik, S. Kuhlmann, C.-L. Kuo, A. T. Lee, E. M. Leitch, A. E. Lowitz, C. Lu, S. S. Meyer, D. Michalik, M. Millea, T. Natoli, H. Nguyen, G. I. Noble, V. Novosad, Y. Omori, S. Padin, Z. Pan, P. Paschos, J. Pearson, C. M. Posada, K. Prabhu, W. Quan, C. L. Reichardt, D. Riebel, B. Riedel, M. Rouble, J. E. Ruhl, B. Saliwanchik, J. T. Sayre, E. Schiappucci, E. Shirokoff, G. Smecher, A. A. Stark, J. Stephen, K. T. Story, A. Suzuki, C. Tandoi, K. L. Thompson, B. Thorne, C. Tucker, C. Umilta, L. R. Vale, K. Vanderlinde, J. D. Vieira, G. Wang, N. Whitehorn, W. L. K. Wu, V. Yefremenko, K. W. Yoon, and M. R. Young, “The design and integrated performance of SPT-3G,” *The Astrophysical Journal Supplement Series*, vol. 258, p. 42,

Feb. 2022.

- [7] CHIME/FRB Collaboration, M. Amiri, K. Bandura, P. Berger, M. Bhardwaj, M. M. Boyce, P. J. Boyle, C. Brar, M. Burhanpurkar, P. Chawla, J. Chowdhury, J. F. Cliche, M. D. Cranmer, D. Cubranic, M. Deng, N. Denman, M. Dobbs, M. Fandino, E. Fonseca, B. M. Gaensler, U. Giri, A. J. Gilbert, D. C. Good, S. Guliani, M. Halpern, G. Hinshaw, C. Höfer, A. Josephy, V. M. Kaspi, T. L. Landecker, D. Lang, H. Liao, K. W. Masui, J. Mena-Parra, A. Naidu, L. B. Newburgh, C. Ng, C. Patel, U. L. Pen, T. Pinsonneault-Marotte, Z. Pleunis, M. Rafiei Ravandi, S. M. Ransom, A. Renard, P. Scholz, K. Sigurdson, S. R. Siegel, K. M. Smith, I. H. Stairs, S. P. Tendulkar, K. Vanderlinde, and D. V. Wiebe, “The CHIME Fast Radio Burst Project: System Overview,” *The Astrophysical Journal*, vol. 863, p. 48, Aug. 2018.
- [8] D. R. DeBoer, A. R. Parsons, J. E. Aguirre, P. Alexander, Z. S. Ali, A. P. Beardsley, G. Bernardi, J. D. Bowman, R. F. Bradley, C. L. Carilli, C. Cheng, E. d. L. Acedo, J. S. Dillon, A. Ewall-Wice, G. Fadana, N. Fagnoni, R. Fritz, S. R. Furlanetto, B. Glendenning, B. Greig, J. Grobbelaar, B. J. Hazelton, J. N. Hewitt, J. Hickish, D. C. Jacobs, A. Julius, M. Kariseb, S. A. Kohn, T. Lekalake, A. Liu, A. Loots, D. MacMahon, L. Malan, C. Malgas, M. Maree, Z. Martinot, N. Mathison, E. Matsetela, A. Mesinger, M. F. Morales, A. R. Neben, N. Patra, S. Pieterse, J. C. Pober, N. Razavi-Ghods, J. Ringuette, J. Robnett, K. Rosie, R. Sell, C. Smith, A. Syce, M. Tegmark, N. Thyagarajan, P. K. G. Williams, and H. Zheng, “Hydrogen Epoch of Reionization Array (HERA),” *Publications of the Astronomical Society of the Pacific*, vol. 129, p. 045001, Mar. 2017.
- [9] C. J. Lonsdale, R. J. Cappallo, M. F. Morales, F. H. Briggs, L. Benkevitch, J. D. Bowman, J. D. Bunton, S. Burns, B. E. Corey, L. Desouza, S. S. Doeleman, M. Derome,

