Cosmology and HI Intensity Mapping

by

John Podczerwinski

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

Doctor of Philosophy

(Physics)

at the

University of Wisconsin–Madison

2024

Date of final oral examination: 08/27/2024

The dissertation is approved by the following members of the Final Oral Committee:
Peter Timbie, Professor, Physics
Moritz Munchmeyer, Associate Professor, Physics
Snežana Stanimirović, Professor, Astronomy
Keith Bechtol, Associate Professor, Physics

ACKNOWLEDGMENTS

I would like to thank Peter for his collaboration, advice and support over the years. He has helped all along the way, providing advice on research, writing and presenting. I feel lucky to have had him as an advisor.

I would like to thank my parents for all of their support during my time in school. I likely would not have made it this far without having them behind me throughout my academic career.

Thank you to my friends, girlfriend and family for making life fun.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This thesis is focused on HI intensity mapping, a proposed method for making three dimensional surveys of the universe. I begin by reviewing basic cosmological theory and common observational techniques such as galaxy surveys and observations of the cosmic microwave background (CMB). I then provide motivation for the use of HI intensity mapping along with a description of how such measurements are performed. I describe my research work, which was focused on developing instrumentation and analysis techniques for HI intensity mapping in the post reionization era of the universe ($z < 6$). A key goal of such observations is to monitor the baryon acoustic oscillations over a wide range of redshifts, providing information about the evolution of dark energy. My work focused mainly on the design of an ultra-wideband antenna and the development of foreground cleaning techniques. For the antenna project, I created an ultra-wideband design that was optimized with intensity mapping systematic effects in mind. This involved the use of a Vivaldi feed with optimized profile and a deep dish with and an elliptical collar. I also investigated the use of absorbing materials on the feed antenna, which helped mitigate systematic effects but resulted in significantly lower sensitivity. The final antenna design achieved satisfactory levels of spectral smoothness, impedance matching and sensitivity to the 21 cm hydrogen line. In simulations, I found that my design efforts helped reduce coupling between neighboring antenna elements, though not to the level desired. For the foreground cleaning projects, I investigated the use of Karhunen-Loève transforms in both map-space and visibility space. My project was the first to incorporate a KL transform in map space. This work, however, did not seem to provide a substantial improvement over other available methods. I then conclude by discussing ideas on the future of HI intensity mapping.

1 INTRODUCTION

This thesis describes work pertaining to 21 cm intensity mapping, a technique used in observational cosmology. In this introductory chapter, I will provide some necessary background to understand the goals and motivation of 21 cm intensity mapping. In chapter 2, I review traditional techniques used to probe Large Scale Structure (LSS). In chapter 3, I describe in greater detail the scientific promise of 21 cm mapping and in chapter 4 I provide information on how the technique works. In chapter 5, I describe systematic effects that are present in 21 cm intensity mapping and summarize attempts that have been made at mitigating them. In chapter 6, I present an ultra-wideband antenna I designed with the previously mentioned systematic effects in mind. In chapter 7, I present the Needlet-Karhunen-Loève (NKL) technique, a method I developed for removing foregrounds from 21 cm intensity maps. In Chapter 8 I describe Visibility Karhunen-Loève (VKL) a technique similar to NKL that works directly on raw interferometer data (visibilities) rather than maps. The thesis then concludes with chapter 9, where I summarize key points made in the thesis and mention some ideas I have for the future of 21 cm intensity mapping.

Let's now review the basic ideas of cosmology along with some questions of interest to researchers in the field. We will begin by reviewing the story of the universe as a simple list.

- The story of the universe told by cosmology begins with **inflation**. During this time, the universe expands exponentially quickly. It is hypothesized that this expansion was caused by one or more quantum fields. At the end of this period, the field(s) driving inflation decay into matter and radiation. This event is referred as **reheating**. Perturbations in the field(s) lead to perturbations in the spatial distribution of matter and radiation. Over time, these perturbations will grow into the structure we see in the universe today. This period is hypothesized to have lasted on the order of 10^{-36} to 10^{-32} seconds.

- After reheating, we enter the era of **Big Bang Nucleosynthesis (BBN)**. During this time, temperatures are still quite hot and matter and radiation are tightly coupled. Once the universe has adequately cooled, interaction rates become low enough that neutral atoms begin to form. This event is referred to as **recombination** and occurs at 375,000 years after inflation ($z \approx 1100$). This results in the Cosmic Microwave Background (CMB) that is observed by cosmologists today.

- After recombination, we enter a period known as the cosmic **Dark Ages**. The dark ages last from $z \approx 1100$ until $z \approx 20$, which corresponds to about 400 million years after the big bang. During this time, structure continues to grow, but we have not yet formed complex objects such as stars or galaxies. This era is referred to as the “dark ages" since it happens after the CMB was emitted but before luminous objects form.

- At $z \approx 20$, structure has grown enough that complex objects such as galaxies and stars begin to form. These newly formed objects emit radiation that ionizes the neutral hydrogen present in the universe and is thus referred to as the “epoch of reionization". From this point on, neutral hydrogen is mostly found in relatively dense pockets within galaxies where it is shielded from the radiation. At around $z \approx 6$ (or about 1 billion years after the big bang) the ionization process is complete, and we enter the final and most recent epoch.
- Lastly, we enter the post-EOR era. This covers $z < 6$ or between 1 and about 13.7 billion years after the big bang. During this time, structure continues to grow and we find, for $z < 1.5$, that the expansion of the universe is once again accelerating. This accelerating expansion is hypothesized to be due to “dark energy" which readers will hear about more later in this thesis.

Next, I will discuss each of these epochs in a bit more detail. However, it would be beneficial to discuss some prerequisites before doing so.

1.1 Some Basic Pre-Requisities

Before discussing the history of the universe, it would be useful to begin by covering some basic pre-requisites for understanding the topic. First, observations of the universe show that it is homogeneous and isotropic to very good approximation on large scales. As such, according to general relativity, it can be described using the Friedmann–Lemaître–Robertson–Walker (FLRW) metric

$$ds^2 = dt^2 - a^2(t)\left[\frac{dr^2}{1-kr^2} + S_k^2(r)d\Omega^2\right]. \tag{1.1}$$

In this formula, we set $c = 1$. Note that k is the curvature of space-time and $S_k^2(r)$ is given by

$$S(k) = \begin{cases} \sqrt{k}^{-1} sin(r\sqrt{k}) & k > 0 \\ r & k = 0 \\ \sqrt{|k|}^{-1} sinh(r\sqrt{|k|}) & k < 0. \end{cases} \tag{1.2}$$

$k = 0$ corresponds to a “flat universe", $k > 0$ to a “closed universe" and $k < 0$ to an “open universe". Current observations show that the universe is very close to being flat, which is to say $k = 0$. When $k = 0$, we find that

$$ds^2 = dt^2 - a^2(t)\left[dr^2 + r^2 d\Omega^2\right]. \tag{1.3}$$

the function $a(t)$ describes the expansion (or possibly contraction) of the universe. This function is referred to as the “scale factor". The effect of a is illustrated nicely in figure 1.1, which is borrowed from (Dodelson (2003)). In this figure, we have two points sitting on the vertices of a grid. At all times, these points will always be separated by one grid unit. We refer to the number of grid units between two points as the “comoving distance", which is denoted by r. However, this grid will be growing over time in the case of our universe. This causes the two grid points to become physically further apart. This physical distance is often referred to as “proper distance". Put simply, we find

$$proper\ distance = comoving\ distance\ *\ a. \tag{1.4}$$

Next, we should discuss what determines the evolution of a with respect to time. To see this we’ll start with Einstein’s equations

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathcal{R} = 8\pi T_{\mu\nu}. \tag{1.5}$$

In this formula, $G_{\mu\nu}$ is “Einstein’s tensor" and $R_{\mu\nu}$ is the “Ricci tensor", which contains information about derivatives of the metric, which is $g_{\mu\nu}$. We also have $\mathcal{R}$, which is the “Ricci scalar". The Ricci scalar can be computed by contracting $R_{\mu\nu}$ with respect to its two indices. Last, we have the energy-momentum tensor $T_{\mu\nu}$, which gives the flux of component μ of the momentum vector across a surface with constant coordinate x_ν. In the case of an isotropic fluid (which we are assuming right now), we will find that $T^0_0 = \rho$ and $T^i_i = P$, where ρ is the energy density and P is the pressure. If we consider the 00 component of Einstein’s equations, we find

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3} \rho. \tag{1.6}$$

Thus, we find that the rate at which the universe expands is dependent upon the energy density. In cosmology, we refer to $\dot{a}/a$ as the “Hubble rate" and denote it as H. It should be noted that the derivation of equation 1.6 ignored curvature and did not include any cosmological constant. To be complete, we should present the full first Friedman equation

$$H^2 = \frac{8\pi G}{3} \rho + \frac{k}{a^2}. \tag{1.7}$$

We now include $\frac{k}{a^2}$, which is a term accounting for the effects of curvature of the expansion. We can also expand ρ into a sum of components

$$\rho(a) = \rho_m(a) + \rho_r(a) + \rho_{DE}(a) = \frac{\rho_m(a=1)}{a^3} + \frac{\rho_r(a=1)}{a^4} + \rho_{DE}(a) \tag{1.8}$$

In this formula, ρ_m is the matter energy density, which scales as a^{-3}. This scaling makes sense as the number of matter particles is conserved while the volume grows as a^3. Thus, the amount of matter per volume should evolve like a^{-3}. We find that the radiation density ρ_r scales as a^{-4}. This is because the expansion of the space both lowers the photon number density (providing a factor of a^{-3}) and also redshifts the photons. This redshifting reduces the energy per photon and provides an additional factor of a^{-1}, giving a^{-4} scaling. The function $\rho_{DE}(a)$ describes the density of “dark energy". The way dark energy evolves over time is still being investigated. One possibility that is commonly discussed and supported by observations is having dark energy be due to a “cosmological constant" with $\rho_{DE} = \frac{\Lambda}{3\pi G}$ for some constant Λ.

However, the universe is not completely homogeneous and isotropic. There also exist perturbations in the distributions of matter and radiation. One often sees these perturbations described in terms of an “overdensity"

$$\delta_s(\mathbf{x}) \equiv \frac{\rho_s(\mathbf{x}) - \overline{\rho}_s}{\overline{\rho}_s}. \tag{1.9}$$

In this formula, δ_s is the overdensity of species s and $\overline{\rho}_s$ is the average density. As we will see later on in this thesis, these perturbations contain a lot of information about the universe and are of utmost interest to observational cosmologists. Cosmologists usually extract information from the statistical properties of δ. The most common approach is to switch the analysis into k-space by considering

$$\delta(\mathbf{k}) = \int \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d^3x. \tag{1.10}$$

The most common statistic to think about is the “power spectrum". For a homogeneous and isotropic field, one should find that

$$\langle \delta(\mathbf{k}) \delta(\mathbf{k}')^* \rangle = (2\pi^3) P(k) \delta^3(\mathbf{k} - \mathbf{k}'). \tag{1.11}$$

On the right side of the equation we have $P(k)$, which is referred to as the “power spectrum". Since $P(k)$ involves two δ terms, we say it tells us about the “two point statistics" of the matter distribution. The power spectrum tells us about the magnitude of perturbations at length scale $\sim \frac{1}{k}$. It is understood that the distribution of matter and radiation is close to Gaussian, meaning that most statistical information about perturbations is contained in $P(k)$. However, there likely exists a small non-Gaussian component to the perturbations. 3-point correlations are the simplest statistic one could use for investigating non-Gaussianity and thus are commonly seen in the literature.

Such density perturbations also give rise to perturbations in the FLRW metric. The simplest way to describe this effect is through scalar perturbations

$$g_{00}(\mathbf{x}, t) = -1 - 2\Psi(\mathbf{x}, t) \tag{1.12}$$

$$g_{0i} = 0 \tag{1.13}$$

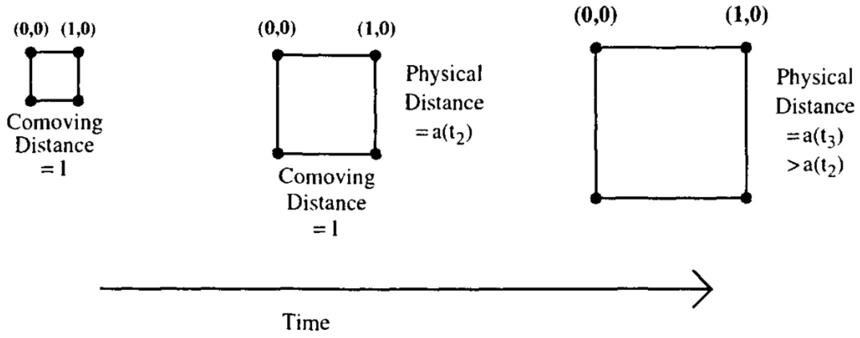

Figure 1.1: An illustration of the scale factor $a(t)$. Borrowed from Dodelson (2003).

$$g_{ij} = a^2\delta_{ij}\left[1 + 2\Phi(\mathbf{x}, t)\right]. \tag{1.14}$$

In this formula, Ψ corresponds to the Newtonian potential and Φ describes perturbations to the scale factor a at different locations. In principle, one should also include “tensor perturbations", which describe gravitational waves. However, as mentioned in (Dodelson (2003)), scalar perturbations are most important in this context. It should be noted that the numerical values for Ψ and Φ are quite small, being on the order of 10^{-4}. As such, one can achieve fairly accurate results working only to linear order in perturbations. This approach eventually breaks down once perturbations become larger.

So far, we have found through Einstein’s equations that there is some relationship between the energy density and the metric. This applies for all of our energy sources, including matter, radiation and dark energy. However, it should also be the case that the constituents of the universe interact with each other. Examples of this include Compton scattering between photons and matter or perhaps Coulomb scattering between electrons and protons. Such interactions are described according to the Boltzmann equation

$$\frac{df}{dt} = C\left[f\right]. \tag{1.15}$$

In this formula, $f(\mathbf{x}, \mathbf{p}, t)$ is the distribution function for the particle species of interest. This function describes the number of particles in a phase space element:

$$N(\mathbf{x}, \mathbf{p}, t) = f(\mathbf{x}, \mathbf{p}, t)d^3xd^3p/(2\pi)^3. \tag{1.16}$$

The term $C\left[f\right]$ describes collisions. Thus, equation 1.15 is saying that the change in the number of particles depends on the collision terms. Without collisions, the total number of particles in phase space will be conserved. No particles are created or destroyed – they simply move around in phase space. However, introducing collisions changes this.

One type of collision one could see is Compton scattering, where

$$e^-(\mathbf{q}) + \gamma(\mathbf{p}) \leftrightarrow e^-(\mathbf{q}') + \gamma(\mathbf{p}'). \tag{1.17}$$

In other words, we are considering having a photon of momentum $\mathbf{p}$ scatter off an electron with momentum $\mathbf{q}$. If $f(p)$ is the distribution function for photons, we would find that, assuming Compton scattering is the only type of collision,

$$\begin{aligned} C\left[f(\mathbf{p})\right] \propto \int \frac{d^3q}{E_e(q)} \int \frac{d^3q'}{E_e(q')} \int \frac{d^3p'}{E(p')} |\mathcal{M}|^2 \\ \times \delta^3(\mathbf{p} + \mathbf{q} - \mathbf{p}' - \mathbf{q}')\delta(E(p) - E_e(q) - E(p') - E_e(q')) \\ \times \left[f_e(\mathbf{q}')f(\mathbf{p}') - f_e(\mathbf{q})f(\mathbf{p})\right]. \end{aligned} \tag{1.18}$$

In this formula, we integrate over all possible momenta for the input electron and output electron and photon. In other words, we are summing over all possible Compton scattering events that might take place. Roughly speaking, the factor of $|\mathcal{M}|^2$ is telling us how likely such a Compton scattering event is to take place. The Dirac delta functions on the second line enforces conservation of momentum and energy. The final line accounts for the densities of electrons and photons at the necessary momenta for this interaction to take place. The first term accounts for the fact that having more electrons and photons at momenta $\mathbf{q}'$ and $\mathbf{p}'$ means that there are more opportunities to have a Compton scattering event that results in a photon with momentum $\mathbf{p}$. The second term of the third line accounts for the fact that Compton scattering can also send a photon from momentum $\mathbf{p}$ to some other momentum $\mathbf{p}'$.

Basically, all of the scattering terms one finds will follow a schematic that looks like the one shown in equation 1.18. In a nutshell, we sum over all possible interactions, enforce energy and momentum conservation and account for both creation and annihilation.

In summary, we find that the energy present in the universe has a relationship to the metric that is described using Einstein's equations. Although the universe is mostly uniform, there will also exist perturbations in the energy density (δ) that lead to metric perturbations (Ψ and Φ). These perturbations are usually small enough that we may work to linear order. Moreover, we find that different constituents of the universe will interact with each other through collisions. The effect of these collisions can be accounted for using Boltzmann's equation 1.15. The Dodelson (2003) textbook includes a nice summary that I'll show in figure 1.2. I added the extra dark energy bubble myself, which I should say is present in the Dodelson and Schmidt (2020). Dark energy has a dashed arrow because it will only interact with the background metric in the case of a cosmological constant.

I would like to define one last important quantity before we begin. Suppose we have a photon

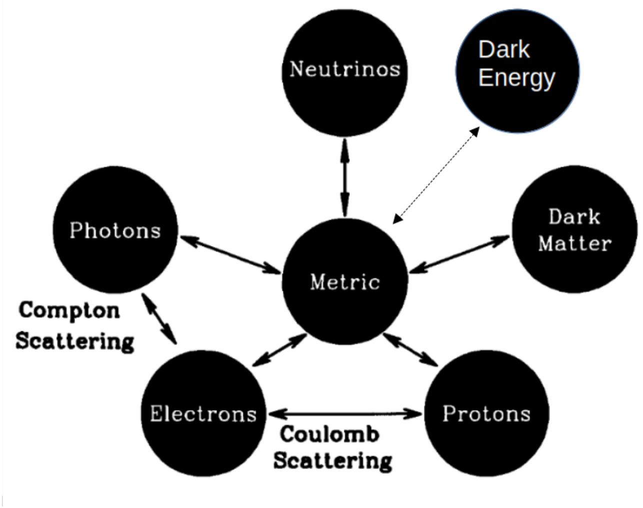

Figure 1.2: Components of the universe, including dark energy. Excludes neutrons and interactions of neutrinos with other constituents, both of which are important during nucleosynthesis. This image is borrowed from (Dodelson (2003)).

moving through space. We should obtain a space-time interval

$$ds^2 = dt^2 - a^2(t)dr^2 = 0. \tag{1.19}$$

This implies that

$$dr = \frac{dt}{a(t)}. \tag{1.20}$$

We then define the “conformal time" as

$$\eta \equiv \int_0^t \frac{dt'}{a(t')}. \tag{1.21}$$

In other words, the conformal time tells us about the comoving distance travelled by a photon between $t = 0$ and the time of interest.