- A. Deshpande, M. R. Gopala, L. J. Greenhill, D. E. Herne, J. N. Hewitt, P. A. Kamini, J. C. Kasper, B. B. Kincaid, J. Kocz, E. Kowald, E. Kratzenberg, D. Kumar, M. J. Lynch, S. Madhavi, M. Matejek, D. A. Mitchell, E. Morgan, D. Oberoi, S. Ord, J. Pathikulangara, T. Prabu, A. Rogers, A. Roshi, J. E. Salah, R. J. Sault, N. U. Shankar, K. S. Srivani, J. Stevens, S. Tingay, A. Vaccarella, M. Waterson, R. B. Wayth, R. L. Webster, A. R. Whitney, A. Williams, and C. Williams, “The Murchison Widefield Array: Design Overview,” *IEEE Proceedings*, vol. 97, pp. 1497–1506, Aug. 2009.
- [10] R. Dutta, S. Kurapati, J. N. H. S. Aditya, O. Bait, M. Das, P. Dutta, K. Indulekha, M. Nandakumar, N. N. Patra, N. Roy, and S. Roychowdhury, “Probing galaxy evolution through HI 21-cm emission and absorption: current status and prospects with square kilometre array,” *Journal of Astrophysics and Astronomy*, vol. 43, Dec. 2022.
- [11] M. Moazzenzadeh and J. T. Firouzjaee, “Searching for the hydrogen 21 cm line in cosmos,” 2021.
- [12] R. Giovanelli and M. P. Haynes, “Extragalactic neutral hydrogen,” in *Galactic and Extragalactic Radio Astronomy* (G. Verschuur and K. Kellermann, eds.), ch. 12, pp. 522–562, Astronomy and Astrophysics Library. Springer, New York, NY, 1988.
- [13] A. A. West, D. A. Garcia-Appadoo, J. J. Dalcanton, M. J. Disney, C. M. Rockosi, Željko Ivezić, M. C. Bentz, and J. Brinkmann, “HI-selected galaxies in the Sloan Digital Sky Survey. I. Optical Data,” *The Astronomical Journal*, vol. 139, p. 315, dec 2009.
- [14] S. Huang, M. P. Haynes, R. Giovanelli, and J. Brinchmann, “The Arecibo Legacy Fast ALFA Survey: The Galaxy Population Detected by ALFALFA,” *The Astrophysical Journal*, vol. 756, p. 113, Sept. 2012.

- [15] J. Wang, B. S. Koribalski, P. Serra, T. van der Hulst, S. Roychowdhury, P. Kamphuis, and J. N. Chengalur, “New lessons from the HI size–mass relation of galaxies,” *Monthly Notices of the Royal Astronomical Society*, vol. 460, pp. 2143–2151, 05 2016.
- [16] E. Nelson, “HI Line Profiles of Galaxies: Tilted Ring Models,” 2008.
- [17] L. Neumann, J. S. den Brok, F. Bigiel, A. Leroy, A. Usero, A. T. Barnes, I. Bešlić, C. Eibensteiner, M. Held, M. J. Jiménez-Donaire, J. Pety, E. W. Rosolowsky, E. Schinnerer, and T. G. Williams, “Spectral stacking of radio-interferometric data,” *Astronomy and Astrophysics*, vol. 675, p. A104, July 2023.
- [18] S. Fabello, *HI properties of massive galaxies from stacking*. PhD thesis, Ludwig Maximilian University of Munich, April 2012.
- [19] M. G. Jones, M. P. Haynes, R. Giovanelli, and E. Papastergis, “When is Stacking Confusing?: The Impact of Confusion on Stacking in Deep HI Galaxy Surveys,” *Monthly Notices of the Royal Astronomical Society*, vol. 455, 2016.
- [20] A. Chowdhury, N. Kanekar, J. N. Chengalur, S. Sethi, and K. S. Dwarakanath, “HI 21-centimetre emission from an ensemble of galaxies at an average redshift of one,” *Nature*, vol. 586, p. 369–372, Oct. 2020.
- [21] J. Delhaize, M. J. Meyer, L. Staveley-Smith, and B. J. Boyle, “Detection of HI in distant galaxies using spectral stacking,” *Monthly Notices of the Royal Astronomical Society*, vol. 433, p. 1398–1410, June 2013.
- [22] R. B. Tully and J. R. Fisher, “A new method of determining distances to galaxies.,” *Astronomy and Astrophysics*, vol. 54, pp. 661–673, Feb. 1977.

- [23] M. A. Zwaan, F. H. Briggs, D. Sprayberry, and E. Sorar, “The HI Mass Function of Galaxies from a Deep Survey in the 21 Centimeter Line,” *The Astrophysical Journal*, vol. 490, pp. 173–186, Nov. 1997.
- [24] M. A. Zwaan, L. Staveley-Smith, B. S. Koribalski, P. A. Henning, V. A. Kilborn, S. D. Ryder, D. G. Barnes, R. Bhathal, P. J. Boyce, W. J. G. de Blok, M. J. Disney, M. J. Drinkwater, R. D. Ekers, K. C. Freeman, B. K. Gibson, A. J. Green, R. F. Haynes, H. Jerjen, S. Juraszek, M. J. Kesteven, P. M. Knezek, R. C. Kraan-Korteweg, S. Mader, M. Marquarding, M. Meyer, R. F. Minchin, J. R. Mould, J. O’Brien, T. Oosterloo, R. M. Price, M. E. Putman, E. Ryan-Weber, E. M. Sadler, A. Schröder, I. M. Stewart, F. Stootman, B. Warren, M. Waugh, R. L. Webster, and A. E. Wright, “The 1000 Brightest HIPASS Galaxies: The H I Mass Function and  $\Omega_{HI}$ ,” *The Astronomical Journal*, vol. 125, pp. 2842–2858, June 2003.
- [25] M. G. Jones, M. P. Haynes, R. Giovanelli, and C. Moorman, “The ALFALFA HI mass function: a dichotomy in the low-mass slope and a locally suppressed ‘knee’ mass,” *Monthly Notices of the Royal Astronomical Society*, vol. 477, p. 2–17, Feb. 2018.
- [26] N. Dalal, M. White, J. R. Bond, and A. Shirokov, “Halo assembly bias in hierarchical structure formation,” *The Astrophysical Journal*, vol. 687, p. 12–21, Nov. 2008.
- [27] M. G. Jones, K. M. Hess, E. A. K. Adams, and L. Verdes-Montenegro, “The HI mass function of group galaxies in the ALFALFA survey,” *Monthly Notices of the Royal Astronomical Society*, vol. 494, p. 2090–2108, Apr. 2020.
- [28] C. Blake and K. Glazebrook, “Probing Dark Energy Using Baryonic Oscillations in the Galaxy Power Spectrum as a Cosmological Ruler,” *The Astrophysical Journal*, vol. 594, pp. 665–673, Sept. 2003.

- [29] B. F. Roukema, “The baryon acoustic oscillation peak: A flexible standard ruler,” in *The Fourteenth Marcel Grossmann Meeting*, WORLD SCIENTIFIC, Nov. 2017.
- [30] D. Collaboration, A. G. Adame, J. Aguilar, S. Ahlen, S. Alam, D. M. Alexander, M. Alvarez, O. Alves, A. Anand, U. Andrade, E. Armengaud, S. Avila, A. Aviles, H. Awan, B. Bahr-Kalus, S. Bailey, C. Baltay, A. Bault, J. Behera, S. BenZvi, A. Bera, F. Beutler, D. Bianchi, C. Blake, R. Blum, S. Brieden, A. Brodzeller, D. Brooks, E. Buckley-Geer, E. Burtin, R. Calderon, R. Canning, A. C. Rosell, R. Cereskaite, J. L. Cervantes-Cota, S. Chabanier, E. Chaussidon, J. Chaves-Montero, S. Chen, X. Chen, T. Claybaugh, S. Cole, A. Cuceu, T. M. Davis, K. Dawson, A. de la Macorra, A. de Mattia, N. Deiosso, A. Dey, B. Dey, Z. Ding, P. Doel, J. Edelstein, S. Eftekharzadeh, D. J. Eisenstein, A. Elliott, P. Fagrelius, K. Fanning, S. Ferraro, J. Ereza, N. Findlay, B. Flaugher, A. Font-Ribera, D. Forero-Sánchez, J. E. Forero-Romero, C. S. Frenk, C. Garcia-Quintero, E. Gaztañaga, H. Gil-Marín, S. G. A. Gontcho, A. X. Gonzalez-Morales, V. Gonzalez-Perez, C. Gordon, D. Green, D. Gruen, R. Gsponer, G. Gutierrez, J. Guy, B. Hadzhiyska, C. Hahn, M. M. S. Hanif, H. K. Herrera-Alcantar, K. Honscheid, C. Howlett, D. Huterer, V. Iršič, M. Ishak, S. Juneau, N. G. Karaçaylı, R. Kehoe, S. Kent, D. Kirkby, A. Kremin, A. Krolewski, Y. Lai, T. W. Lan, M. Landriau, D. Lang, J. Lasker, J. M. L. Goff, L. L. Guillou, A. Leauthaud, M. E. Levi, T. S. Li, E. Linder, K. Lodha, C. Magneville, M. Manera, D. Margala, P. Martini, M. Maus, P. McDonald, L. Medina-Varela, A. Meisner, J. Mena-Fernández, R. Miquel, J. Moon, S. Moore, J. Moustakas, N. Mudur, E. Mueller, A. Muñoz-Gutiérrez, A. D. Myers, S. Nadathur, L. Napolitano, R. Neveux, J. A. Newman, N. M. Nguyen, J. Nie, G. Niz, H. E. Noriega, N. Padmanabhan, E. Paillas, N. Palanque-Delabrouille, J. Pan, S. Penmetsa, W. J. Percival, M. Pieri, M. Pinon, C. Poppett, A. Porredon, F. Prada,