1.2 Nucleosynthesis and Dark Matter Production

As mentioned earlier, we say that the inflationary epoch ended with reheating, where the inflation field(s) decayed into matter, antimatter and radiation. Something interesting happens immediately: the matter and antimatter produced during reheating begins to annihilate. The rate of these annihilations will decrease as the universe expands, eventually becoming insignificant. This is referred to as “freeze-out". Let’s now suppose that matter and antimatter were created in equal abundances in the early universe. In such a case, we'd find a post-freeze-out proton to photon ratio of

$$\eta = \frac{n_p}{n_\gamma} \sim 10^{-19}, \tag{1.22}$$

where n_p and n_γ are the number densities of protons and photons respectively. However, it is known from CMB observations that $\eta \sim 10^{-10}$. This difference can be explained by a small asymmetry in the production of baryons and anti-baryons. Thus, we have a simple observation with important consequences for fundamental physics.

As one could guess from the name, this is also the epoch in which the first nuclei begin to form. The lightest such nucleus to form (discounting hydrogen) is deuterium. Deuterium was formed in the reaction $n + p \leftrightarrow D + \gamma$. Starting from the Boltzmann equation and assuming kinetic equilibrium, one can show that ratio of number densities for deuterium and baryons goes like

$$\frac{n_D}{n_b} \sim \eta_b (\frac{T}{m_p})^{3/2} e^{B_D/T}. \tag{1.23}$$

In this formula, η_b is the ratio of baryons to photons and B_D is the binding energy of deuterium. Recall that observations show $\eta_b \sim 10^{-10}$, meaning that there will not be appreciable amounts of deuterium until after the temperature of the universe drops below the deuterium binding energy $B_D = 2.22$ MeV. Once the temperature drops below this value, one will find the light elements beginning to form. We show the resulting abundances in figure 1.3. This figure shows that helium-4 will be the most abundant element as the temperature drops below $T \sim 0.1$ MeV. In this figure, we see that the abundance drops significantly as temperatures drop below $T \sim 0.1$ MeV. Moreover, the fraction of helium-4 over total number of baryons rises to about 0.22. This has two important implications. For one, the significant helium fraction will be later used to generate carbon-12 in stars. Second, the helium-4 fraction 0.22 implies that there will still be significant amounts of free protons that can be used to form hydrogen once we reach recombination.

It should also be noted that this figure does not show abundances for any nuclei heavier than helium-4. This is due to the fact that the next energetically favorable isotope after helium-4 is carbon-12. This is illustrated in figure 1.4. Thus, in order to produce carbon-12, we would require 3 helium-4 nuclei. Such a reaction will not occur appreciably unless a high density of helium-4 is present. This will occur later, during the epoch of reionization when stars begin to form.

Recall that these light nuclei are produced at $T \sim 0.1$ MeV, which corresponds to about 100 seconds after the big bang. The next important step will be the formation of the first atoms, which will occur when temperatures drop below $T \sim 1/4$ eV. This event is referred to as “recombination".

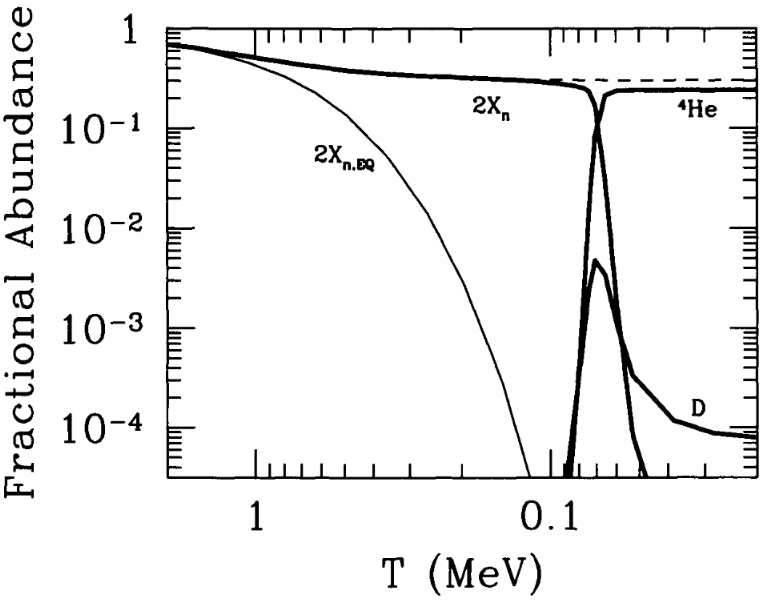

Figure 1.3: A plot of modeled abundances during the nucleosynthesis epoch. Once again, I have borrowed a plot from Dodelson (2003). In this plot, X_n refers to the ratio $n_n/(n_n + n_p)$. The dashed line provides fractional abundance assuming only weak interactions $p + e^- \leftrightarrow n + \nu_e$. The variable $X_{n,EQ}$ refers to the ratio assuming equilibrium is maintained throughout.

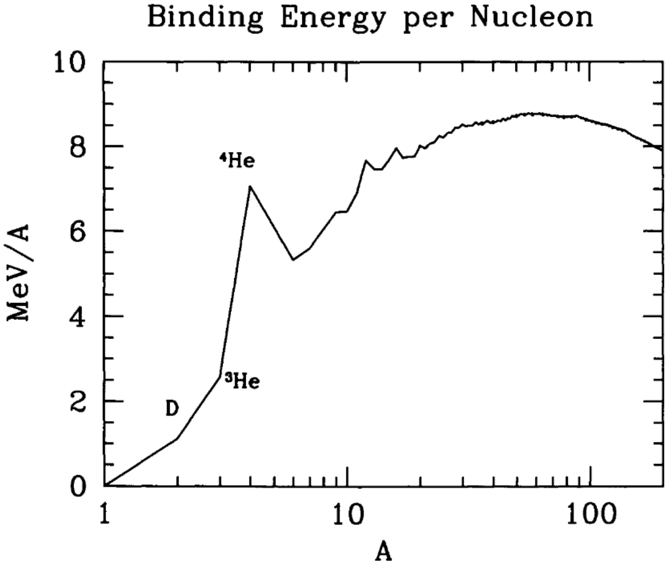

Figure 1.4: A plot showing binding energy per nucleon as a function of number of nucleons. This illustrates the isotopes between helium-4 and carbon-12 are not energetically favorable. Note that this figure was borrowed from Dodelson.

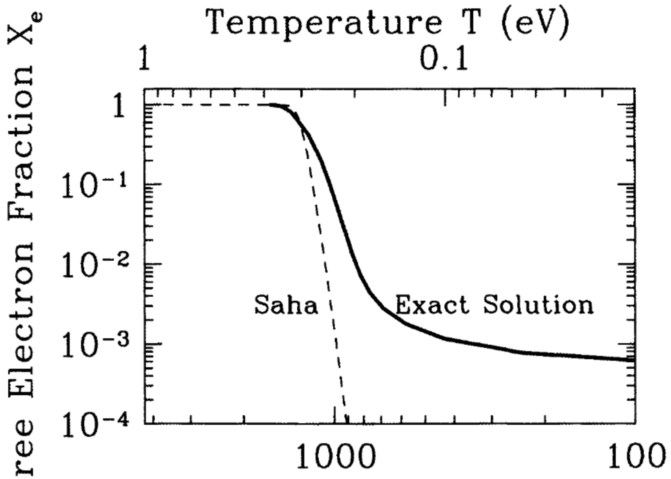

Figure 1.5: A figure from Dodelson (2003) illustrating hydrogen abundance. Note that the “Saha" curve refers to the predicted abundance assuming that the interaction rate between electrons and photons is high enough to maintain chemical equilibrium. The solid curve shows the numerical solution one would obtain without using the assumption of equilibrium.

1.3 Recombination

Although it is energetically favorable for electrons and nuclei to form atoms, we will not find appreciable abundances of any atoms until temperatures drop below $\sim 1\,eV$. This is due to the fact that at higher temperatures any atom that forms will be almost immediately be broken apart by Compton scattering. However, once the temperature has dropped appreciably below the hydrogen energy, one will find that recombination can occur! Once again, this reaction can be understood by using the Boltzmann equation

$$\frac{dX_e}{dt} = X_H \beta - X_e^2 n_b \alpha^{(2)}. \tag{1.24}$$

In this formula, the variable $X_e = \frac{n_e}{n_e + n_H}$ describes the abundance of free electrons. We also have $X_H = \frac{n_H}{n_e + n_H}$ which is the abundance of hydrogen. The first term here encodes the ionization of hydrogen while the second term encodes the production of hydrogen. The superscript above the α is meant to denote that we are considering the production of excited state hydrogen, since the production of ground state hydrogen will also result in ionizing photons.

It can be shown that the rate of Compton scattering in the universe is directly proportional to the electron fraction X_e. As a result, we find that Compton-scattering between electrons and photons drops sharply at the time of recombination. After the decoupling caused by recombination, the photons “free stream" through the universe, scattering minimally. Eventually, some of these photons will reach the instruments of cosmologists on earth.

1.4 The Cosmic Microwave Background

As mentioned in the previous section, photons were tightly coupled to electrons and baryons before recombination occurred. After recombination and decoupling, these photons proceeded to free stream through the universe. From this, one can expect that observation of these photons today would provide us with information about the universe at the time of recombination. The CMB is a rich topic with many subtleties and ties to cosmological parameters. As such, I summarize the most basic model for it and provide some explanation of the features that one sees in its power spectrum.

The CMB observed today is almost entirely isotropic, following a blackbody spectrum with temperature $2.7\,K$. However, there do exist small fluctuations in the CMB. These fluctuations are of order $10^{-5}\,K$ and carry with them much information about cosmological parameters. Since the CMB photon distribution features only small spatial perturbations, we can describe its statistics using the distribution function

$$f(\mathbf{x}, p, \hat{p}, \eta) = \left[exp(\frac{p}{T(t)\left[1 + \tilde{\Theta}(\mathbf{x}, \hat{p}, \eta)\right]} \right]^{-1}. \tag{1.25}$$

In this formula, $\tilde{\Theta}$ encodes the small spatial perturbations that one finds in the distribution of photons (aka the CMB). One can interpret $\tilde{\Theta}$ as telling us about fractional temperature perturbations $\frac{\delta T}{T}$ in the photon distribution. Note that we are ignoring polarizaton for now, although it will become important a bit later in this section. The photon perturbation $\tilde{\Theta}$ has no dependence on the magnitude of momentum p. This is due to the fact that the absolute value of momentum of a photon is usually affected minimally by Compton scattering. It is also helpful to Fourier transform $\tilde{\Theta}$, giving $\Theta(\mathbf{k}, \hat{p}, \eta)$.

It turns out that the $\hat{p}$ dependence of Θ is actually rather simple, depending only on $\mu \equiv \frac{\hat{p} \cdot \mathbf{k}}{k}$. So, we now are dealing with a function $\Theta(k, \mu, \eta)$. It will also prove useful to express Θ in terms of Legendre polynomial terms

$$\Theta_l(k, \eta) \equiv \frac{1}{(-i)^l} \int_1^1 \frac{d\mu}{2} \mathcal{P}_l(\mu) \Theta(k, \mu, \eta). \tag{1.26}$$

Now that we have these prerequisites out of the way, let's consider the evolution of photons before recombination. As always in cosmology, we track the evolution of structure using the Boltzmann equations. In this case, we will also be aided by making a few observations. During this era, photons are tightly coupled to electrons and protons. This coupling reduces the mean free path of photons and in turn washes out structures at small scales. At larger scales, this scattering will suppress Θ_l for $l > 1$. I'll now provide some intution for this. Consider a large-scale mode for which $1/k$ is much larger than the mean free path. In this case, an observer will only see photons arriving from distances much smaller than the wavelength of the mode. As such, they will see photons arriving from all directions with more or less equal temperature. This implies that little angular variation is seen and higher order Θ_l terms ought to be small. Armed with this knowledge, we end up with a pair of coupled equations for the photon distribution

$$\dot{\Theta}_0 + k\Theta_1 = -\dot{\Phi} \tag{1.27}$$

$$\dot{\Theta}_1 - \frac{k\Theta_0}{3} = \frac{k\Psi}{3} + \dot{\tau}\left[\Theta_1 - \frac{iv_b}{3}\right]. \tag{1.28}$$

In these equations, $\dot{\tau}$ is the scattering rate and v_b is the velocity of baryons. One can then massage the equation a bit, turning the coupled system into a single second order equation. Interestingly, this provides us with the equation of a damped and driven harmonic oscillator

$$\frac{d^2\Theta_0}{d\eta^2} + \frac{\dot{R}}{1+R}\frac{d\Theta_0}{d\eta} + k^2c_s^2\Theta_0 = F(k,\eta). \tag{1.29}$$

In this formula, R gives the baryon to photon ratio, c_s provides the speed of sound in the photon-baryon fluid and F is a forcing term that depends on Φ, Ψ and R. Although simplified, this equation does tell us something important. First, we see that perturbations in the distribution of photons should approximately behave as a damped, driven harmonic oscillator. Secondly, we see that these oscillations depend directly on the abundance of baryons along with the metric perturbations Φ and Ψ. In Figure 1.6, I show the estimated evolution of the monopole $\Theta_0 + \Psi$ as a function of conformal time η. Thus, before recombination, the k-modes of Θ_0 will oscillate over time. Moreover, these modes will not oscillate at the same rate nor be in phase with each other. Then, once recombination is reached, these oscillations cease due to decoupling of photons from matter, leaving the characteristic “acoustic peaks" seen in the CMB power spectrum. It should be noted that this oscillatory behavior is also responsible for the Baryon Acoustic Oscillations (BAO) measured by lower redshift Large Scale Structure (LSS) surveys.

Thus, the perturbations seen in the CMB today provide us with information about the state of the universe at the time of recombination. CMB surveys usually concern themselves with the “angular power spectrum" of the CMB, denoted C_l, which is defined via

$$\Theta(\mathbf{x}, \hat{p}, \eta) = \sum_{\ell=1} \sum_{m=-\ell}^{\ell} a_{\ell m}(\mathbf{x}, \eta) Y_{\ell m}(\hat{p}) \tag{1.30}$$

$$C_\ell \delta_{\ell\ell'} \delta_{mm'} = \langle a_{\ell m} a^*_{\ell' m'} \rangle. \tag{1.31}$$

Note in equation 1.30 that the photons observed by scientists on earth are located here on earth!, so $\mathbf{x}$ is denoting the location of the observers. On the other hand, $\hat{p}$ tells us the direction from which

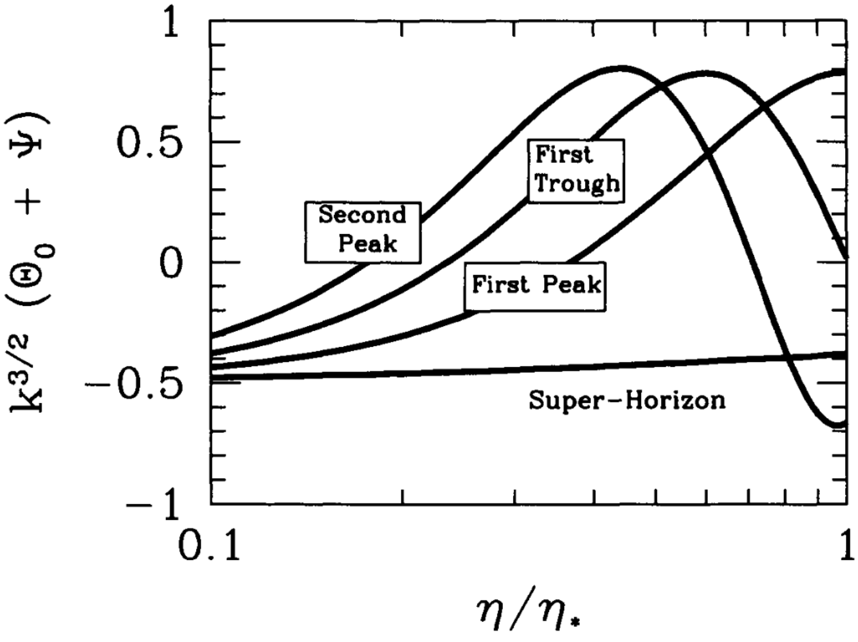

Figure 1.6: A plot of photon oscillations borrowed from Dodelson (2003). Note that $\eta_\star$ is the conformal time of recombination. Thus, the horizontal axis of the plot ends at recombination.

the photons arrive. It should be noted that C_l is really telling as about scales $\ell \approx k\eta_0$ at the time of recombination. In figure 1.7, I show a plot of $D_l = l(l+1)C_l$, borrowed from Planck Collaboration et al. (2020b)

Inflation

For a moment, let's assume the universe has been composed only of matter, radiation and perhaps dark energy for its entire history. Upon observing the CMB, we notice that it is very nearly homogeneous, with perturbations about a factor of 10^5 smaller than the average temperature. This would cause some confusion, as most of the observed sky was never in causal contact under our assumptions. This is illustrated in figure 1.8. This plot shows light cones of two points on the CMB last scattering surface. Since their light cones do not intersect at $\eta = 0$, we know that they should not have been in causal contact at last scattering. More generally, figure 1.8 shows that any points separated by comoving distance l_{co} greater than the conformal time at last scattering (η_{ls}) will not have been in causal contact at the time of last scattering. From our point of view on Earth, this corresponds to parts with angular separations about 1° or larger.

This may cause one to ask, “Why is the CMB so uniform if most regions were not in causal contact?" This question is referred to as the “horizon problem". Moreover, observations have shown that the universe is very close to flat. There are many values the curvature could have taken on,

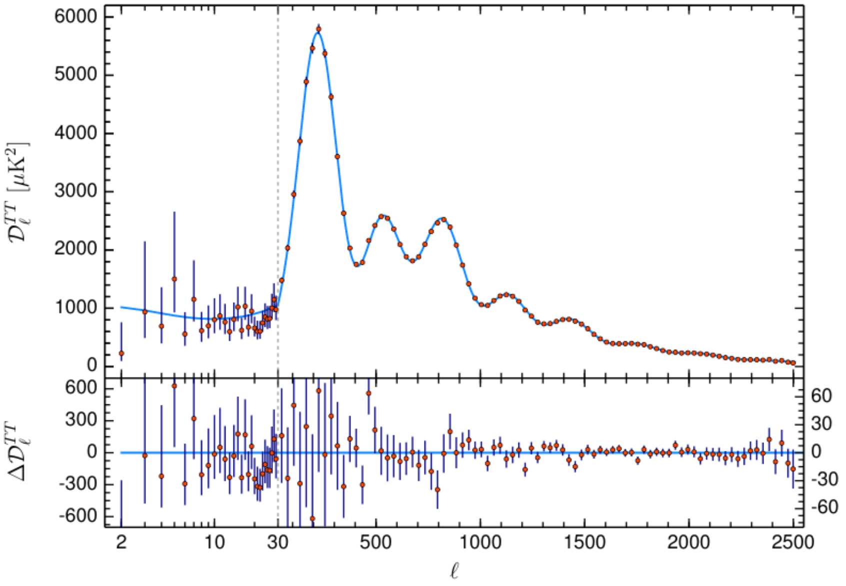

Figure 1.7: A plot from Planck Collaboration et al. (2020b) showing measured values of the angular power spectrum, along with the best fit ΛCDM model and residuals.

but for some reason it is very close to 0. This lead cosmologists to ask “Why is the universe so flat?" This question is referred to as the “flatness problem." One way cosmologists have attempted to resolve the horizon problem is by proposing the idea of “inflation".