A. Pérez-Fernández, I. Pérez-Ràfols, D. Rabinowitz, A. Raichoor, C. Ramírez-Pérez, S. Ramirez-Solano, M. Rashkovetskyi, M. Rezaie, J. Rich, A. Rocher, C. Rockosi, N. A. Roe, A. Rosado-Marin, A. J. Ross, G. Rossi, R. Ruggeri, V. Ruhlmann-Kleider, L. Samushia, E. Sanchez, C. Saulder, E. F. Schlafly, D. Schlegel, M. Schubnell, H. Seo, A. Shafieloo, R. Sharples, J. Silber, A. Slosar, A. Smith, D. Sprayberry, T. Tan, G. Tarlé, P. Taylor, S. Trusov, L. A. Ureña-López, R. Vaisakh, D. Valcin, F. Valdes, M. Vargas-Magaña, L. Verde, M. Walther, B. Wang, M. S. Wang, B. A. Weaver, N. Weaverdyck, R. H. Wechsler, D. H. Weinberg, M. White, J. Yu, Y. Yu, S. Yuan, C. Yèche, E. A. Zaborowski, P. Zarrouk, H. Zhang, C. Zhao, R. Zhao, R. Zhou, T. Zhuang, and H. Zou, “Desi 2024 vi: Cosmological constraints from the measurements of baryon acoustic oscillations,” 2024.

- [31] F. B. Abdalla and S. Rawlings, “Probing dark energy with baryonic oscillations and future radio surveys of neutral hydrogen,” *Monthly Notices of the Royal Astronomical Society*, vol. 360, p. 27–40, June 2005.
- [32] A. Begum, J. N. Chengalur, I. D. Karachentsev, M. E. Sharina, and S. S. Kaisin, “FIGGS: Faint Irregular Galaxies GMRT Survey – overview, observations and first results,” *Monthly Notices of the Royal Astronomical Society*, vol. 386, p. 1667–1682, May 2008.
- [33] Noordermeer, E., van der Hulst, J. M., Sancisi, R., Swaters, R. A., and van Albada, T. S., “The Westerbork HI survey of spiral and irregular galaxies - III. HI observations of early-type disk galaxies,” *Astronomy and Astrophysics*, vol. 442, no. 1, pp. 137–157, 2005.

- [34] J. M. Cannon, R. Giovanelli, M. P. Haynes, S. Janowiecki, A. Parker, J. J. Salzer, E. A. K. Adams, E. Engstrom, S. Huang, K. B. W. McQuinn, J. Ott, A. Saintonge, E. D. Skillman, J. Allan, G. Erny, P. Fliss, and A. Smith, “The survey of HI in extremely low-mass dwarfs (SHIELD),” *The Astrophysical Journal*, vol. 739, p. L22, Aug. 2011.
- [35] J. J. Salzer, N. M. Tresser, J. L. Rosenberg, and S. Stevenson, “An Optically Unbiased Look at the Local Universe: Properties of the ADBS HI-Selected Sample of Galaxies.,” in *American Astronomical Society Meeting Abstracts*, vol. 207 of *American Astronomical Society Meeting Abstracts*, p. 179.19, Dec. 2005.
- [36] D. G. Barnes, L. Staveley-Smith, W. J. G. de Blok, T. Oosterloo, I. M. Stewart, A. E. Wright, G. D. Banks, R. Bhathal, P. J. Boyce, M. R. Calabretta, M. J. Disney, M. J. Drinkwater, R. D. Ekers, K. C. Freeman, B. K. Gibson, A. J. Green, R. F. Haynes, P. te Lintel Hekkert, P. A. Henning, H. Jerjen, S. Juraszek, M. J. Kesteven, V. A. Kilborn, P. M. Knezek, B. Koribalski, R. C. Kraan-Korteweg, D. F. Malin, M. Marquarding, R. F. Minchin, J. R. Mould, R. M. Price, M. E. Putman, S. D. Ryder, E. M. Sadler, A. Schröder, F. Stootman, R. L. Webster, W. E. Wilson, and T. Ye, “The HI Parkes All Sky Survey: southern observations, calibration and robust imaging,” *Monthly Notices of the Royal Astronomical Society*, vol. 322, pp. 486–498, Apr. 2001.
- [37] M. J. Meyer, M. A. Zwaan, R. L. Webster, L. Staveley-Smith, E. Ryan-Weber, M. J. Drinkwater, D. G. Barnes, M. Howlett, V. A. Kilborn, J. Stevens, M. Waugh, M. J. Pierce, R. Bhathal, W. J. G. de Blok, M. J. Disney, R. D. Ekers, K. C. Freeman, D. A. Garcia, B. K. Gibson, J. Harnett, P. A. Henning, H. Jerjen, M. J. Kesteven, P. M. Knezek, B. S. Koribalski, S. Mader, M. Marquarding, R. F. Minchin, J. O’Brien, T. Oosterloo, R. M. Price, M. E. Putman, S. D. Ryder, E. M. Sadler, I. M. Stewart,

- F. Stootman, and A. E. Wright, “The HIPASS catalogue - I. Data presentation,” *Monthly Notices of the Royal Astronomical Society*, vol. 350, pp. 1195–1209, June 2004.
- [38] O. I. Wong, E. V. Ryan-Weber, D. A. Garcia-Appadoo, R. L. Webster, L. Staveley-Smith, M. A. Zwaan, M. J. Meyer, D. G. Barnes, V. A. Kilborn, R. Bhathal, W. J. G. De Blok, M. J. Disney, M. T. Doyle, M. J. Drinkwater, R. D. Ekers, K. C. Freeman, B. K. Gibson, S. Guovich, J. Harnett, P. A. Henning, H. Jerjen, M. J. Kesteven, P. M. Knezek, B. S. Koribalski, S. Mader, M. Marquarding, R. F. Minchin, J. O’Brien, M. E. Putman, S. D. Ryder, E. M. Sadler, J. Stevens, I. M. Stewart, F. Stootman, and M. Waugh, “The Northern HIPASS catalogue - data presentation, completeness and reliability measures,” *Monthly Notices of the Royal Astronomical Society*, vol. 371, p. 1855–1864, Oct. 2006.
- [39] R. Giovanelli, M. P. Haynes, B. R. Kent, P. Perillat, A. Saintonge, N. Brosch, B. Catinella, G. L. Hoffman, S. Stierwalt, K. Spekkens, M. S. Lerner, K. L. Masters, E. Momjian, J. L. Rosenberg, C. M. Springob, A. Boselli, V. Charmandaris, J. K. Darling, J. Davies, D. Garcia Lambas, G. Gavazzi, C. Giovanardi, E. Hardy, L. K. Hunt, A. Iovino, I. D. Karachentsev, V. E. Karachentseva, R. A. Koopmann, C. Marinoni, R. Minchin, E. Muller, M. Putman, C. Pantoja, J. J. Salzer, M. Scodeggio, E. Skillman, J. M. Solanes, C. Valotto, W. van Driel, and L. van Zee, “The Arecibo Legacy Fast ALFA Survey. I. Science Goals, Survey Design, and Strategy,” *The Astrophysical Journal*, vol. 130, pp. 2598–2612, Dec. 2005.
- [40] D. A. Roshi, N. Aponte, E. Araya, H. Arce, L. A. Baker, W. Baan, T. M. Becker, J. K. Breakall, R. G. Brown, C. G. M. Brum, M. Busch, D. B. Campbell, T. Cohen, F. Cordova, J. S. Deneva, M. Devogele, T. Dolch, F. O. Fernandez-Rodriguez, T. Ghosh, P. F. Goldsmith, L. I. Gurvits, M. Haynes, C. Heiles, J. W. T. Hessel,

- D. Hickson, B. Isham, R. B. Kerr, J. Kelly, J. J. Kiriazes, J. Lautenbach, M. Lebron, N. Lewandowska, L. Magnani, P. K. Manoharan, J. L. Margot, S. E. Marshall, A. K. McGilvray, A. Mendez, R. Minchin, V. Negron, M. C. Nolan, L. Olmi, F. Paganelli, N. T. Palliyaguru, C. A. Pantoja, Z. Paragi, S. C. Parshley, J. E. G. Peek, B. B. P. Perera, P. Perillat, N. Pinilla-Alonso, L. Quintero, H. Radovan, S. Raizada, T. Robishaw, M. Route, C. J. Salter, A. Santoni, P. Santos, S. Sau, D. Selvaraj, A. J. Smith, M. Sulzer, S. Vaddi, F. Vargas, F. C. F. Venditti, A. Venkataraman, H. Verkouter, A. K. Virkki, A. Vishwas, S. Weinreb, D. Werthimer, A. Wolszczan, and L. F. Zambrano-Marin, “The Future Of The Arecibo Observatory: The Next Generation Arecibo Telescope,” 2021.
- [41] M. P. Haynes, R. Giovanelli, B. R. Kent, E. A. K. Adams, T. J. Balonek, D. W. Craig, D. Fertig, R. Finn, C. Giovanardi, G. Hallenbeck, K. M. Hess, G. L. Hoffman, S. Huang, M. G. Jones, R. A. Koopmann, D. A. Kornreich, L. Leisman, J. Miller, C. Moorman, J. O’Connor, A. O’Donoghue, E. Papastergis, P. Troischt, D. Stark, and L. Xiao, “The Arecibo Legacy Fast ALFA Survey: The ALFALFA Extragalactic HI Source Catalog,” *The Astrophysical Journal*, vol. 861, p. 49, July 2018.
- [42] J. Kang, M. Zhu, M. Ai, H. Yu, and C. Sun, “Extragalactic HI Survey with FAST: First look at the Pilot Survey Results,” *Research in Astronomy and Astrophysics*, vol. 22, p. 065019, June 2022.
- [43] S. Yahya, P. Bull, M. G. Santos, M. Silva, R. Maartens, P. Okouma, and B. Bassett, “Cosmological performance of SKA HI galaxy surveys,” *Monthly Notices of the Royal Astronomical Society*, vol. 450, p. 2251–2260, May 2015.
- [44] N. Maddox, B. S. Frank, A. A. Ponomareva, M. J. Jarvis, E. A. K. Adams, R. Davé, T. A. Oosterloo, M. G. Santos, S. L. Blyth, M. Glowacki, R. C. Kraan-Korteweg,