I will now illustrate how inflation provides an answer to the horizon problem. Let $c = 1$ and consider two points separated by a comoving distance $r > (aH)^{-1}$. This implies that $aHr = \dot{a}r > 1$. Since $\dot{a}r$ gives the velocity of one object relative to the other, this tell us that the two points are receding away from each other at a speed greater than $c = 1$. Thus, points separated by a comoving distance greater than $(aH)^{-1}$ will not be in causal contact. The quantity $(aH)^{-1}$ is commonly referred to as the “Hubble radius".

Let’s continue considering the horizon problem. One solution to is to assume that the Hubble radius was extremely large at early times, ensuring that all points on the last scattering surface were in causal contact before recombination. Then, some mechanism causes the Hubble radius to shrink down to the value we observe at the time of recombination. One such mechanism for shrinking the Hubble radius is inflation. In the popular “single field slow roll" inflation scenario, the scale factor grows exponentially, which implies that the Hubble parameter becomes constant. Thus, we have

$$a(t) = a_e e^{H(t-t_e)}, \tag{1.32}$$

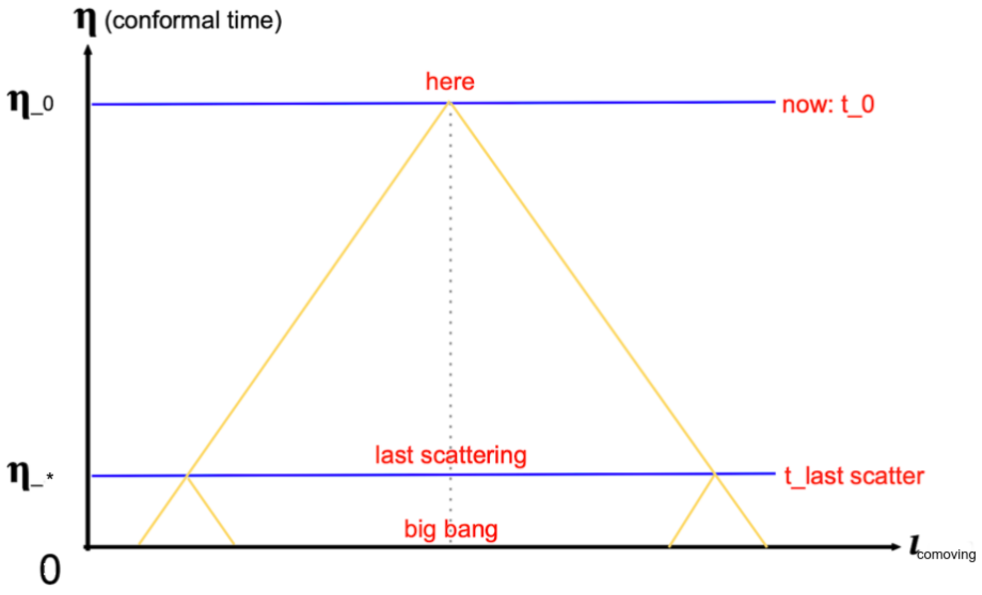

Figure 1.8: A figure illustrating the horizon problem, borrowed from wikipedia `https://en.wikipedia.org/wiki/Horizon_problem`.

where t_e is the time at the end of inflation. We also have

$$H = \frac{\dot{a}}{a} = constant. \tag{1.33}$$

According to the first Friedmann equation, this exponential growth in a corresponds to constant energy density. Since a is growing and H is constant, we find that inflation causes the Hubble radius $(aH)^{-1}$ to shrink over time. Assuming adequate shrinkage (60 e-fold according to Dodelson (2003)), one will find that the horizon issue is fixed. This is illustrated nicely in figure 1.9, which is borrowed from Dodelson (2003). In this plot, one finds that the Hubble radius at early times is large enough to ensure that the scales we observe in CMB and LSS surveys are susceptible to microphysical processes at early times. Inflation then shrinks the Hubble radius to a very small value. After inflation, we enter radiation domination, leading the Hubble radius to increase again, eventually reaching the value observed in CMB surveys. Thus, inflation has solved the horizon problem.

In general, inflation models require the Hubble radius to shrink over time. The most simple way to achieve this is to assume the presence of a single scalar field ϕ with potential $V(\phi)$ slowly rolling into a minimum. This is referred to as “single field slow roll" inflation. In this scenario, the scalar field provides essentially constant energy density as $V(\phi)$ varies little over time. The era of inflation ends when ϕ reaches a potential energy minimum and decays into matter, antimatter and radiation. This event is referred to as “reheating". In addition to being an explanation for what has already

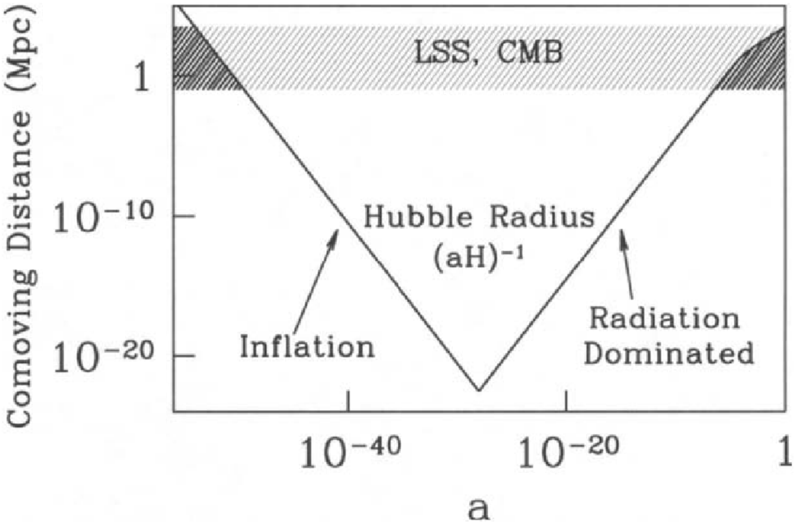

Figure 1.9: a plot showing how the horizon problem is fixed by inflation in a way that is consistent with CMB observations. The grey band in this plot corresponds the distance observed in CMB and LSS surveys (Dodelson (2003)).

been observed, inflation theories also provide predictions about the structure of the universe. I will now summarize some of these predictions for the single field slow roll hypothesis

1. Due to the uncertainty principle, the inflation field should have perturbations $\phi + \delta\phi(x)$. These perturbations provide the seeds for the structure we see in the universe today. We can characterize this structure according to the effect that it has on the metric. For the single field slow roll case, we find scalar curvature perturbations given by

$$P_{\mathcal{R}}(k) = 2\pi^2 \mathcal{A}_s k^{-3} (\frac{k}{k_p})^{n_s - 1}. \tag{1.34}$$

In this formula, k_p is a “pivot" scale chosen by the analyst, $\mathcal{A}_s$ sets the amplitude of perturbations and n_s is the “spectral index" for scalar perturbations. The model also predicts tensor perturbations

$$P_T(k) = 2\pi^2 \mathcal{A}_T k^{-3} (\frac{k}{k_p})^{n_T}. \tag{1.35}$$

In this formula, n_T is the spectral index for tensor perturbations. This power spectrum given in the first equation provides the seeds for the structure we see in the universe today. The power spectrum in the second equation sources primordial gravitational waves. Such gravtiational waves could in principle be detected in the CMB polarization, but as of now no detection has been made. Such waves could also leave an imprint in the matter power spectrum at very small scales, which could in principle be detected by observing the 21 *cm* line of hydrogen during the dark ages.

2. Some models, including single field slow roll, predict that there should be some degree of non-Gaussianity. Such non-Gaussianity may be detectable in LSS probes such HI intensity mapping. As of now, this non-Gaussianity has never been detected.

Thus, one finds that the inflation model solves the horizon problem and makes testable predictions about the matter power spectrum along with the presence of primordial gravitational waves.

1.5 Growth of Matter Perturbations

One can recast 1.34 to show that inflation establishes initial perturbations in the metric (via the function Φ). These metric perturbations $\Phi(\mathbf{k},\eta)$ will in turn source perturbations in the dark matter overdensity $\delta(\mathbf{k},\eta)$. Once these matter perturbations are established, they should evolve in a way that schematically looks like

$$\ddot{\delta} = \delta\left[Gravity - Pressure\right]. \tag{1.36}$$

This schematic is showing what one would likely expect, which is that gravity causes perturbations to grow while pressure tries to pull them apart. The idea here is that in regimes with low pressure, one will find perturbations growing. In regimes with significant pressure, one will obtain oscillatory behavior. We present modeled values for the power spectrum at $z = 0$ in Figure 1.10. Note that the curve labeled ΛCDM corresponds to the standard cosmological theory. One will note here that the power spectrum curve increases linearly as a function of k until $k \approx 0.15hMpc^{-1}$ at which time it begins to shrink. We'll now provide some explanation for why one sees such features appearing in the power spectrum.

In Dodelson (2003), he models the evolution of Φ over time, then relates those results into statements about the matter distribution. In particular, we relate Φ at the end of inflation to Φ at late times using

$$\Phi(\mathbf{k},a) = \frac{9}{10}\Phi_p(\mathbf{k})T(k)\frac{D_1(a)}{a}. \tag{1.37}$$

To be more precise, a "late time" is a time for which $a >> a_{eq}$, with a_{eq} being the scale factor at matter-radiation equality.

In this formula, Φ_p describes primordial metric perturbations. $T(k)$ is the "transfer function", which is defined as

$$T(k) \equiv \frac{\Phi(k,a_{late})}{\Phi(k_{large-scale},a_{late})}. \tag{1.38}$$

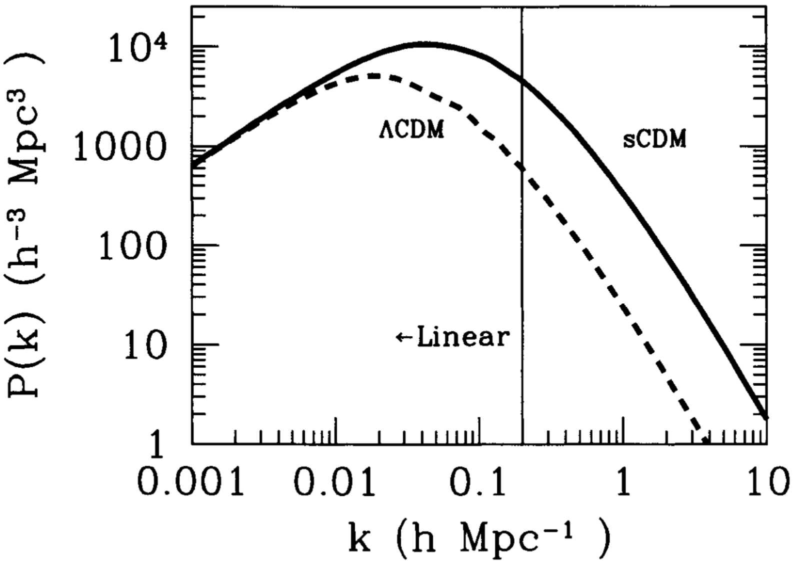

Figure 1.10: Simulated matter power spectra at present day, borrowed from Dodelson (2003). This figure does not account for non-linear effects. The "sCDM" curve corresponds to a cosmology that does not have a cosmological constant.

This function encodes the k dependence of the evolution of Φ and is equal to 1 at large scales. D_1 is the "growth factor" that describes the growth of modes due to gravitational effects. In particular, it is defined as

$$\frac{D_1(a)}{a} \equiv \frac{\Phi(a_{late})}{\Phi(a)}. \tag{1.39}$$

In this way, we can relate the physics of the early universe (right at the end of inflation) to those of the late universe.

At late times, once the mode is well within the horizon, we can relate the potential Φ to δ using Poisson's equation. After some massaging, we obtain

$$P(k,a) = 2\pi^2\delta_H^2\frac{k^n}{H_0^{n+3}}T^2(k)(\frac{D_1(a)}{a=1})^2. \tag{1.40}$$

So, tracking the evolution of Φ gives us a way to relate the power spectrum at later times to the primordial perturbations established by inflation. This equation also makes clear why $D_1(a)$ is referred to as the "growth function". It is describing the growth of matter perturbations with respect to time.

We can model the evolution of these modes to linear order using the Boltzmann equations along with a formula relating Φ to the radiation and matter overdensities. In particular, the Boltzmann equations are

$$\dot{\Theta}_{r,0} + k\Theta_{r,1} = -\dot{\Phi} \tag{1.41}$$

$$\dot{\Theta}_{r,1} - \frac{k}{3}\Theta_{r,0} = -\frac{k}{3}\Phi \tag{1.42}$$

$$\dot{\delta} + ikv = -3\dot{\Phi} \tag{1.43}$$

$$\dot{v} + \frac{\dot{a}}{a}v = ik\Phi \tag{1.44}$$

Note that the r subscript is used to denote radiation perturbations. Moreover, the variable v refers to the bulk velocity of dark matter. We then complete the set of equations by relating Φ to the overdensities. One has many options for this. Two such options are

$$k^2\Phi + 3\frac{\dot{a}}{a}(\dot{\Phi} + \frac{\dot{a}}{a}\Phi) = 4\pi G a^2 \left[\rho_{dm}\delta + 4\rho_r\Theta_{r,0}\right] \tag{1.45}$$

and

$$k^2\Phi = 4\pi G a^2 \left[\rho_{dm}\delta + 4\rho_r\Theta_{r,0} + \frac{3aH}{k}(i\rho_{dm}v + 4\rho_r\Theta_{R,1})\right]. \tag{1.46}$$

These equations can both be derived from Einstein's equations. This system of equations will give us good understanding of the features seen in Fig. 1.10. However, it should be understood that this system of equations includes perturbations only to linear order. Over time, as structure grows, one will find that this linear treatment begins to break down, no longer accurately describing the distribution of matter. Such non-linear modes are difficult to analyze, being corrupted by more complicated astrophysics. For this reason, higher redshift surveys provide us with more easily interpreted information about smaller scales. This will be discussed a bit more later in this section.

It should also be noted that these equations are neglecting dark energy. As such, any results derived from them will be only approximately true at $a = 1$. If accounted for, one would find that the presence of significant dark energy at low redshifts will suppress the growth of structure.

However, we can still understand the general features in the power spectrum using this set of equations. Let's now begin understanding the features.

Why does $P(k)$ increase linearly for low k?

Although we are interested in δ, we will instead solve for the potential Φ, then relate this result back to δ later on. Consider a k mode with length scale large enough that it enters the horizon after the the epoch of matter-radiation equality (a_{eq}). While the mode $\delta(k)$ is larger than the horizon, it will still undergo some evolution due to its interaction with the metric. Under these assumptions, we end up with following system of equations relating matter and radiation to the metric.

$$\dot{\Theta}_{r,0} = -\dot{\Phi} \tag{1.47}$$

$$\dot{\delta} = -3\dot{\Phi} \tag{1.48}$$

$$3\frac{\dot{a}}{a}(\dot{\Phi} + \frac{\dot{a}}{a}\Phi) = 4\pi G a^2 \left[\rho_{dm}\delta + 4\rho_r\Theta_{r,0}\right] \tag{1.49}$$

The first two equations show that $\delta = 3\Theta_{r,0}$ at large scales. Note that there is no constant offset between the two due to the initial conditions provided by inflation. This result can be be plugged into the third equation, and after much massaging, one obtains the result

$$\Phi = \frac{\Phi(0)}{10}\frac{1}{y^3}\left[16\sqrt{1+y} + 9y^3 + 2y^2 - 8y - 16\right]. \tag{1.50}$$

In this formula, $y = \frac{a}{a_{eq}} = \frac{\rho_{dm}}{\rho_r}$. At later times (large y), we will find that

$$\Phi = \frac{9}{10}\Phi(0). \tag{1.51}$$

Thus, large scale modes will go through a 10 percent drop as they pass through matter-radiation equality. Moreover, this effect is independent of k for any such large scale mode.

After passing through the horizon (at times later than $a >> a_{eq}$), one will find that Φ will become close to constant with respect to time. Although we won't derive it here, the fact that Φ stays constant during matter domination can be explained using our system of equations, excluding any contributions from radiation perturbations.

By relating this result to equation 1.37, we find that the transfer function $T(k)$ does not vary with k at large scales. If one assumes single field slow roll inflation, the value of n in equation 1.40 should equal 1, providing the linear increase at large scales seen in figure 1.10. Thus, the linear increase of $P(k)$ with respect to k at large scales is explained by fact that each large scale metric perturbation evolves identically with respect to time. This then provides the linear behavior seen in $P(k)$ via Poisson's equation.

Why does $P(k)$ decrease for high k?

Let's consider a mode that enters the horizon before a_{eq}. As the horizon crossing occurs before a_{eq}, we can neglect any influence from the dark matter overdensity δ on the evolution of Φ at this time. Once we understand Φ, we can estimate δ by taking Φ as an external driving force. As explained earlier, we know that only the monopole and dipole moments of the radiation perturbations should be significant. With these assumptions, we obtain equations

$$\Phi = \frac{6a^2H^2}{k^2}\left[\Theta_{r,0} + \frac{3aH}{k}\Theta_{r,1}\right] \tag{1.52}$$

and

$$-\frac{3}{k\eta}\dot{\Theta}_{r,1}\left[1 + \frac{3}{k^2\eta^2}\right] = -\dot{\Phi}\left[1 + \frac{k^2\eta^2}{6}\right] - \Phi\frac{k^2\eta}{3}. \tag{1.53}$$

With some massaging, one can obtain a spherical Bessel function solution

$$\Phi = 3\Phi_p j_1(k\eta/\sqrt{3}). \tag{1.54}$$

So, a mode that enters the horizon during radiation domination will evolve like j_1. This result can qualitatively be explained as follows. Since there is significant radiation pressure during radiation domination, one will find oscillatory behavior in the radiation perturbations. Since perturbations are not growing, one will find Φ shrinking due to dilution caused by the expansion of the universe. This decaying oscillatory behavior is described using the j_1 function. The fact that the potential shrinks in this way implies that the transfer function $T(k)$ ought to be significantly less than 1 at small scales. Moreover, as k increases, one can expect even more severe decay in Φ. This is because such modes will both enter the horizon sooner (allowing more time to decay before a_{eq}) and also because the decay will happen more quickly due to the higher value of k.

In Dodelson (2003) , the author goes through a lengthy derivation relating the decaying Φ solution to the evolution of δ with respect to time for a small scale mode, eventually deriving a transfer function. Although this derivation is interesting, I'll skip over it and show the resulting transfer function. We have

$$T(k) = \frac{12k_{eq}^2}{k^2} ln\left[\frac{k}{8k_{eq}}\right] \tag{1.55}$$

Note that this logarithmic scaling is due to the fact that δ grows logarithmically during radiation domination, somewhat suppressing the decrease. Relating this back to equation 1.40, and assuming once again that $n = 1$, we should find that the power spectrum decreases as k increases. Speaking roughly, we can say that the small scale modes of the power spectrum are suppressed due to the fact that they enter the horizon at a time when radiation pressure suppresses the growth of structure.

Non-Linear Physics

Note that equations 1.41 through 1.46 shown earlier in this section are only correct to *linear* order in the perturbations. However, as perturbations grow one will find that these linear approximations no longer apply. To understand the transition into non-linearity, one can imagine convolving the matter

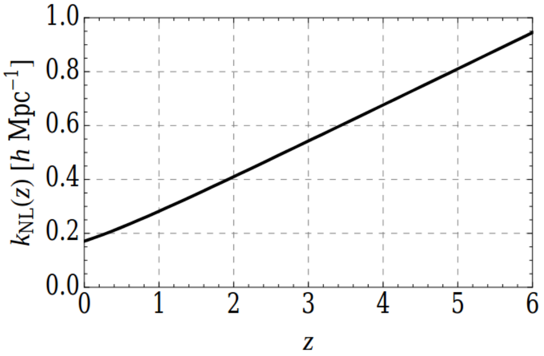

Figure 1.11: The non-linear scale k_{nl} as a function of redshift. This plot was borrowed from Ansari et al. (2019).

over-density field $\delta(\mathbf{x})$ with a window function of finite width $W(\mathbf{x})$. Let's use $\delta_W(\mathbf{x})$ to denote the convolved overdensity. One can relate δ_W to the linear power spectrum via

$$\langle \delta_W(\mathbf{X})^2 \rangle = \frac{1}{2\pi^2} \int dln(k) k^3 P_L(k) |W(k)|^2. \tag{1.56}$$

As one makes $W(\mathbf{x})$ narrower, one will find $W(k)$ becoming broader, increasing the size of perturbations. Thus, if one looks at a sufficiently small scale, they will find large perturbations beginning to appear. As an approximation, one can take k_{nl} to be the smallest value of k for which

$$\Delta_L(k_{nl}) = 1. \tag{1.57}$$

Note here that $\Delta(k) = \frac{k^3 P(k)}{2\pi^2}$ is the “dimensionless power spectrum" which gives the magnitude of perturbations in a logarithmic bin of k-space. I have used Δ_L to denote the dimensionless spectrum predicted by linear theory. So, we are saying that the non-linear regime begins when dimensionless perturbations exceed 1. In figure 1.11, I show the estimated value of k_{nl} for redshifts 0 to 6. This plot shows k_{nl} increasing as a function of z. This result is expected since higher z corresponds to earlier times and thus smaller perturbations.

Once a scale becomes non-linear, it can't be accurately understood using the results described in this section. One must instead use a perturbative approach or N-body simulations. As such, non-linear modes become significantly more difficult to analyze and relate to the ΛCDM model. As we will see later, this motivates the use of observations at higher redshifts, as the earlier universe will have more linear modes available. While we're on the topic of higher redshifts, let's move on to the dark ages.

1.6 Dark Ages

So far we have discussed the universe up until recombination, along with physical processes that have determined the structure within it. Next, we enter the dark ages, which last from recombination until the beginning of reionization . During this epoch, decoupling has occurred and photons free stream as they like. Meanwhile, objects such as stars or galaxies have not yet formed and only light atoms exist in significant abundances.

This era of the universe may seem boring at first glance, but its simplicity is what makes it interesting. Since we are still at early times, perturbations have not had much opportunity to grow and we find that k_{nl} should be quite large, corresponding to very small scales. As such, small scales during this era can be related to linear theory, making analysis simple. This will be discussed in more depth later.

1.7 Epoch of Reionization

The dark ages end when density perturbations become large enough to allow for the formation of more complex objects like stars and galaxies. This occurs around redshift $z \sim 20$, or hundreds of millions of years after the big bang. As mentioned earlier, the universe up until this point has had negligible concentrations of atoms heavier than Helium-4. Now, however, we will find heavier elements being produced in significant concentrations in the newly formed stars. Of course, the fusion occurring in these stars raises the temperature, leading to the emission of ionizing photons. This radiation in turn ionizes the neutral gas of the universe, hence the term “reionization". This process begins with the formation of “ionization bubbles" around the newly formed stars and galaxies. In Fig. 1.3, I present a figure illustrating this process. This figure shows small bubbles of ionized gas around the newly formed objects. These regions then grow over time as the photons propagate outwards.

In figure 1.13, I show another viewpoint of the reionization process. This figure shows the brightness temperature of 21 *cm* radiation as a function of redshift. This 21 *cm* radiation is emitted by neutral hydrogen (HI). In this figure, we can see the reionization process taking place as the HI is ionized, suppressing the emission of the 21 *cm* line for decreasing redshifts.

I’d like to caution readers not to interpret figure 1.13 as saying the 21 *cm* is completely eliminated during reionization. Within galaxies, there will still exist relatively dense clouds of HI that self shield from the ionizing radiation. This HI will be still be observable after the reionization process is complete.

After some time, we will find that the reionization process has run its course, and the diffuse HI of the universe has been throughly ionized. This happens around $z \sim 6$, or about 1 billion years after

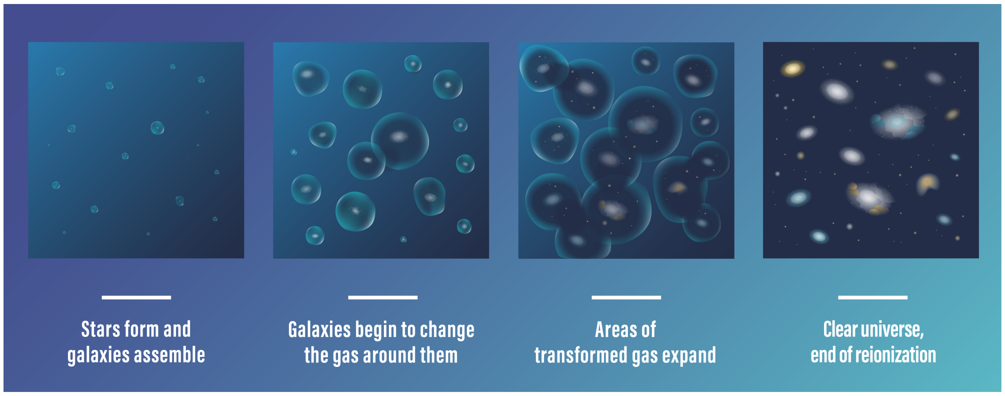

Figure 1.12: A graphic illustrating the process of reionization. This image is borrowed from `https://webbtelescope.org`.

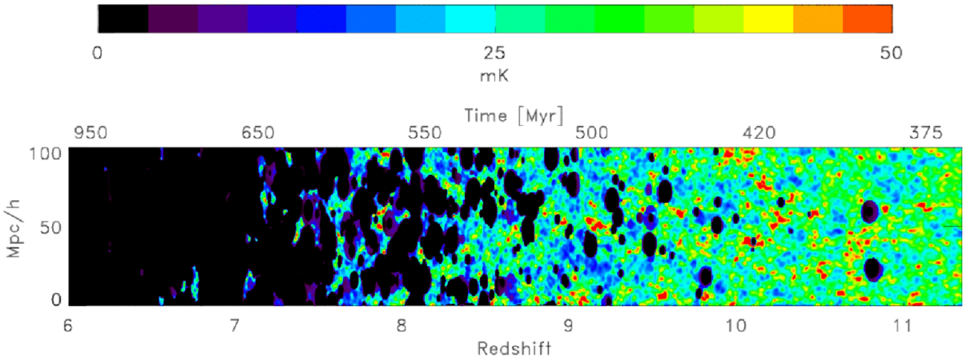

Figure 1.13: Simulated 21 *cm* emission during reionization. This figure is borrowed from (Thomas et al. (2009)).

the big bang.

1.8 Post-EOR Universe

Once reionization ends, the newly minted objects will continue evolving and we will find gravity continuing to cause the growth of the matter overdensity $\delta(k)$. Moreover, we will find that the expansion of the universe will begin to accelerate, at least at low redshifts. By this, I mean in particular that $\ddot{a} > 0$. This accelerating expansion is the aspect of the post-EOR most relevant for this thesis, so I will focus on that aspect.

One can find how to achieve accelerating expansion by checking the second Friedmann equation. This equation states that

$$\frac{\ddot{a}}{a} \propto -(\rho + 3P). \tag{1.58}$$

where P is the pressure and ρ is the energy density. Thus, one achieves accelerating expansion when

$$P < \frac{1}{3}\rho. \tag{1.59}$$

This is expansion is often times described in terms of the “equation of state"

$$w = \frac{P}{\rho}. \tag{1.60}$$

For substances in the universe, we find that the density will be governed by the fluid equation

$$\frac{\partial \rho}{\partial t} + \frac{\dot{a}}{a}\left[3\rho + 3P\right] = 0. \tag{1.61}$$

One can massage this equation and integrate to obtain

$$\rho \propto exp\left[-3\int^{a} \frac{da'}{a'}\left[1 + w(a')\right]\right]. \tag{1.62}$$

It’s possible that the dark energy density is constant with time, only becoming appreciable at recent times due to the dilution of matter and radiation. In this case, one would find that $w(a) = -1$. Having a constant energy density is consistent with current observations. For instance, **?** shows w being close to -1 assuming no redshift dependence in the dark energy equation of state. Alternatively, one could also find that $w(a)$ varies with time and perhaps only became appreciably negative in recent times. Such scenarios have not yet been ruled out by observations, which are scant for $z > 2$.

1.9 What Are Questions of Interest in Cosmology?

Now that I’ve supplied a very brief summary of the universe’s timeline, I’d like to emphasize that the story here is of course not complete. We have much left to learn! Here’s a brief list with questions that are of interest in the field.

1. What’s the nature of dark energy?
2. What’s causing the Hubble tension?
3. What’s the nature of inflation?

I’ll now go through each of these in a bit more detail, describing the question and providing some observational evidence related to each.

What's the nature of dark energy?

As mentioned earlier, cosmological observations show accelerating expansion. As of now, it is unclear what may cause such expansion. For instance, one could suppose that vacuum energy may provide a uniform density that leads to the accelerating expansion seen. However, theoretical models result in an energy density much larger than the observed density of dark energy. There also exist theories involving scalar fields, which are described as “quintessence models". Such models predict that there may be some redshift evolution present in the dark energy equation of state.

So far, information about the accelerating expansion comes from type Ia supernovae, galaxy surveys and CMB observations. First, these observations show that currently only about 30 percent of the universe's energy is in the form of matter, with negligible contributions from radiation or curvature. As such, dark energy likely constitutes about 70 percent of the universe's current energy density. However, the story does not end here; we may also try to see if dark energy evolves over time.

Type Ia supernovae are useful in this context since they serve as “standard candles", which is to say they have a well understood luminosity. One can then estimate the “luminosity distance" of a supernova as

$$d_L \equiv (\frac{L}{4\pi f})^{1/2} \propto \int \frac{dz}{H(z)}. \tag{1.63}$$

In this formula, L is the luminosity and f is the measured flux. Thus, measuring type Ia supernovae as a function of redshift gives us some handle on the evolution of $H(z)$ over time, which in turn provides information on the acceleration. These supernovae provide us with information about $H(z)$ out to a redshift of $z \sim 1$.

The Scolnic analysis also incorporated results from Baryon Acoustic Oscillations (BAO) observed in galaxy surveys. BAO wiggles are a consequence of the oscillations discussed in the CMB section. Roughly speaking, the BAO provides a boost in the matter distribution at separations with comoving distance $r_s \approx 100Mpc/h$, where r_s is the “comoving sound horizon". Since the BAO is a feature with a well known length scale, we refer to it as a “standard ruler". One way to derive information about $H(z)$ from these wiggles is to measure modes perpendicular to the line of sight. In this case, one can measure an “angular diameter distance"

$$d_A = \frac{r_s}{\theta} = \frac{r(z)}{1+z} = \frac{1}{1+z}\int \frac{dz}{H(z)}. \tag{1.64}$$

In this formula, $r(z)$ is the comoving distance to redshift z. Thus, we use the fact that the length-scale of the BAO is well understood to extract information about how the expansion rate of the universe evolves. It should be noted that a similar analysis can be done using BAO wiggles along the line of sight. As of now, measurement of the BAO via galaxy surveys provides us with information about

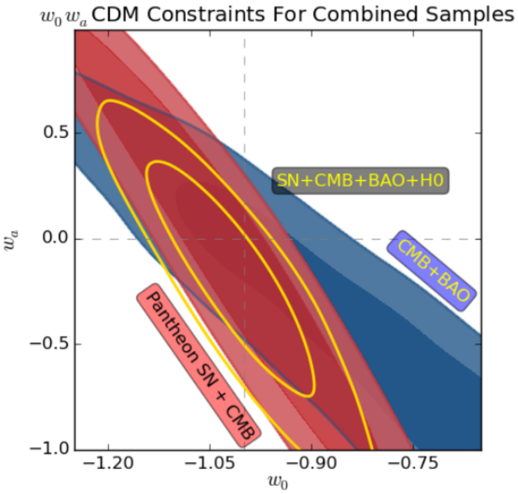

Figure 1.14: Limits from Scolnic et al. (2018), both at 68 and 95 percent confidence. The yellow curves combine all available information.

$H(z)$ out to $z \sim 2.5$.

As already mentioned, acoustic oscillations also appear in the CMB, providing us with another standard ruler.

The analysis of Scolnic et al. (2018), assumes a dark energy equation of state with the form

$$w(z) = w_0 + w_a \frac{z}{1+z}. \tag{1.65}$$

This gives a simple model, and allows us to estimate whether or not the equation of state has any redshift evolution. In figure, 1.14, I show the limits set in Scolnic et al. (2018), which combined data from supernovae, galaxy and CMB observations.

In other words, the authors of the Scolnic paper estimate $w_0 = -1.007 \pm 0.089$ and $w_a = -0.222 \pm 0.407$. Thus, the observations of Scolnic et al. (2018) are consistent with dark energy being a cosmological constant with equation of state $w = -1$.

More recently, the Dark Energy Spectroscopic Instrument (DESI) collaboration has published results on this subject as well (Adame et al. (2024)). In this paper, the authors analyzed a combination of data from DESI and other probes, finding a preference for $w_0 > -1$ and $w_a < 0$. In other words, their analysis seems to be preferring dynamic dark energy over a cosmological constant. Depending on the data sets chosen, they find discrepancies with ΛCDM ranging from 2.5 to 3.9σ. In Fig. 1.15, I present a plot of dynamical dark energy estimates from Adame et al. (2024).

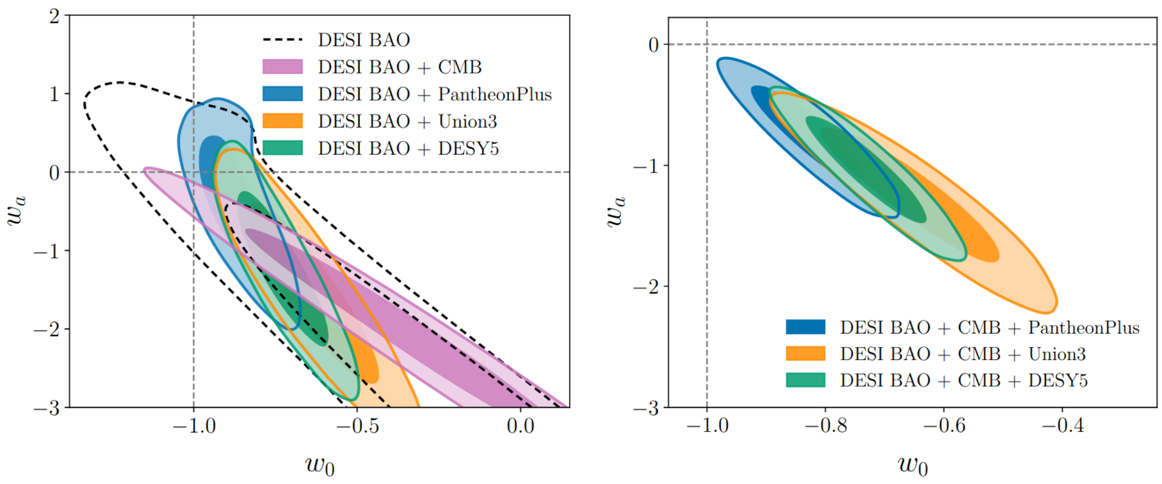

Figure 1.15: Constraints set on w_a and w_0 given at the 65% and 95% levels. These constraints were set by combining the DESI BAO data with data from other probes.

Hubble Tension

The observations in Scolnic also lead to an interesting discrepancy in H_0, the value of the Hubble constant at present day. If one estimates H_0 using standard candles at low redshift, they will obtain estimates for H_0 around 73 $kms^{-1}Mpc^{-1}$. For instance, Riess et al. (2022) estimates $H_0 = 73.04 \pm 1.04\, kms^{-1}Mpc^{-1}$ using cepheids and type Ia supernovae. However, if one estimates H_0 using CMB observations, they will obtain estimates closer to 68 $kms^{-1}Mpc^{-1}$. For instance, the Planck 2018 results provide an estimate $67.4 \pm 0.5 kms^{-1}Mpc^{-1}$. This discrepancy is consistently seen across all surveys.

At this point, the discrepancy has reached a 4 to 6σ significance (Hu and Wang (2023)). Data from both CMB and local distance ladder measurements have been reanalyzed, providing still discrepant results. However, it is still possible that the Hubble tension is due to some sort of incorrect assumption being made in the analyses. Other proposals to explain the Hubble tension involve new physics. As one would expect, many hypotheses have been made to explain why this tension has appeared. Such hypotheses can be nicely classified into a list (Hu and Wang (2023))

1. Late time deformations of the Hubble expansion rate $H(z)$.
2. Deformations of the Hubble expansion rate H(z) with additional interactions/degrees of freedom.
3. Deformations of the Hubble expansion rate H(z) with inhomogeneous or anisotropic modifications.

4. Late time modifications: Transition or recalibration of the SNe Ia absolute luminosity.
5. Early time modifications of sound horizon.

As one can tell from this list, there are a lot of possible explanations for the Hubble tension. It seems to me that this diverse set of theories highlights the fact that we don't really know what's going on! This indicates that more observations will need to be made to resolve this issue.

What is the nature of inflation?

As mentioned earlier, inflation is a popular theory for the state of the universe at early times. It is quite well motivated due to the fact that it answers both the horizon problem and also the "flatness problem". The flatness problem is the difficulty of explaining why the curvature of the universe is so close to 0, which without a mechanism such as inflation, would seem to require some fine tuning.

There also exist a large number of hypotheses for what may have caused inflation. For example, Martin et al. (2014) lists 139 different single field slow roll inflation models! There of course also exist inflation models beyond single-field slow roll, although I won't discuss them in this thesis.

In any case, the fact that inflation provides solutions to the flatness problem and the horizon problem makes it a well motivated idea. However, we do not as of now know which inflation theory (if any) is correct. In the next chapter, I will review how CMB polarization and LSS can be used to test models of inflation.