- W. Mulaudzi, B. Namumba, I. Prandoni, S. H. A. Rajohnson, K. Spekkens, N. J. Adams, R. A. A. Bowler, J. D. Collier, I. Heywood, S. Sekhar, and A. R. Taylor, “MIGHTEE-HI: The H I emission project of the MeerKAT MIGHTEE survey,” *Astronomy and Astrophysics*, vol. 646, p. A35, Feb. 2021.
- [45] A. A. Ponomareva, M. J. Jarvis, H. Pan, N. Maddox, M. G. Jones, B. S. Frank, S. H. A. Rajohnson, W. Mulaudzi, M. Meyer, E. A. K. Adams, M. Baes, K. M. Hess, S. Kurapati, I. Prandoni, F. Sinigaglia, K. Spekkens, M. Tudorache, I. Heywood, J. D. Collier, and S. Sekhar, “MIGHTEE-HI: the first MeerKAT HI mass function from an untargeted interferometric survey,” *Monthly Notices of the Royal Astronomical Society*, vol. 522, p. 5308–5319, May 2023.
- [46] F. Sinigaglia, G. Rodighiero, E. Elson, M. Vaccari, N. Maddox, B. S. Frank, M. J. Jarvis, T. Oosterloo, R. Davé, M. Salvato, M. Baes, S. Bellstedt, L. Bisigello, J. D. Collier, R. H. W. Cook, L. J. M. Davies, J. Delhaize, S. P. Driver, C. Foster, S. Kurapati, C. d. P. Lagos, C. Lidman, P. E. Mancera Piña, M. J. Meyer, K. M. Mogotsi, H. Pan, A. A. Ponomareva, I. Prandoni, S. H. A. Rajohnson, A. S. G. Robotham, M. G. Santos, S. Sekhar, K. Spekkens, J. E. Thorne, J. M. van der Hulst, and O. I. Wong, “MIGHTEE-HI: Evolution of HI scaling relations of star-forming galaxies at  $z < 0.5$ ,” *The Astrophysical Journal Letters*, vol. 935, p. L13, Aug. 2022.
- [47] A. R. Duffy, M. J. Meyer, L. Staveley-Smith, M. Bernyk, D. J. Croton, B. S. Koribalski, D. Gerstmann, and S. Westerlund, “Predictions for ASKAP neutral hydrogen surveys: ASKAP surveys,” *Monthly Notices of the Royal Astronomical Society*, vol. 426, p. 3385–3402, Oct. 2012.