2 OBSERVATIONAL METHODS FOR LARGE SCALE STRUCTURE

Now that I've provided a brief summary of cosmology, it will be beneficial to introduce observational methods used to learn about the universe. In the previous chapter, I did mention the use of supernovae and other standard candles for estimating the Hubble constant. However, for simplicity, I'll now narrow my focus to probes of Large Scale Structure (LSS), which are most relevant to the work that I have done. These LSS probes include CMB surveys, galaxy surveys and intensity mapping (IM). I'll begin this chapter by providing brief descriptions of CMB experiments and galaxy redshift surveys, along with some of the cosmology we have learned from these methods. Then, I will provide an in depth description of intensity mapping (IM).

2.1 Traditional Probes of LSS

Currently, a lot of what we know about cosmology comes from observations of the CMB and of the distribution of galaxies and quasars at lower redshifts. These techniques are highly developed and have been producing interesting results for decades. Although I don't work on these sorts of experiments, I think it would be beneficial to discuss them for the sake of getting a more complete picture of observational cosmology.

CMB Measurements

I will begin the CMB discussion by providing a brief description of how CMB observations are often made. I'll focus mostly on the Planck satellite and the BICEP array, located in Antarctica.

How are CMB Measurements Conducted?

The discovery of the CMB took place in 1964 (Penzias and Wilson (1965)). In this experiment the authors saw what appeared to be isotropic and unpolarized radiation with a temperature of $3.5 \pm 1\,K$. The instrument used by Penzias and Wilson was relatively simple. They had a horn antenna looking at the sky with a beam-width of about 1.5°. This antenna was then connected to a receiver measuring the power coming out of the antenna. This power was then used to estimate the temperature of the CMB.

As mentioned earlier, the CMB is very close to isotropic, with anistropies of order $10^{-5}\,K$. As such, an instrument more sensitive than that of Penzias and Wilson must be used to measure the anisotropies.

As one can imagine, CMB experiments have developed a lot over the years. First, while Penzias and Wilson performed a ground-based experiment, modern CMB anisotropy experiments are sometimes placed on satellites orbiting the earth. Placing the instrument on a satellite is done to avoid troublesome atmospheric effects. Examples of satellite-based CMB experiments include Planck, LiteBIRD and the proposed PICO experiment. There are still a significant number of modern ground-based CMB experiments, including the Simons Observatory (Ade et al. (2019)) and the Atacama Cosmology Telescope (ACT) (Choi et al. (2020)). A highly notable CMB instrument is the Planck satellite (Aghanim et al. (2020)). In figure 2.1, I show a picture of the focal plane of the High Frequency Instrument (HFI) used on the Planck satellite. One can tell already from this figure that the complexity of CMB experiments had increased significantly since the early days. While Penzias and Wilson used a single horn antenna at a single frequency, the HFI of Planck has a large array of horns covering frequencies from 100 to 857 GHz. These modern instruments also measure the power output from their antennas differently than Penzias and Wilson. While Penzias and Wilson used “coherent" detectors, current and future CMB telescopes use “incoherent" detectors such as “bolometers" and Kinetic Inductance Detectors (KIDs), at least when operating at frequencies greater than 100 GHz. Bolometers measure the intensity of incoming radiation by measuring the resistance of a thermistor in thermal contact with an object that absorbs CMB photons. In figure 2.2, I show a picture of a bolometers used in Planck. This bolometer on the left is referred to as a “spider web bolometer" and is designed to measure only the unpolarized signal. On the right, the bolometer has a rectangular grid, which is metallized in only one direction. This directionality provides such a bolometer with sensitivity to the polarization of the CMB. As such, the Planck satellite has provided us with excellent data about the temperature and polarization of the CMB.

One typically sees the polarization described in terms of “E-modes" and “B-modes". I will skip the details, but E-mode polarization in the CMB can be caused by both tensor and scalar perturbations in the metric. On the other hand, primordial B-modes in the CMB can only be sourced by tensor perturbations. Such tensor perturbations are predicted by inflationary models. E-modes have been detected by Planck and the information gained from them has been used to reduce uncertainties in cosmological parameters. On the other hand, primordial B-modes sourced by tensor modes, if they exist, would be quite small. As of now, CMB B-modes sourced by tensor perturbations have never been observed by Planck or any other CMB probe. It should be noted that experiments such as Planck do observe non-zero B-mode contributions. However, the B-modes that have been observed were not a feature of the CMB itself but rather an artifact of gravitational lensing, caused by structure along the line of sight out to the last scattering surface of the CMB.

It should be noted that there do exist instruments specifically seeking to measure primordial B-modes in the CMB. The most notable among these is the series of experiments performed by the BICEP/KECK collaboration (Yoon et al. (2006)) (Ade et al. (2014)). B-mode instruments such as the series of BICEP/KECK telescopes are ground-based and observe small patches of the sky for long amounts of time. This is different than telescopes such as Planck that seek to measure the

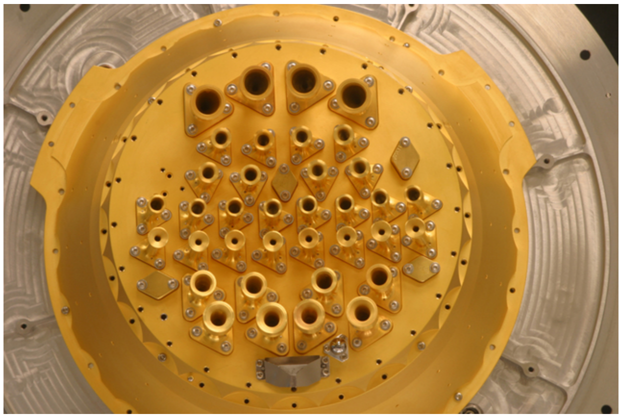

Figure 2.1: A picture of the focal plane of the Planck HFI. Borrowed from `https://planck.ipac.caltech.edu/page/high_frequency_instrument`.

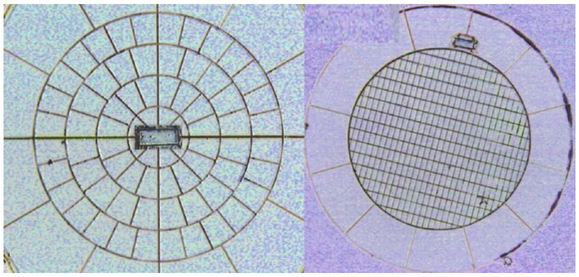

Figure 2.2: A picture of bolometers used on the Planck satellite. The bolometer on the left is used to measure the CMB temperature while the one on the right is sensitive to the polarization of the CMB. This image is borrowed from Lamarre et al. (2010).

whole sky. The idea is to search for primordial B-modes by maximizing sensitivity within some small patch of the sky.

As mentioned earlier, primordial B-modes have never been observed in the CMB. Rather, experiments such as KECK have only set upper limits on the level of such polarization.

What can be learned from such CMB measurements?

Measurements of the CMB temperature anisotropies and E-mode polarization have been an extremely useful tool for testing and fleshing out ΛCDM cosmology. I'll now show a list giving some examples of what we can learn from the CMB. These examples are all provided in Planck Collaboration et al. (2020b).

- Observations are consistent with slow roll inflation with scalar spectral index close to 1.
- Constrains curvature to be extremely small.

- Consistent with BBN given ΛCDM
- Constrains the sum of neutrino masses.
- Helps constrain $w_0 \approx -1$
- Probes for sterile neutrinos and dark matter annihilation.
- Provides constrains Ω_m and Ω_Λ, Ω_k, Ω_b and Ω_c.
- Provides an estimate of H_0, although this is in tension with local distance ladder estimates.

This list only contains a subset of what is described in Planck Collaboration et al. (2020b). In any case, It's clear here that the CMB provides us with information about matter (both baryons and CDM), curvature, dark energy, neutrinos, inflation and BBN. Observing the temperature and E-modes of the CMB has been a very useful tool for cosmologists, but does not tell the whole story, as one could guess by looking at the "open questions" section of this thesis.

On the other hand, CMB B-mode measurements have put constraints on the parameter

$$r(k) \equiv \frac{\Delta_h^2}{\Delta_{\mathcal{R}}^2} \propto V. \tag{2.1}$$

In this formula, Δ_h is the dimensionless power spectrum of tensor metric perturbations and $\Delta_{\mathcal{R}}$ is the same but for scalar perturbations. The variable V gives the potential energy of the inflation field, telling one the energy scale at which inflation occurs. As one would imagine, a larger value of r, the tensor to scalar ratio, would imply stronger gravitational waves and thus more B-mode polarization in the CMB. In the literature, one often finds estimates of r evaluated at $k = 0.05\, hMpc^{-1}$. As of October 2021, an analysis (Ade et al. (2021)) combining results from WMAP and KECK constrained $r(0.05) < 0.036$. For context, single field slow roll models predict $r \geq 0.01$. The BICEP array (Hui et al. (2018)) is expected to reduce these limits to $r(0.05) < 0.005$. Moreover, the Simons Observatory is forecasting an uncertainty of $\sigma(r) = 0.003$. Such results would definitively falsify slow roll inflation models.

Galaxy Surveys

Even though it was brief, the previous section of this chapter showed the power of CMB observations. However, it should be emphasized here that CMB observations are measuring the universe as it was at recombination, along with some effects due to gravitational lensing in between. As such, measurements at lower redshifts are required to obtain the full story, especially for dark energy, which becomes dominant after $z = 1$. One way to do this is through galaxy redshift surveys. The

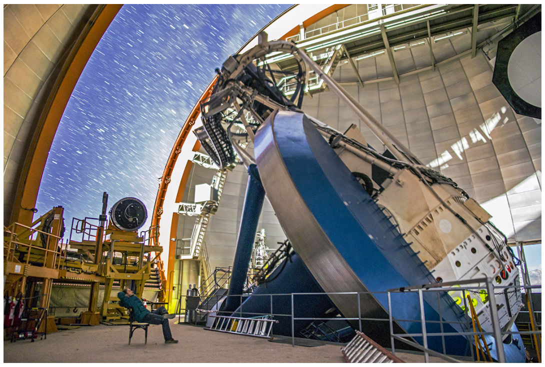

Figure 2.3: An image of the DES instrument, borrowed from their website.

main idea here, which I will discuss more later, is that galaxies are a “tracer" of the total matter distribution. In other words, looking at the galactic power spectrum P_g will tell us about the matter power spectrum P_m, giving us a direct connection between galaxies and cosmological theory. In this section, I will first provide a brief summary of how such surveys are conducted, followed by a review of the benefits that they have for both cosmology and astrophysics.

How do Galaxy Redshift Surveys Work?

Roughly speaking, galaxy surveys are conducted as follows. First, of course, one needs to capture an image of the sky. Such images are captured from ultra-violet wavelengths up to the infrared. For example, the Dark Energy Survey (DES) observes from 400 to 1000 nm (and Flaugher (2005)). In general, such instruments use some combination of lenses and mirrors to focus incoming light onto some focal plane. In the case of DES, the largest optical component (setting the aperture size) is a $4m$ diameter mirror. In figure 2.3, I show a picture of DES borrowed from their website. In general, incoming light at the focal plane will end up being detected by CCDs, much like on a digital camera. In the case of DES, this light is broken upon into 5 sub-bands using filters placed over the CCDs. This is shown in figure 2.4. Presuming the galaxies have features broad enough to be seen in the crude filters, one can then estimate a redshift. This is referred to as a “photometric redshift". Alternatively, as is the case with Dark Energy Spectroscopic Survey (DESI), one can fill their focal plane with optical fibers (Levi et al. (2013)). These optical fibers are then fed to spectrographs that partition the incoming radiation by frequency. One can then find the atomic spectral lines in the received data, which allows for one to make an estimate of “spectroscopic redshift". Naturally, this second method produces more accurate estimates of redshift, but is more complicated. In figure 2.5, I show a picture of the DESI focal plane borrowed from their website.

In addition to collecting spectra, one must also identify galaxies to be observed. This selection is sometimes done by setting a brightness threshold on observed objects. in the case of DESI, a

Figure 2.4: A picture of the SDSS filters, borrowed from their website. These filters are placed over CCD detectors.

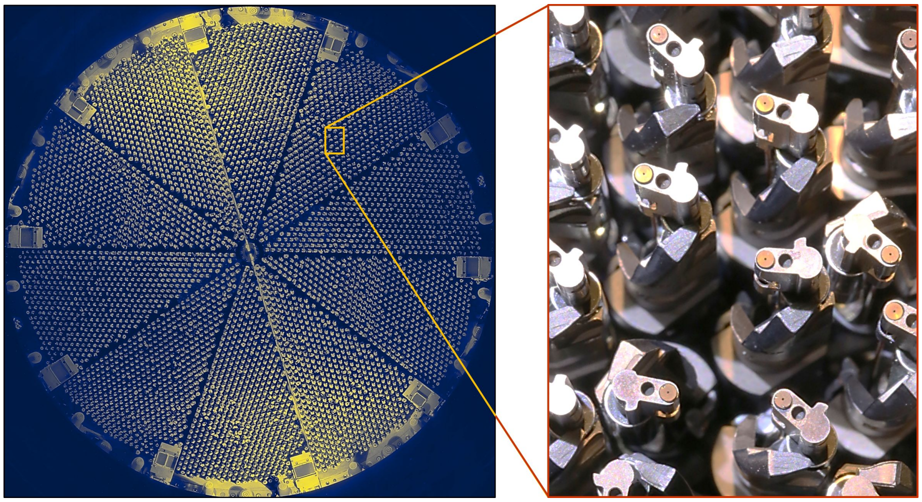

Figure 2.5: Focal plane of the DESI instrument. These fibers are placed to capture images of target galaxies in the plane.

separate telescope is used to identify galaxies to be observed.

After identifying galaxies and estimating their redshift, one obtains a 3-D a map (θ, ϕ, z) of the galaxy distribution. From this map, one may estimate a galaxy power spectrum. For linear scales, we relate P_g to the linear matter power spectrum P_L as (Dodelson (2003))

$$P_g(k, \mu_k, z) = P_L(k, z) \left[b_1(z) + f(z)\mu_k^2\right]^2 + P_{SN}. \tag{2.2}$$

In this formula, b_1 is a bias term accounting for the fact that galaxies do not exactly trace the distribution of matter. This dependence of course may be complicated, but for small perturbations one can simplify it as $\delta_g \approx b(z)\delta_m$. The term $f\mu_k^2$ accounts for the fact that galaxies may have some additional velocity in addition to the Hubble flow. This effect is referred to as “redshift space distortions". The term μ_k is defined as the cosine of the angle between the k-mode of interest and the line of sight. Roughly, this is saying that modes closer to the line of sight will be more severely affected by redshift space distortions. Lastly, the P_{SN} term accounts for shot noise due to the discrete nature of galaxy surveys. Thus, measuring the galactic power spectrum provides one with information about the matter power spectrum, up to bias, RSD and noise terms.

Now that I’ve provided a brief summary of how galaxy surveys work, I’ll move on to discussing the sort of science we can learn from them.

What can be learned from galaxy redshift surveys?

In the previous chapter, we saw that Planck provides estimates for ΛCDM parameters. However, cosmology of course still benefits from having other observational methods. I’ll now discuss some of these.

One thing not discussed in the previous section is that CMB observations have some degeneracy. These degeneracies are discussed somewhat in Tegmark (1997). One example is that CMB observations by themselves are degenerate in Ω_Λ and Ω_k. This degeneracy can be broken however by considering supernovae, lensing and galaxy surveys. In fact, the Planck survey report from 2018 states “CMB power spectrum measurements suffer from a “geometric degeneracy” (see Efstathiou and Bond 1999) which limits their ability to constrain certain extensions to the base cosmology (for example, allowing Ω_K or w_0 to vary). Planck lensing measurements partially break the geometric degeneracy, but it is broken very effectively with the addition of baryon acoustic oscillation (BAO) measurements from galaxy surveys."

Just from this quote, we see one major benefit of galaxy surveys – they can be combined with CMB measurements to produce precise estimates of the curvature density and dark energy equation of state.

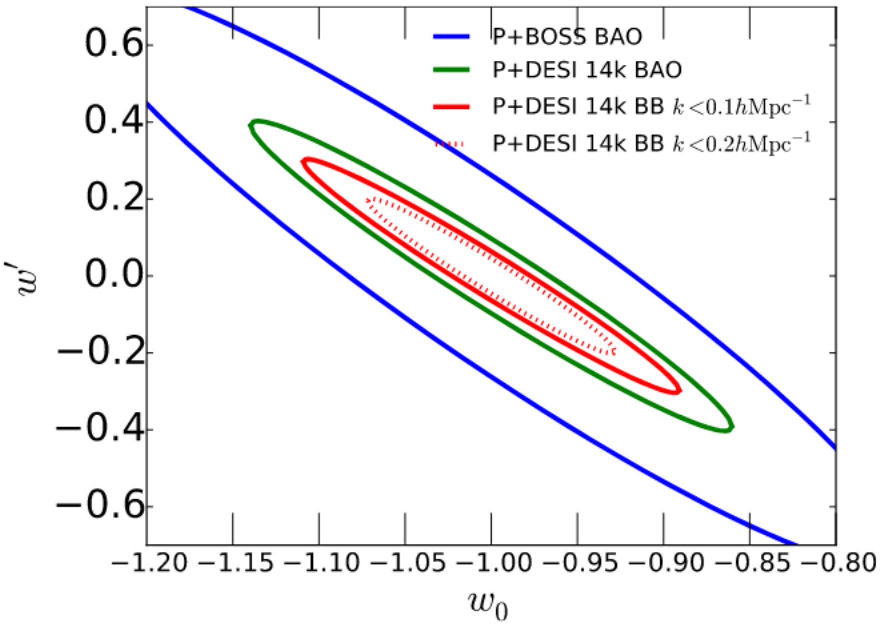

Figure 2.6: A plot of projected dark energy constraints, borrowed from DESI Collaboration et al. (2016). All curves include Planck data. The acronym “BB" standards for the Broad Band power spectrum obtained from lyman-α forest observations.

In addition to breaking degeneracies, galaxy surveys (among other lower redshift observations) provide a double check of CMB observations. By and large, ΛCDM parameters tend to be in good agreement whether estimated from galaxy surveys or CMB data. The fact that we obtain similar results from two completely different approaches should provide some comfort. However, of course, there also exists the Hubble tension. The Hubble tension strongly points towards either new physics or perhaps some unaccounted for systematic. Any new physics we may learn from the H_0 tension would not have been possible without lower redshift observations to complement the CMB.

On the topic of dark energy, it should be noted that galaxy surveys also will help tighten constraints on dark energy parameters. In the DESI science goals paper (DESI Collaboration et al. (2016)), they provide a nice figure showing forecasted dark energy constraints when combined with Planck data. I show this plot in figure 2.6. As one can see, adding more and more galaxy observation data provides us with tighter and tighter dark energy constraints.

Such galaxy survey instruments can also be used to learn about other astrophysical phenomena. For instance, the Sloan Digital Sky Survey (SDSS) instrument is currently being used to gather information about super massive holes (Kollmeier et al. (2019)). The instrument has also been used to perform some measurements of active galactic nuclei (Coffey, D. et al. (2019)).