- [48] B. S. Koribalski, L. Staveley-Smith, T. Westmeier, P. Serra, K. Spekkens, O. I. Wong, K. Lee-Waddell, C. D. P. Lagos, D. Obreschkow, E. V. Ryan-Weber, M. Zwaan, V. Kilborn, G. Bekiaris, K. Bekki, F. Bigiel, A. Boselli, A. Bosma, B. Catinella, G. Chauhan, M. E. Cluver, M. Colless, H. M. Courtois, R. A. Crain, W. J. G. de Blok, H. Dénes, A. R. Duffy, A. Elagali, C. J. Fluke, B. Q. For, G. Heald, P. A. Henning, K. M. Hess, B. W. Holwerda, C. Howlett, T. Jarrett, D. H. Jones, M. G. Jones, G. I. G. Józsa, R. Jurek, E. Jütte, P. Kamphuis, I. Karachentsev, J. Kerp, D. Kleiner, R. C. Kraan-Korteweg, Á. R. López-Sánchez, J. Madrid, M. Meyer, J. Mould, C. Murugeshan, R. P. Norris, S. H. Oh, T. A. Oosterloo, A. Popping, M. Putman, T. N. Reynolds, J. Rhee, A. S. G. Robotham, S. Ryder, A. C. Schröder, L. Shao, A. R. H. Stevens, E. N. Taylor, J. M. van der Hulst, L. Verdes-Montenegro, B. P. Wakker, J. Wang, M. Whiting, B. Winkel, and C. Wolf, “WALLABY - an SKA Pathfinder HI survey,” *Astrophysics and Space Science*, vol. 365, p. 118, July 2020.
- [49] R. Ansari, “Current status and future of cosmology with 21cm Intensity Mapping,” 2022.
- [50] V. MacKay, M. Lai, P. Shmerko, D. Wulf, L. Belostotski, and K. Vanderlinde, “Low-cost, Low-loss, Ultra-wideband Compact Feed for Interferometric Radio Telescopes,” *Journal of Astronomical Instrumentation*, 2023.
- [51] A. Liu and J. R. Shaw, “Data analysis for precision 21 cm cosmology,” *Publications of the Astronomical Society of the Pacific*, vol. 132, p. 062001, Apr. 2020.
- [52] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L.

- Wright, “Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results,” *The Astrophysical Journal Supplement*, vol. 208, p. 19, Oct. 2013.
- [53] H. Shin, *CHANNELIZATION TECHNIQUES FOR WIDEBAND RADIOS*. PhD thesis, University OF Minnesota, May 2017.
- [54] K. A. Oman, “The ALFALFA HI velocity width function,” *Monthly Notices of the Royal Astronomical Society*, vol. 509, p. 3268–3284, Nov. 2021.
- [55] M. Tegmark, “How to make maps from cosmic microwave background data without losing information,” *The Astrophysical Journal*, vol. 480, p. L87–L90, May 1997.
- [56] L. Zhang, M. Zhang, and B. Wang, “A Model Estimator for Noisy Compact Emission Recovery in Radio Synthesis Imaging,” *The Astrophysical Journal*, vol. 166, p. 53, Aug. 2023.
- [57] F. Rahman, P. Chingangbam, and T. Ghosh, “Statistical properties of galactic synchrotron temperature and polarization maps - a multi-frequency comparison,” *Journal of Cosmology and Astroparticle Physics*, vol. 2024, p. 036, Jan. 2024.
- [58] C. G. T. Haslam, C. J. Salter, H. Stoffel, and W. E. Wilson, “A 408-MHZ All-Sky Continuum Survey. II. The Atlas of Contour Maps,” *Astronomy and Astrophysics Supplement*, vol. 47, p. 1, Jan. 1982.
- [59] M.-A. Miville-Deschénes, N. Ysard, A. Lavabre, N. Ponthieu, J. F. Macías-Pérez, J. Aumont, and J. P. Bernard, “Separation of anomalous and synchrotron emissions using WMAP polarization data,” *Astronomy and Astrophysics*, vol. 490, p. 1093–1102, Sept. 2008.

- [60] J. J. Condon, W. D. Cotton, E. W. Greisen, Q. F. Yin, R. A. Perley, G. B. Taylor, and J. J. Broderick, “The NRAO VLA Sky Survey,” *The Astronomical Journal*, vol. 115, pp. 1693–1716, May 1998.
- [61] A. S. Cohen, W. M. Lane, W. D. Cotton, N. E. Kassim, T. J. W. Lazio, R. A. Perley, J. J. Condon, and W. C. Erickson, “The VLA Low-Frequency Sky Survey,” *The Astronomical Journal*, vol. 134, pp. 1245–1262, Sept. 2007.
- [62] T. Di Matteo, R. Perna, T. Abel, and M. J. Rees, “Radio Foregrounds for the 21 Centimeter Tomography of the Neutral Intergalactic Medium at High Redshifts,” *The Astrophysical Journal*, vol. 564, p. 576–580, Jan. 2002.
- [63] M. G. Santos, A. Cooray, and L. Knox, “Multifrequency analysis of 21 centimeter fluctuations from the era of reionization,” *The Astrophysical Journal*, vol. 625, p. 575–587, June 2005.
- [64] R. A. N. Brooks, K. A. Oman, and C. S. Frenk, “The north–south asymmetry of the ALFALFA HI velocity width function,” *Monthly Notices of the Royal Astronomical Society*, vol. 522, p. 4043–4058, Apr. 2023.
- [65] D. Tamburro, H.-W. Rix, A. K. Leroy, M.-M. M. Low, F. Walter, R. C. Kennicutt, E. Brinks, and W. J. G. de Blok, “What is driving the HI velocity dispersion?,” *The Astronomical Journal*, vol. 137, p. 4424–4435, Apr. 2009.
- [66] C. Bacchini, F. Fraternali, G. Pezzulli, G. Iorio, A. Marasco, and V. Nipoti, “Evidence for supernova feedback sustaining gas turbulence in nearby star-forming galaxies,” *Proceedings of the International Astronomical Union*, vol. 17, p. 199–202, Aug. 2021.

- [67] A. K. Leroy, F. Walter, E. Brinks, F. Bigiel, W. J. G. de Blok, B. Madore, and M. D. Thornley, “The Star Formation Efficiency in Nearby Galaxies: Measuring Where Gas Forms Stars Effectively,” *The Astronomical Journal*, vol. 136, pp. 2782–2845, Dec. 2008.
- [68] S. Kazemi, S. Yatawatta, S. Zaroubi, P. Lampropoulos, A. G. de Bruyn, L. V. E. Koopmans, and J. Noordam, “Radio interferometric calibration using the SAGE algorithm,” *Monthly Notices of the Royal Astronomical Society*, vol. 414, p. 1656–1666, Apr. 2011.
- [69] J. S. Dillon, M. Lee, Z. S. Ali, A. R. Parsons, N. Orosz, C. D. Nunhokee, P. La-Plante, A. P. Beardsley, N. S. Kern, Z. Abdurashidova, J. E. Aguirre, P. Alexander, Y. Balfour, G. Bernardi, T. S. Billings, J. D. Bowman, R. F. Bradley, P. Bull, J. Burba, S. Carey, C. L. Carilli, C. Cheng, D. R. DeBoer, M. Dexter, E. de Lera-Acedo, J. Ely, A. Ewall-Wice, N. Fagnoni, R. Fritz, S. R. Furlanetto, K. Gale-Sides, B. Glendenning, D. Gorthi, B. Greig, J. Grobbelaar, Z. Halday, B. J. Hazelton, J. N. Hewitt, J. Hickish, D. C. Jacobs, A. Julius, J. Kerrigan, P. Kittiwisit, S. A. Kohn, M. Kolopanis, A. Lanman, T. Lekalake, D. Lewis, A. Liu, Y.-Z. Ma, D. MacMahon, L. Malan, C. Malgas, M. Maree, Z. E. Martinot, E. Matsetela, A. Mesinger, M. Molewa, M. F. Morales, T. Mosiane, S. Murray, A. R. Neben, B. Nikolic, R. Pascua, N. Patra, S. Pieterse, J. C. Pober, N. Razavi-Ghods, J. Ringuette, J. Robnett, K. Rosie, M. G. Santos, P. Sims, C. Smith, A. Syce, M. Tegmark, N. Thyagarajan, P. K. G. Williams, and H. Zheng, “Redundant-baseline calibration of the hydrogen epoch of reionization array,” *Monthly Notices of the Royal Astronomical Society*, vol. 499, p. 5840–5861, Oct. 2020.
- [70] J. L. Sievers, “Calibration of quasi-redundant interferometers,” 2017.

- [71] L. Feng, *A matched filter technique for slow radio transient detection and first demonstration with the Murchison Widefield Array*. PhD thesis, Massachusetts Institute of Technology, 2016.
- [72] R. Vio, P. Andreani, and W. Wamsteker, “Some good reasons to use matched filters for the detection of point sources in CMB maps,” *Astronomy and Astrophysics*, vol. 414, p. 17–21, Jan. 2004.
- [73] J. Roman, M. Rangaswamy, D. Davis, Q. Zhang, B. Himed, and J. Michels, “Parametric adaptive matched filter for airborne radar applications,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 677–692, 2000.
- [74] G. Turin, “An introduction to matched filters,” *IRE Transactions on Information Theory*, vol. 6, no. 3, pp. 311–329, 1960.
- [75] J. Janquart, *Gravitational waves signal analysis: Matched filtering, typical analyses and beyond*. PhD thesis, Université de Liège, 2020.
- [76] M. Bazdresch, “An elucidating proof that a matched filter is optimum,” *International Journal of Electrical Engineering & Education*, vol. 59, pp. 60 – 70, 2018.
- [77] A. Saintonge, “The Arecibo Legacy Fast ALFA Survey. IV. Strategies for Signal Identification and Survey Catalog Reliability,” *The Astronomical Journal*, vol. 133, p. 2087–2096, Mar. 2007.
- [78] I. Zubeldia, A. Rotti, J. Chluba, and R. Battye, “Understanding matched filters for precision cosmology,” *Monthly Notices of the Royal Astronomical Society*, vol. 507, p. 4852–4863, Sept. 2021.