A Brief Glance at Intensity Mapping

In the last section, we saw that galaxy surveys are a highly useful way of probing large scale structure at lower redshifts. Such measurements break the degeneracies present when analyzing the CMB and serve as a useful double check. Put simply, parameters derived from galaxy or CMB observations ought to be in agreement. Overall, there has been good agreement found between the two approaches, with the Hubble tension being a single glaring disagreement.

However, galaxy surveys cannot tell us everything. One obvious limitation is the fact that galaxy surveys cannot observe the universe during eras in which there were no galaxies! A quick google search of the most distant galaxy ever observed turns up JADES-GS-z13-0, which was observed by JWST at $z \approx 13.2$ (Robertson et al. (2023)). However, galaxy redshift surveys usually start struggling at redshifts significantly lower than this. One problem is that higher redshifts correspond to galaxies further away, implying lower flux. This reduces the observable population of galaxies. This reduced galaxy population naturally increases the shot noise term of equation 2.2, which scales as 1 over the number of galaxies observed. Moreover, the background emission in the sky becomes brighter at longer wavelengths, making observations more difficult as redshift increases (although this is also an issue for Intensity Mapping). Lastly, CCD detectors cannot be used at these infrared wavelengths. The detectors used instead are less sensitive and more expensive. In my quick review of the literature, the highest redshift galaxy survey I could find is the proposed Megamapper experiment which would observe at $2 < z < 5$ (Schlegel et al. (2022)).

One way to overcome these issues with redshift scaling is to use a technique called Line Intensity Mapping (IM) or (LIM), which is the main focus of this thesis. In IM, one does not concern themselves with resolving individual objects. Rather, one measures unresolved emission (or absorption) of various spectral lines. The spectral line of most interest for this thesis is the 21 cm line emitted by neutral hydrogen (HI). However, one can measure other lines, such as those from CO or CII. In figure 2.7, I show a nice image from NASA/Lambda which illustrates the idea behind IM at post-EOR redshifts. The dots represent galaxies while the color represents what is seen by the hypothetical IM instrument.

Since IM does not depend on the presence of galaxies or other bright objects, it can in principle be used to measure LSS at any redshift for which the spectral line of interest is detectable. For instance, HI intensity mapping could in principle be used to estimate the matter power spectrum from present day all the way back to $z \approx 150$, long before the formation of the first luminous objects. Although HI is present out to recombination, the 21 cm signal is not observable at $z > 150$ for reasons that will be discussed later.

Moreover, IM techniques allow for large volumes of space to be probed relatively easily. As mentioned earlier, galaxies surveys must change the type of detector used as they change from optical to infrared wavelengths. On the other hand, HI intensity mapping experiments operate

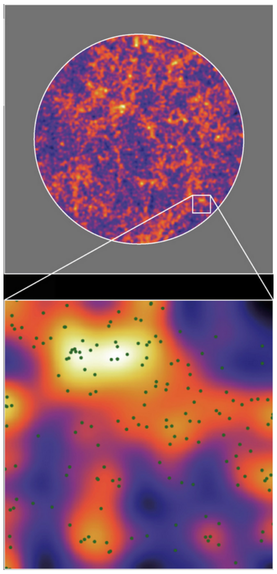

Figure 2.7: An illustration of the idea behind IM at post-EOR redshifts. This image is borrowed from NASA/LAMBDA.

at radio wavelengths, meaning one can measure large radial volumes simply by using broadband receivers and antennas. However, it should be clarified that operating over such a wideband does come with a performance penalty compared to a more narrow band instrument.

In addition, eliminating the requirement to resolve individual objects helps greatly with shot noise at higher redshifts. As mentioned earlier, the population of bright galaxies decreases with redshift, increasing shot noise. On the other hand, HI intensity mapping measures the collected emission from all galaxies in a region, even the dim ones. This results in a lower shot noise contribution. This is illustrated in figure 2.8, which is borrowed from Sailer et al. (2021). This plot shows Megamapper losing sensitivity as redshift increases while the proposed Packed Ultra wide-band Mapping Array (PUMA) instrument remains more stable. Note that PUMA is a proposed HI intensity mapping (HIM) instrument.

These last few paragraphs were intended to give some motivation for why an observational cosmologist would be interested in IM. However, it should be noted that IM is not without its drawbacks. In particular, HIM suffers from systematics such as bright foregrounds, mutual coupling of antenna elements and RFI. Moreover, there exists a degeneracy between $T_b(z)$, the average HI brightness temperature of the HI, and σ_8, which is defined as the integral of the dimensionless power spectrum over a $8hMpc$ cube

$$\sigma_8^2 \equiv \int \Delta^2(k) dln(k) \tag{2.3}$$

. I will discuss these effects in greater detail in the following chapter. What should be emphasized here is that HI intensity mapping does have some weaknesses of its own, making it potentially less

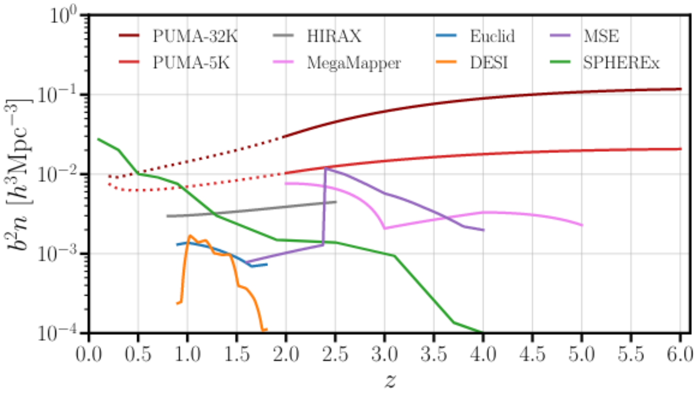

Figure 2.8: A plot of “effective number density" of various surveys as a function of redshift, effectively a stand-in for the noise level of the instrument. For PUMA, noise is mostly thermal rather than shot. Borrowed from Sailer et al. (2021).

sensitive than galaxy surveys when estimating certain parameters.

3 HI INTENSITY MAPPING IN DETAIL

As mentioned earlier, one can perform IM experiments observing spectral lines from various elements or molecules. However, my work these last few years has been focused specifically on HI intensity mapping. As such, from this point forward I will focus specifically on HI intensity mapping.

3.1 The 21cm line

I'd like to start by talking a bit about the 21 cm line. This line comes about due to coupling between the spins of the proton and electron that make up the HI atom. In particular, the $1s$ state of hydrogen will have a slightly lower energy when in the spin-0 state compared to the spin-1 state. This splitting is roughly $5.8\,\mu eV$, which corresponds to a photon of wavelength $\lambda = 21\,cm$.

We describe the ratio of spin-1 to spin-0 HI atoms via a "spin-temperature", denoted as T_s. The equation for the ratio reads

$$\frac{n_1}{n_0} = 3exp(-\frac{h\nu_{21}}{k_B T_S}). \tag{3.1}$$

I will not present a derivation, but note that the HI brightness temperature $T_b(\hat{r}, z)$ is then described via

$$T_b(\hat{r}, z) = (1 - e^{-\tau_{21}(\hat{r},z)})\frac{T_s(\hat{r}, z) - T_{CMB}(z)}{1+z}. \tag{3.2}$$

The optical depth, denoted as τ_{21}, is expected to be quite small, leading to brightness temperatures on the order of mK for post-EOR surveys. Moreover, the 21 cm line will be either emissive or absorptive depending on how the spin temperature compares to the CMB temperature. In figure 3.1, I show a plot with various temperatures as a function of redshift. Note here that these are global temperatures and T_{gas} represents the kinetic temperature of the HI gas. The dashed and solid lines represent two different models for the HI spin temperature. This plot illustrates that for $z > 150$, $T_{CMB} \approx T_S$ and thus $T_b \approx 0$ at very high redshifts. At low redshifts, we find that $T_S > T_{CMB}$ and the 21 cm is seen through emission rather than absorption. For intermediate redshifts, we will see the 21 cm line through absorption and we will have $T_b < 0$.

Now that we have some understanding of the 21 cm line and its temperature brightness $T_b(z)$, we can move on to discussing the 21 cm power spectrum.

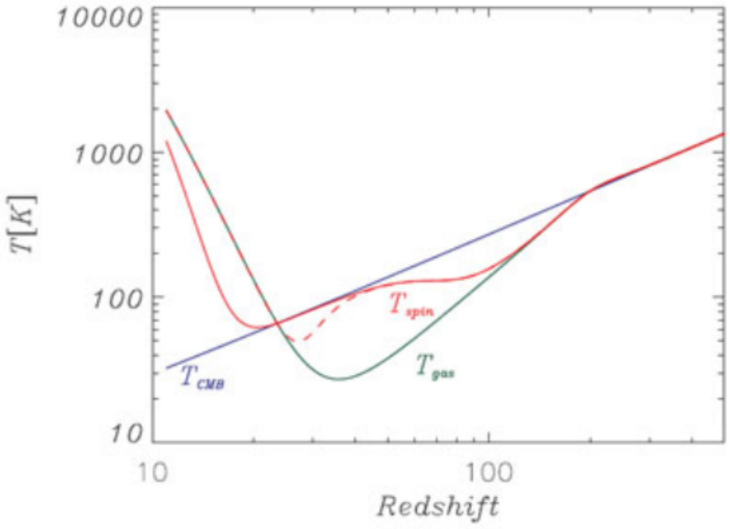

Figure 3.1: A plot of various cosmic temperatures as a function of redshift. Borrowed from `https://ned.ipac.caltech.edu/level5/March14/Zaroubi/Zaroubi4.html`.

3.2 The 21cm Power Spectrum

In HIM, the goal is to measure the spatial distribution of emission (or absorption) of the 21 *cm* line of neutral hydrogen. The HI is in fact a tracer of the matter overdensity, just as galaxies are. As such we can relate the power spectrum observed by an HIM instrument to the matter power spectrum via

$$P_{21}(\mathbf{k}, z) = T_b^2(z) \left[(b_{HI}(z) + f(z)\mu)^2 P_L(\mathbf{k}, z) + P_{SN} \right] + P_N. \tag{3.3}$$

Note that during the EOR, the relationship between P_{21} and P_L is more complicated than what is shown here. In this formula, $T_b(z)$ is the mean brightness temperature of the 21 *cm* line and P_N is a term accounting for thermal noise in the instrument. As was the case with galaxy surveys, we assume a linear bias for small perturbations ($\delta_m = b_{HI}\delta_{HI}$).

This 21 *cm* power spectrum, which is a 2-point statistic, is the main tool used for extracting cosmological information from HI intensity mapping data. However, it should be noted that 3-point statistics of the HI can also be used to search for primordial non-Gaussianity.

Thus, 21 *cm* intensity mapping tells us about the statistics of matter in the universe from present day back to $z \approx 150$, up to bias and RSD terms. In Ansari et al. (2019), it is stated that the degeneracy between these parameters could be broken either through cross-correlations or perhaps by modeling the power spectrum in the mildly non-linear regime.

Let's now review some of the new physics that could be learned from 21 cm intensity mapping the different observable eras.

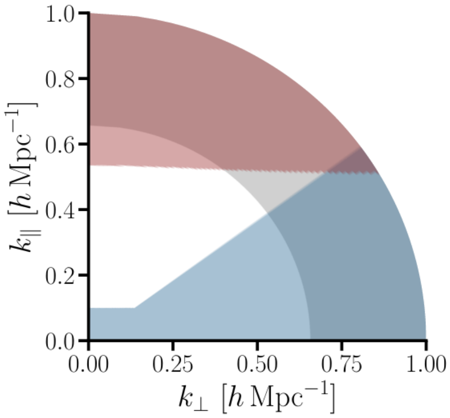

Figure 3.2: A plot showing the assumptions of Sailer et al. (2021). This plot is for $z = 3.9$. Blue regions represent modes assumed to be inaccessible to PUMA under pessimistic assumptions. Grey represents nonlinear modes and red represents modes inaccessible to MegaMapper due to Finger of God (FoG)-like effects.

3.3 HI Intensity Mapping in the Post-EOR era

When considering 21 cm intensity mapping at $z < 6$, the work presented in Sailer et al. (2021) provides excellent forecasting and summaries for the future of cosmology in the post-EOR epoch. Particularly, the Sailer paper focuses on the proposed PUMA telescope, which would measure half the sky over a redshift range $0.3 < z < 6$. A detailed description of PUMA can be found in Ansari et al. (2019). On the other hand, the Sailer et al. paper also focuses on the proposed MegaMapper instrument, which would perform a galaxy redshift survey over about 1/3 of the sky for redshifts $2 < z < 5$. This paper highlights the promise of future LSS surveys and shows that galaxy surveys and intensity mapping probes complement each other at such redshifts.

Rather than going in depth into the Sailer et al paper, I will just summarize some key results. It should be noted here that Sailer et al make separate “pessimistic" and “optimistic" forecasts for PUMA. The difference between the optimistic and pessimistic cases is whether the previously mentioned foregrounds can be effectively cleaned from the data. In figure 3.2, I show a plot from Sailer et al. (2021) illustrating their pessimistic assumption. The BAO features of interest are located at $k < 0.4 Mpc^{-1}$ and will thus still be mostly observable even under this “pessimistic assumption". To me this “pessimistic" assumption seems somewhat optimistic, since, from the analyses I describe in my thesis, it appears that more $k_{||}$ modes will likely be lost due to beam chromaticity and polarized foregrounds.

In any case, here is a brief list of some key results from Sailer et al. (2021)

- Megamapper and PUMA would put constraints on angular diameter distance and $H(z)$ out to $z = 6$. In the case of optimistic foregrounds, PUMA would provide sub percent level errors.

In the case of pessimistic foregrounds, we get errors on level of 1 percent at z =6. Such results are of interest for constraining dynamical dark energy models.

- Megamapper and PUMA would provide estimates of σ_8 for the whole post-EOR epoch. In this case, Megamapper will provide more precise estimates than PUMA. This is due to the degeneracy between T_b and $f\sigma_8$ present in HI measureuments. It should be emphasized, however, that only PUMA would be capable of providing estimates of σ_8 for $z > 5$.

- Interestingly, Megamapper and PUMA could also help constrain neutrino physics! These post-EOR LSS surveys would constrain M_ν, the sum of the neutrino masses. When combined with CMB measurements, the forecast error would be at the level of 20 meV. This would be a significant improvement over the current limit of $120\,meV$. In addition to providing an excellent probe of the mass, these LSS probes, when combined with Planck, would provide a tighter constraint on N_{eff}, the effective number of neutrino degrees of freedom. A value of N_{eff} significantly different than the ΛCDM value could indicate new neutrino physics. Based on the Sailer forecasts, these results don't seem to be impacted much by pessimistic foreground assumptions.

- Sailer et al consider a relatively simple model for primordial non-Gaussianity. In this model, the level of non-Gaussianity is parameterized using a number f_{NL}^{loc}. Currently, analysis of Planck data provides an estimate of $f_{NL}^{loc} = -0.9 \pm 5.1$. In other words, Planck data is consistent with no primordial non-Gaussianity. Sailer et al forecast an error of $\sigma(f_{NL}^{loc}) = 1$ after LSS survey data is included, a factor of 5 improvement. For this flavor of primordial non-Gaussianity, pessimistic foreground assumptions would severely hamper the sensitivity of PUMA. This is due to the fact that this type of non-Gaussianity leaves an imprint on length scales expected to be contaminated by foregrounds.

- Sailer et al also consider the presence of sinuosoidal features in the primordial power spectrum. These would take the form

$$P_m(k) -> [1 + A_{lin} sin(\omega_{lin} k + \phi_{lin})]\, P_m(k). \tag{3.4}$$

  Such power spectrum features would correspond to scenarios in which there exists a sharp feature in the inflationary potential at some length scale. In this case, Megamapper + DESI would provide errors $\sigma(A_{lin}) \sim 7 * 10^{-4}$. On the other hand PUMA + DESI with optimistic foreground estimates would provide $\sigma(A_{lin}) \sim 2 \times 10^{-4}$. This error would increase by a factor of two with pessimistic foreground assumptions.

- Lastly, these future LSS surveys would help constrain ΛCDM parameters.

3.4 HI IM during the EOR

During the EOR, 21 *cm* fluctuations will depend upon both cosmology and astrophysics. As such, it would be more difficult to extract cosmological information from EOR observations as compared to post-EOR observations. However, such measurements would certainly provide us with interesting information about the astrophysics of the EOR. To illustrate this, I'll begin by presenting the following model of 21 *cm* fluctuations from Mirocha et al. (2022):

$$\delta T_{21}(\mathbf{x}, z) = T_0(z)(1 + \delta_b)(1 + \alpha b(\mathbf{x}, z))(1 + \delta_\phi). \tag{3.5}$$

This formula introduces several new variables which I'll now introduce. The variable $T_0(z)$ is a normalization factor that is equal to a cosmological term times $1 - T_{CMB}/T_s$. We also have δ_b which is the baryon overdensity. b is a binary term accounting for whether or not a bubble is present, while α accounts for the effect of the bubble on δT_{21}. For an ionized bubble, $\alpha = -1$ corresponding to suppression of 21 *cm* emission or absorption. The variable δ_ϕ describes perturbations in the temperature contrast and is defined via spin-temperature perturbations as

$$\delta_\phi \equiv \frac{T_{CMB}}{T_s - T_{CMB}} \delta T_s. \tag{3.6}$$

In figure 3.1, we see that during the EOR, T_s and T_{CMB} are at the same order of magnitude and T_s varies significantly as a function of redshift (at least for the dashed line model). This will leave an imprint in δT_{21} via the $T_0(z)$ and δ_ϕ terms.

As we exit the EOR, ionized regions will fill the entire space, leading to $b = 1$ everywhere. Moreover, we will find T_{CMB}/T_s approaching 0. These two events greatly simplify equation 3.5 and remove the astrophysical complications. Note that $\alpha = -1$ is really a simplification, since ionized regions will still contain some neutral hydrogen, as mentioned earlier in this thesis. Once T_s shoots upwards, our model for δT_{21} will tend towards the expected post-EOR behavior.

Due to these additional complications, experiments observing the 21 *cm* signal during the EOR tend to be focused on extracting astrophysics information from the data. However, it is possible to extract some cosmological information from the EOR, though this often seems to be a secondary goal. In figure 3.3, I show some forecasted constraints published by the Hydrogen Epoch Of Reionization Array (HERA) collaboration in 2016 (DeBoer et al. (2017)). Although not up to date, I think this provides some idea of what EOR observations are looking to achieve.

In this figure, ζ represents the ionizing efficiency of galaxies. This efficiency describes the ratio of ionizing photon production rate over the non-ionizing UV luminosity density. This figure also include R_{mfp}, the mean free path of an ionizing photon in an H-II region. The variable T_{vir}^{min} is the minimum Virial temperature of an ionizing galaxy. Lastly, f_{esc} describes the fraction of ionizing

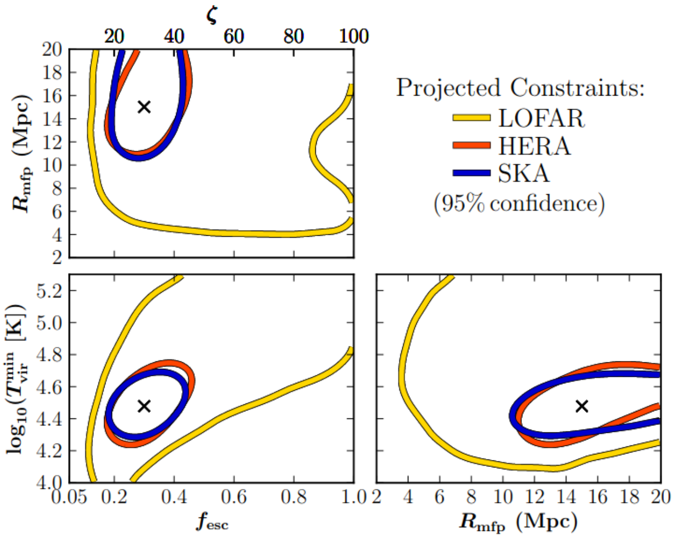

Figure 3.3: Forecasted limits for HERA from DeBoer et al. (2017).

photons that escape from galaxies into the intergalatic medium.

Assuming HERA is representative of other EOR surveys, we see that the main motivation for making 21 *cm* observations during the EOR is to understand the reionization process. This is done by constraining parameters such as those described here.

3.5 HI IM during the Dark Ages

Luckily, the relationship between the 21 *cm* signal and cosmology once again becomes simpler during the dark ages. Complex astrophysics does not yet play a role, and modes have not had too much time to grow, leading to less non-linearity than in the post-EOR epoch. Moreover, small scale perturbations in the 21 cm signal will not suffer from the diffusion damping seen at small scales in the CMB

Although HI will be present in significant abundances any time after recombination, there will only be significant emission of absorption of the 21 *cm* line for $z < 150$. This is a consequence of the fact that the HI spin temperature will be kept close to the CMB for higher redshifts via Compton scattering. This is illustrated in figure 3.1. Despite this limitation, performing observations of the 21 *cm* line at redshifts $30 < z < 150$ would in fact provide us with a wealth of cosmological information.

It should also be noted that dark ages observations would certainly present a challenge for experimentalists. For one, such an array would not do well on Earth due to atmospheric effects present at the low frequencies required for such observations. Thus, it would require the construction of a large telescope outside of Earth's atmosphere, perhaps on the Moon. Moreover, the previously mentioned issue of foregrounds will become even more severe as redshift increases.

Even though there are challenges, the scientific promise of dark ages observations still provides enough motivation to investigate the possibility. The ability to measure linear modes down to very small scales makes dark ages observations a unique cosmological probe. Put simply, a well conducted and precise dark ages telescope would provide an unrivaled view of primordial physics. I'll now provide a brief list with examples of the power that a future dark ages 21 *cm* survey could have.

- In Muñoz et al. (2015) it is stated that a survey from $30 \leq z \leq 100$ would be able to constrain local primordial non-Gaussianity with $\sigma = 0.03$. This would be a roughly factor of 30 improvement over the limits described in Sailer et al. (2021).
- In Ansari et al. (2019), there is a summary of constraints that could be provided on the tensor to scalar ratio by a dark ages 21cm experiment. Supposing one had an array with baselines reaching hundreds of kilometers, they would be able to constrain r much more effectively than what is possible with CMB surveys. However, it is unclear to me whether or not it will be feasible to build a huge 21 *cm* array on the moon.
- In Chen et al. (2016), it is argued that a future dark ages observation would make an excellent probe of primordial features on small scales. This would provide us with information about the physics of inflation that would not be detectable using CMB observations due to diffusion damping.
- In Cole and Silk (2020), it is argued that an adequately large value of the primordial power spectrum at small scales would imply the creation of primordial black holes, a dark matter candidate. Such scales would be visible and linear during the dark ages.

I'm sure this list is not complete, but it does provide some motivation for the sort of physics that could be probed using a dark ages instrument.

4 HI INTENSITY MAPPING IN PRACTICE

As mentioned earlier, HI IM looks at the redshifted 21 cm line of hydrogen. This corresponds to frequencies

$$\nu(z) = \frac{1420.8}{1+z}\, MHz. \tag{4.1}$$

As such, the HI IM approach observes radio frequency radiation. This of course leads to antennas being used to observe the 21 cm line.

There are a lot of factors to consider when designing a telescope for HIM. For instance, one would like for their instrument to have large sky coverage and a wide bandwidth, maximizing the volume observed. One would also like for their instrument to have adequate angular resolution (for instance, to see the BAO) and a high signal-to-noise ratio at the scales of interest. As I will explain later, it is also particularly important to keep systematics under control when designing instrumentation for HIM.

In this chapter, I will discuss the sorts of instruments used in HIM mapping and how the measurement process works.

4.1 What Sorts of Telescopes are Used for HIM?

In this section, I'll provide some discussion of the sorts of telescopes one sees being used for HIM.

The most common instrument one sees being used is a reflective dish paired with a feed antenna. Such dish-based designs are commonly used for several reasons. For one, they have a relatively small field of view (FOV). A small FOV is desirable since it allows one to use what is called the “flat sky approximation". In the flat-sky approximation, one uses the fact that they are looking at a small patch of the sky to make simplifications to their analysis. Moreover, in the context of arrays, the use of dishes provides some opportunity to decrease systematic effects. I will discuss this more later. One will see these dish-based instruments being used in both post-EOR and EOR contexts.

Among dish-based instruments, one will find some that use a single large antenna element. One example of this is the Green Bank Telescope (GBT) (Switzer et al. (2013)), which I show in figure 4.1. Note that the GBT uses what is called an “off-axis" design in which one uses an asymmetrical reflector that has the focus located off to the side. Additional examples of telescopes using a single large dish include The FAST telescope (Bigot-Sazy et al. (2016)) and the BINGO telescope (Wuensche (2019)). It should be noted that telescopes using a large single antenna are not the most common approach to HIM.

Motivated by sensitivity and cost, radio-interferometers are typically used instead of single dishes. In interferometry, an array consisting of many antenna elements is used. The signals from

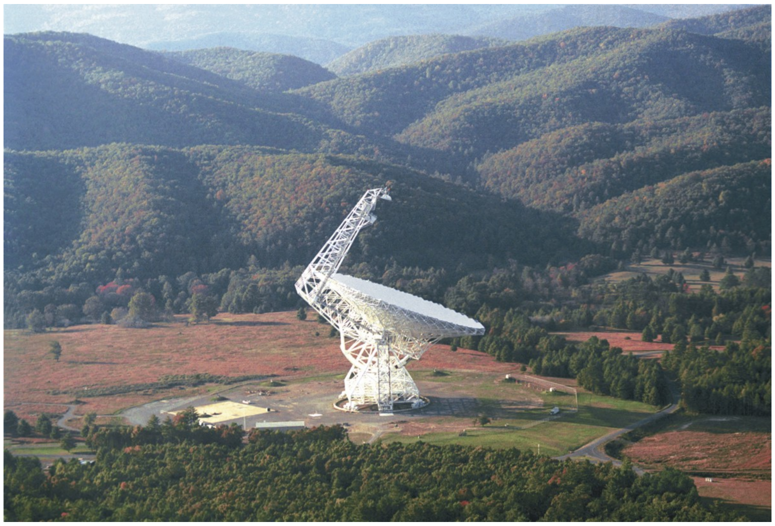

Figure 4.1: A picture of the GBT, borrowed from `https://en.wikipedia.org/wiki/Green_Bank_Telescope`.

pairs of antennas are combined to form “visibilities". The distance between the antennas sets the angular scale probed by a particular visibility. By having redundant antenna pairs, one can measure the same power spectrum mode many times, increasing sensitivity. In order to maximize sensitivity and access to large angular scales, one finds these instruments often taking on “large N small D" designs, in which a large number of small diameter dishes are used.

Examples of dish-based interferometric instruments include Tianlai (Chen (2012)), HERA (DeBoer and HERA (2015)), HIRAX (Saliwanchik et al. (2021)), the proposed CHORD telescope (Vanderlinde et al. (2019)) and the proposed PUMA telescope (Ansari et al. (2019)). In figure 4.2, I show a picture of the dishes used in in the HERA telescope.

It should be noted that one could also look at the signals from their dishes one at a time. I refer to this approach as “single dish mode". This approach eschews small-scale angular modes in favor of a more simple analysis. One example of such an instrument is the MeerKAT experiment (Santos et al. (2016)).

However, not all HIM instruments are dish-based. One alternative is to instead have long cylindrical reflectors with many feeds placed along the axis. Such an approach is being used in experiments such as Tianlai and CHIME (Bandura et al. (2014)). Cylindrical reflectors are also incorporated into the proposed CHORD instrument (Vanderlinde et al. (2019)). In figure 4.3, I show a picture of a cylindrical reflector used in CHIME. Such cylindrical reflectors were considered since they provide a beam that is wide along one axis and narrow along another, somewhat like a fan. This type of beam eliminates the need for repointing of the antennas that one encounters in dish-based instruments when surveying large portions of the sky. Moreover, the large field of view seen in cylindrical reflectors has also made them useful as a probe of Fast Radio Bursts (FRB). The

Figure 4.2: A picture of the HERA telescope. The metal objects hanging over the dishes are the feed antennas.

Figure 4.3: A picture of the cylindrical reflectors used in the CHIME experiment. This picture is borrowed from the McGill University website.

downside to using a cylindrical reflector, however, is that it will include many closely packed feed antennas. Naturally, these feeds will strongly interact, complicating beam patterns and most likely introducing a significant noise offset.

One last option to mention is the use of phased arrays. One experiment that makes use of phased arrays is the LOw-Frequency ARray (LOFAR) (van Haarlem et al. (2013)). In figure 4.4, I show a picture of the LOFAR experiment, which consists of a collection of phased arrays. Phased arrays work by combining the signals from many relatively small antenna elements to form a single beam pattern. Depending on the relative phases chosen in the summation, we can focus the phased array

Figure 4.4: A picture of the LOFAR telescope

on different patches of the sky. In principle, this would allow one to have many beams on the sky at once, although this would perhaps be expensive in terms of signal processing.

4.2 The Sky Signal and its Measurement With a Single Antenna

The Sky Signal

In the context of HI intensity mapping, we suppose that the incoming electric field from direction $\hat{n}$ at time t is described via

$$d\mathbf{E}(t) = \mathbf{e}(\hat{n}, \nu) exp(-2\pi i \nu t) d\nu d\Omega. \tag{4.2}$$

In this formula, **e** describes the incoming electric field per spectral frequency per solid angle. Note that **e** will have a different phase depending on where the measuring antenna is located. For now, I will ignore this phase, but it will become important when I discuss interferometry. Since the signal from the radio sky is due to many different sources, we can describe it as “incoherent". As such, we imagine the incoming electric field as being drawn from a Gaussian distribution that is stationary with time. Moreover, we assume that emission from different lines of sight and different frequency bins are uncorrelated. This means that for polarization components c and d, we find

$$\langle e_c(\hat{n},\nu)e_d^*(\hat{n'},\nu')\rangle = \frac{2k_B Z_0}{\lambda^2} T_{cd}(\hat{n},\nu)\delta(\hat{n}-\hat{n}')\delta(\nu-\nu'). \tag{4.3}$$

In this formula, T_{cd} is a "coherence matrix" with units of temperature. The variable Z_0 is the impedance of free space. This matrix describes the properties of the incoming sky signal.

I'll now illustrate how this correlation is estimated. First, suppose that we observe over some finite time window centered at time $t = 0$. Let's denote this window by $W(t)$. Now, if we use W to perform a short time Fourier transform (STFT), we find that

$$\langle e_{c,W}(\hat{n},\nu)e_{d,W}^*(\hat{n'},\nu)\rangle = \delta(\hat{n}-\hat{n}')\frac{2k_B Z_0}{\lambda^2}\int d\nu'' T_{cd}(\hat{n},\nu-\nu'')|W(\nu'')|^2. \tag{4.4}$$

Thus, computing two point statistics of windowed electric fields provides us with an estimator of the true underlying statistics. We can then form an estimate of the underlying statistics by computing this correlation for many time chunks, then averaging. The chunk size chosen will have an impact on noise levels and frequency resolution. I will discuss more details on this topic later.

From Sky Signal to Voltage to Power to Temperature

Now, I'll explain how the sky signal becomes a voltage at the output of the antenna. We usually model the receiving antenna as a voltage generator with impedance Z_A. The voltage produced by the antenna will be given by integrating over contributions from all lines of sight and summing over polarizations:

$$v_{open}(\nu) = \int \sum_c l_c(\hat{n},\nu)e_c(\hat{n},\nu)d\Omega. \tag{4.5}$$

In this formula, $\mathbf{l}$ is referred to as the "effective length" of the antenna. The resulting voltage is referred to as v_{open}, since it is the voltage that would be measured across an open circuit load attached to the antenna terminals.

On the other hand, if we attach a receiver with impedance Z_L, we would obtain a receiver voltage

$$v_{rcv}(\nu) = \frac{Z_L}{Z_A + Z_L} v_{open}(\nu). \tag{4.6}$$

The power per unit frequency absorbed by the receiver due to the antenna is then given by

$$P_{rcv} = \frac{1}{2}\mathfrak{Re}(v_{rcv}I_L^*) = \frac{1}{2}\frac{R_L}{|Z_A + Z_L|^2}|v_{open}|^2. \tag{4.7}$$

In this formula, I_L is the current at the receiver. We can say that, on average

$$|v_{open}|^2 = \frac{2k_B Z_0}{\lambda^2} \int \sum_{cd} l_c l_d^* T_{cd} d\Omega. \tag{4.8}$$

Plugging this into our power equation, we find that the average receiver power is

$$P_{rcv}(\nu) = \frac{1}{2} \frac{R_L}{|Z_A + Z_L|^2} \frac{2k_B Z_0}{\lambda^2} \int \sum_{cd} l_c l_d^* T_{cd} d\Omega. \tag{4.9}$$

Let's now define a telescope beam term:

$$B_{ab} = \frac{1}{2} \frac{R_L}{|Z_A + Z_L|^2} \frac{2Z_0}{\lambda^2} l_c l_d^*. \tag{4.10}$$

Plugging B_{ab} into our power equation gives

$$P_{rcv}(\nu) = k_B \int \sum_{cd} B_{cd} T_{cd} d\Omega = k_B T_p. \tag{4.11}$$

Thus, we see that the power at the receiver can be expressed in terms of a temperature T_p. In the HIM context, we will normalize T_p to get the "antenna temperature". I'll now show how this works.

Suppose that emission from the sky is unpolarized. In such a case, we can eliminate the cross terms and write our beam as

$$B_{unpol} = \frac{1}{2} \frac{R_L}{|Z_A + Z_L|^2} \frac{2Z_0}{\lambda^2} |\mathbf{l}|^2 \tag{4.12}$$

and

$$T_p = \int B_{unpol}(\hat{n}, \nu) T(\hat{n}, \nu) d\Omega. \tag{4.13}$$

We will define the "antenna temperature" as

$$T_A = \frac{T_p}{\Omega_i}. \tag{4.14}$$

In this formula, I have introduced the beam solid angle variable

$$\Omega_i \equiv \int B_{unpol} d\Omega. \tag{4.15}$$

Note that the subscript i will later be used to label the antenna being considered. Supposing the sky is at uniform temperature T_{sky}, then we will find that $T_A = T_{sky}$. Thus this normalization makes it easy to compare the sky temperature to the temperature at our receiver.

From Sky Signal to Voltage, in Practice

In practice, the voltage measured by the receiver will also include a contribution from noise. As such, we will describe the measured voltage as

$$v_{rcv}(t) = s(t) + n(t). \tag{4.16}$$

In this formula, v_{rcv} is the voltage at input of the receiver, $s(t)$ is the contribution from the sky and $n(t)$ is the contribution from noise. In practice, we cannot measure the incoming signal at all times. Rather, we collect discrete samples, giving

$$v_{rcv}[t_a] = s[t_a] + n[t_a]. \tag{4.17}$$

In this formula, a is used to index the samples. The sampling rate is usually chosen to ensure that $\nu_{sample} > 2(\nu_{max} - \nu_{min})$. In other words, we need to make sure that we are sampling at or above the Nyquist limit. Once we have our samples, we break the time stream into chunks of length N, much like what was described in our discussion around equation 4.4. We can then perform a discrete Fourier transform (DFT) on one of our windows (which begins at time t_A) to obtain

$$\tilde{v}_{rcv}[\nu_b; t_A] = \sum_a v_{rcv}[t_{A+a}] w[a] exp(\frac{-2\pi i b a}{N}). \tag{4.18}$$

In this equation, the function $w[c]$ represents the window function used. One can show that this DFT results in frequency bins of width ν_{sample}/N. Thus, larger time windows will provide us with finer frequency resolution. However, such large windows will also decrease the number of spectra collected, decreasing sensitivity.

It should be noted that I describe this process using a DFT since this is the most conceptually simple approach. However, most interferometers use polyphase filter banks (Price) to form their spectra. According to Liu and Shaw (2020), this approach provides better frequency localization but also leads to correlations between samples.

4.3 Single Dish

The simplest method for HI intensity mapping is what I will refer to as the “single dish" approach. I’ve put “single dish" in quotes since this description can also include situations where one averages the auto-correlation signals observed from many dishes. One early notable example of a single dish observation is the use of the Green Bank Telescopes (GBT) in 2009. However, there also exist telescopes that work by averaging auto-correlation signals from many dishes. Examples of this include MeerKat (Santos et al. (2016)) and SKA-Mid (Norris et al. (2014)), which is currently under construction.

The Single Dish Measurement Process

In the single dish measurement process, our physics comes out of taking the square of the antenna voltage

$$\hat{V}_{ii}[\nu_b] = \frac{1}{M}\sum_{A=1}^{M} \tilde{v}_i\,[\nu_b; t_A]\,\tilde{v}_i^*[\nu_b; t_A]. \tag{4.19}$$

In this equation, M represents the number of spectra collected and i is an index labeling the antennas in the array, assuming that multiple antennas are used. The expectation value for this estimator will come out to be

$$\langle \hat{V}_{ii}[\nu_b]\rangle = \alpha \int d\nu d\Omega \sum_{cd} B_{ii}^{cd}(\hat{n},\nu) T_{cd}(\hat{n},\nu) |W(\nu-\nu_b)|^2 + \int d\nu N(\nu) |W(\nu-\nu_b)|^2 \tag{4.20}$$

In this formula, $N(\nu)$ is the variance of receiver voltage noise and $W(\nu - \nu_b)$ is the Fourier transform of the time window $w(t)$. We also have defined

$$\alpha(\nu_b) = \frac{2|Z_L|^2 k_B}{R_L}. \tag{4.21}$$

As of now, our estimator has units of voltage squared. We can convert this into units of antenna temperature by scaling by a factor

$$\beta = \frac{1}{\alpha\Omega_i}. \tag{4.22}$$

We can now define a new estimator

$$\langle \hat{\mathcal{V}}_{ii}[\nu_b]\rangle \equiv \langle \beta[\nu_b]\hat{V}_{ii}[\nu_b]\rangle. \tag{4.23}$$

Our new variable $\mathcal{V}_{ii}$ has antenna temperature units, allowing us to easily relate our measured quantity to the sky. I will refer to quantities such as $\mathcal{V}_{ii}$ as “autocorrelations". Supposing that none of the quantities of interest vary significantly over the width of a frequency bin, we can say that

$$\langle \hat{\mathcal{V}}_{ii}[\nu_b]\rangle \approx \frac{1}{\Omega_i}\int B_{ii}^{ab}(\hat{n},\nu_b) T_{ab}(\hat{n},\nu_b) d\Omega + T_{rcv,A}(\nu_b). \tag{4.24}$$

The variable $T_{rcv,A} \approx N(\nu_b)\beta$ is the receiver noise variance in antenna temperature units. Thus we can say that $\hat{\mathcal{V}}_{ii}$ is an estimator of the theoretical auto-correlation

$$\mathcal{V}_{ii} = \frac{1}{\Omega_i} \int B_{ii}^{ab} T_{ab} d\Omega. \tag{4.25}$$

This auto correlation will have variance

$$var(\hat{\mathcal{V}}_{ii}[\nu_b]) = \frac{T_{sys}^2(\nu_b)}{M}. \tag{4.26}$$

In this case, we find that the “system temperature" is given by

$$T_{sys} \equiv T_A + T_{rcv,A}. \tag{4.27}$$

It should be noted that a poor impedance match between antenna and receiver will decrease the amount of power supplied to the receiver from the sky. In the case of equation 4.27, this would correspond to an increase in $T_{rcv,A}$.

Weaknesses of Single Dish Observations

I began the discussion of IM techniques by talking about single dishes due to the relative simplicity. However, taking a single dish approach does come with disadvantages. For one, it becomes difficult to probe small angular scales. Recall that angular resolution for a circular aperture goes like

$$\theta \sim \lambda / D \tag{4.28}$$

where D is the diameter of the aperture. As such, one needs to build larger and larger diameter dishes to probe smaller scales. In practice, this will make single dish instruments ineffective at observing the BAO as redshift increases. This is illustrated nicely in figure 4.5, which is borrowed from Bull et al. (2015). This figure illustrates how a single dish instrument with $D = 15\,m$ will be become unable to measure BAO scales as the redshift increases.

Another disadvantage of the single dish approach can be seen in equation 4.24. In this equation, we see that the autocorrelation of a single dish has a contribution $T_{rcv,A}$ from the receiver noise. In a typical HIM context, one will find $T_{rcv,A}$ on the order 50 - 100 K. On the other hand, the HI signal is expected to be on the order of $1\,mK$ for post-EOR observations. As such, this receiver noise is much larger than the signal we would wish to observe. This means that a single dish instrument would require a high level of stability combined with an appropriate scanning scheme to allow for the noise pedestal to be subtracted out.

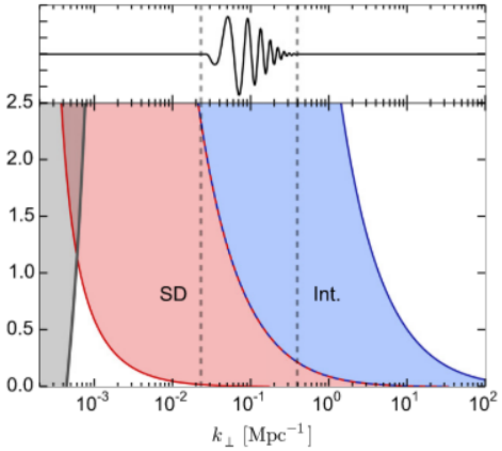

Figure 4.5: Illustration of BAO features and the ability to measure them from Bull et al. (2015). The black curve shows the BAO features as a function of $k_\perp$. The red shaded area describes $k_\perp$ modes probable by a single dish with $D = 15\,m$. The blue shaded region describes an interferometer with baselines of length $b > 15\,m$.

These two issues motivate the use of interferometers, which I will discuss in the next section.

4.4 Interferometers

In practice, most HIM experiments are interferometer-based. Examples include Tianlai (Chen (2012)), CHIME (Bandura et al. (2014)), HERA (DeBoer and HERA (2015)), LOFAR (van Haarlem et al. (2013)), the proposed PUMA experiment (Ansari et al. (2019)) and others. In interferometry, one's instrument consists of a collection of relatively small antenna elements. For example, in the case of Tianlai and HIRAX, these dishes are $6\,m$ in diameter, which of course is much smaller than the $100\,m$ diameter used by GBT. As in the single dish case, our science comes from squaring signals. In particular, we form a “visibility" by correlating signals from antennas i and j. The act of correlating signals from pairs of antennas will allow us to probe smaller angular scales and will also (at least in theory) eliminate the noise pedestal. Let's suppose that antennas i and j are located at positions $\mathbf{r}_i$ and $\mathbf{r}_j$. Due to their different locations, an incoming planewave will have a different phase as it arrives at antenna i as compared to antenna j. As such, we will update equation 4.5 to read

$$v_{open,i} = \int \sum_c l_c(\hat{n}, \nu) e_c(\hat{n}, \nu) e^{-\frac{2\pi i}{\lambda}(\mathbf{r}_i \cdot \hat{n})}. \tag{4.29}$$

In this equation, $\mathbf{e}$ now describes the incoming wave as it is at $\mathbf{r} = 0$. The exponential term encodes the phase difference of the wave between locations $\mathbf{r} = 0$ and $\mathbf{r}_i$. Now, our true visibility

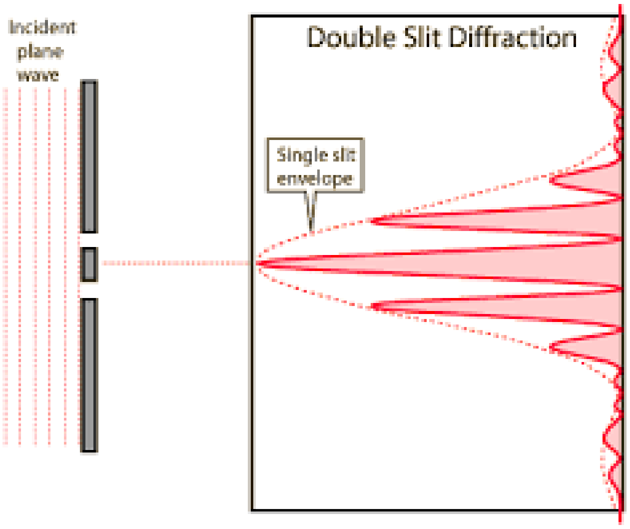

Figure 4.6: Double slit cartoon borrowed from `http://hyperphysics.phy-astr.gsu.edu/hbase/phyopt/slits.html`.

will look like

$$\mathcal{V}_{ij}(\nu) = \int B_{ij}^{cd}(\hat{n}, \nu_b) T_{cd}(\hat{n}, \nu_b) d\Omega. \tag{4.30}$$

In this equation, we have

$$B_{ij}^{cd}(\hat{n}, \nu_b) = \sqrt{\alpha_i \alpha_j} l_{c,i} l_{d,j}^{*} e^{-\frac{2\pi i}{\lambda}(\hat{n}\cdot(\mathbf{r}_i - \mathbf{r}_j))}, \tag{4.31}$$

where, as in our single dish discussion, the α prefactors are there to normalize our visibilities to antenna temperature units. It should be noted that the exponential term in equation 4.31 is extremely important. This is what gives an interferometer its ability to probe small angular scales! I'll describe this through comparison, then mathematically. First, note that a pair of antennas such as we have is essentially a double slit apparatus. However, instead of slits we have the effective apertures of the antenna elements. In figure 4.6, I show a double slit cartoon borrowed from hyperphysics.

Just as in the double slit case, radiation from a pair of antennas will produce an interference pattern on the sky. As we increase the distance between the antennas, we will find that the fringes of the interference pattern become closer together, corresponding to finer angular resolution on the sky. Thus, different baseline lengths will allow us to probe different angular scales, with longer baselines probing the finer scales.

Now that we have seen a illustration, I'll provide some mathematical explanation. Oftentimes, such discussions revolve around the flat-sky approximation, which depends on one having a small beam pattern relative to the curvature of the sky. I don't prefer this approach, since it does not apply

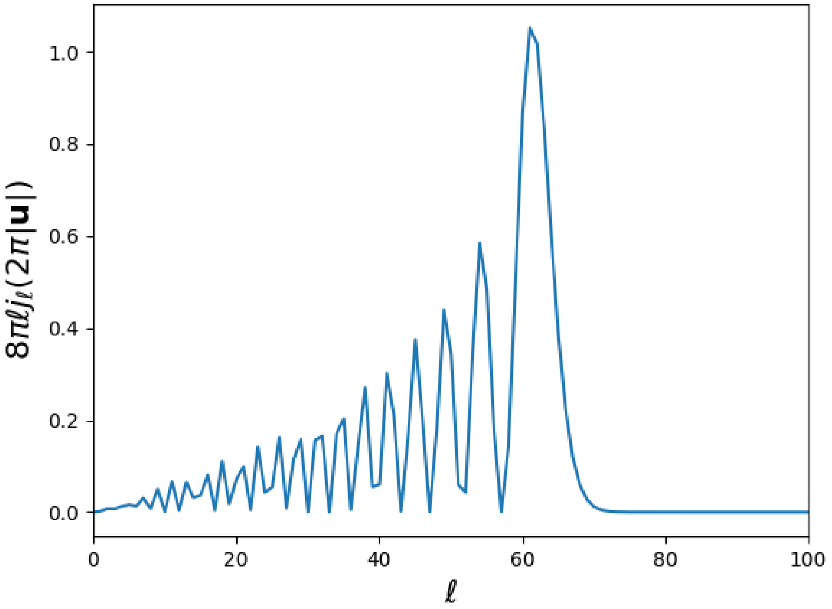

Figure 4.7: Illustration of sensitivity in l-space for a baseline with $2\pi|\mathbf{u}| = 65$.

to instruments with larger beam areas such as CHIME. In particular, I'll summarize the discussion from Shaw et al. (2015). Let's say that $\mathbf{u} = \frac{\mathbf{r}_i - \mathbf{r}_j}{\lambda}$. We can then rewrite our exponential term as $e^{-2\pi i \mathbf{u}\cdot\hat{n}}$. Rewriting this expression as a series of spherical harmonics, we find

$$e^{-2\pi i \mathbf{u}\cdot\hat{n}} = \sum_{lm} \left[4\pi i^l j_l(2\pi|\mathbf{u}|) Y^*_{lm}(\hat{u})\right] Y_{lm}(\hat{n}). \tag{4.32}$$

For $l >> 1$, we can approximate the spherical harmonic coefficients as

$$|a_{lm}|^2 \approx 8\pi l j_l(2\pi|\mathbf{u}|)^2 J_m(l sin(\theta))^2. \tag{4.33}$$

In this formula, θ is the angle between the baseline vector and the z axis. In figure 4.7, I show a plot illustrating the l dependence of this expression. This plot is for a baseline with $2\pi|\mathbf{u}| = 65$. Note that sensitivity peaks around $l \approx 65$ and drops sharply for $l > 65$.

In general, we can say that for $l >> 1$, a baseline will be most sensitive to l-modes around $l \approx 2\pi|\mathbf{u}|$ and will be not at all sensitive to modes with $l > 2\pi|\mathbf{u}|$. Similarly, we can say that such a baseline will be sensitive only to $m < |l sin(\theta)|$. Supposing our z-axis points towards the north pole, which is standard for radio interferometers, we will find that baselines oriented more towards the E-W axis will be more sensitive to higher m values than N-S baselines.

Thus, we see how correlating pairs of antennas can be used to probe small angular scales without needing to build one enormous antenna. Now that the importance of the exponential phase term has been illustrated, let's return to considering our visibility. From data, we can form an estimator in the same way as in equation 5.59:

$$\hat{\mathcal{V}}_{ij} = \frac{1}{M}\sum_{A=1}^{M} x_i[\nu_b, t_A]x_j^*[\nu_b, t_A]. \tag{4.34}$$

In this formula, $x_{i,j} = \sqrt{\alpha_{i,j}}v_{i,j}$. Then, assuming adequately large time windows, we find that our estimator becomes

$$\langle\hat{\mathcal{V}}_{ij}\ [\nu_b]\rangle \approx \int \sum_{cd} B_{ij}^{cd} T_{cd}d\Omega + \sqrt{\alpha_i\alpha_j}\langle n_i n_j^*\rangle. \tag{4.35}$$

Since n describes receiver noise, we should find that n_i and n_j are uncorrelated, meaning that $\langle n_i n_j^*\rangle = 0$. Eliminating the noise term gives

$$\langle\hat{\mathcal{V}}_{ij}\ [\nu_b]\rangle \approx \int \sum_{cd} B_{ij}^{cd} T_{cd}d\Omega. \tag{4.36}$$

Thus, we find that taking an interferometric approach helps eliminate the pesky noise pedestal that one would see in a single dish approach, which is shown in equation 4.24. I think it's important to note here that this result was reached assuming no correlation between n_i and n_j. In truth, this will not quite be the case. This is because a real interferometer will include some coupling between antennas. I will describe this in more detail later on in the thesis.

Despite eliminating the noise pedestal, noise will still contribute to the variance of $\mathcal{V}_{ij}$ as

$$var(\mathcal{V}_{ij}) = \frac{T_{i,sys}T_{j,sys}}{M}. \tag{4.37}$$

Thus, we see that using an interferometer allows one to probe small angular scales (and thus large $k_\perp$) without needing to build enormous and expensive antenna elements. We have also seen that using cross correlations between antennas allows one to eliminate, or at least minimize, the noise pedestal seen in single dish telescopes. These two factors are the main reasons why most HIM experiments are interferometer-based rather than single dish. However, as we will see later during the systematics sections, there is a lot more left to discuss when comparing interferometers to single dishes!

4.5 Signal Processing and Calibration

In the previous two sections, I discussed everything in terms of the voltages at the receivers of the antennas, multiplied by some normalization factor. In practice, there is in fact much more signal processing involved. To illustrate this, I show the signal chain used in the dish array of the Tianlai telescope in figure 4.8. As one can see, this signal chain involves cabling, amplification, bandpass filters and down-conversion of frequencies via mixing. As such, one needs to account for all these effects in some way if they want to relate their data to the sky signal. This is typically done through what is referred to as "direction independent calibration". Roughly speaking, the usual approach in this technique is to assume that the effect of all this signal processing takes the form of a multiplicative function, i.e.

$$\mathcal{V}_{ij}^{meas}(\nu_b; t_A) = g_i(\nu_b, t_A) g_j^*(\nu_b, t_A) \mathcal{V}_{ij}^{true}(\nu_b) + n_{ij}(\nu_b, t_A). \tag{4.38}$$

This equation is saying that the measured visibility is equal to a term proportional to the true visibility plus some contribution from receiver noise. The functions g_i and g_j are referred to as the "direction independent gains" of antennas i and j. These gains encode the effect of the signal chain on the visibility. I've included time dependence in the gains g_i since they may of course drift over time. In the calibration process, one attempts to estimate the values of the $g_i(\nu)$. One approach to this is done by estimating $\mathcal{V}_{ij}^{true}$ using models of the sky and of the antenna beams. One then estimates the g_i by comparing their measured visibilities to the modeled visibilities. Another approach to calibration is through what is called redundant calibration. In redundant calibration, one assumes that each antenna has an identical beam pattern. As such, V_{ij}^{true} should be the same for all pairs of antennas with equal separation vector. This reduces the number of parameters to be estimated, allowing for an estimation of the gains without need for modeling the sky. The issue with this approach, of course, is that the antennas will most likely not all have the same beam pattern.

4.6 A spherical harmonic point of view

Before continuing, I'd like to highlight a nice trick that makes analysis of visibilities easier. This trick comes from (Shaw et al. (2015)) and allows us to obtain an expression for the visibilities without needing to use any integrals. I'll spare the details, but the core of the idea is that we express both the beam B_{ij}^{cd} and the sky T_{cd} in terms of sums of spherical harmonics. We then find that the visibility is given by

$$V_{ij} = \sum_{lm} \left[B_{ij,lm}^{T} a_{lm}^{T} + B_{ij,lm}^{E} a_{lm}^{E} + B_{ij,lm}^{B} a_{lm}^{B} + B_{ij,lm}^{V} a_{lm}^{V} \right] + n_{ij}. \tag{4.39}$$

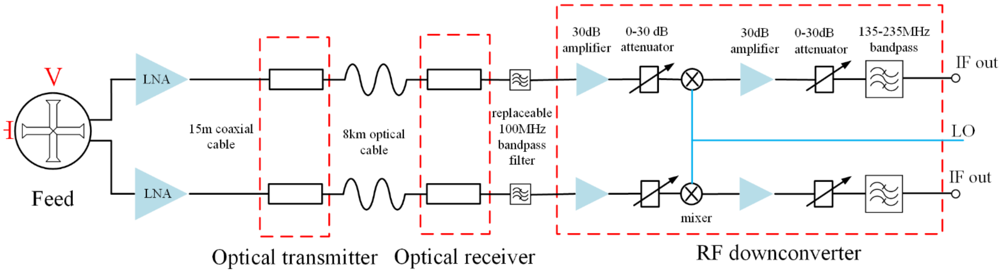

Figure 4.8: Signal chain of Tianlai instrument borrowed from Wu et al. (2021). This image shows two identical signal chains, one for each polarization of an antenna in the Tianlai dish array. The IF (intermediate frequency produced by a frequency downconverter) signal will be sent to an Analog to Digital Converter (ADC). The digitized signals are then correlated to form visibilities.

In this formula, T, E, V, B correspond to different polarization components of the sky. T corresponds to the total intensity, E and B correspond to linear polarizations and V encodes circular polarization. In the context of HIM, the HI signal will be contained in T. On the other hand, foregrounds will be concentrated in the T,B and E and components. For a particular baseline, we need to only sum up to $l_{max} = 2\pi|\mathbf{u}|$ and $|m_{max}| = l_{max} sin(\theta)$. Since we now have a sum rather than an integral, we can express this as a matrix equation

$$V_{ij} = \mathbf{B}_{ij}\mathbf{a} + n_{ij}. \tag{4.40}$$

In this equation, $\mathbf{a}$ is a column vector containing the spherical harmonic coefficients of the sky. Similarly, $\mathbf{B}_{ij}$ is a row vector containing the $B_{lm}^{T,E,B,V}$ coefficients. This equation is for a single baseline at a single frequency channel. However, there is no reason we can't repackage things to obtain a single matrix equation for multiple baselines at multiple frequencies.

As we will see later, having a matrix equation rather than an integral will make map making and power spectrum estimation much simpler.

Many HIM telescopes perform “drift scanning", in which the telescope pointing remains fixed as the sky passes overhead. The other option is to use a “tracking" scheme where the telescope pointing is adjusted to continuously observe a fixed patch of sky as the Earth rotates. For drift scan instruments, this matrix approach comes with another nice trick. In particular, we'll be able decompose our visibility into uncorrelated Fourier space components. To see this, we can begin by rewriting our equation as one that is a function of azimuthal angle during drift scanning. We obtain

$$V_{ij}(\phi) = \sum_{lm} \left[B_{ij,lm}^{T}(\phi) a_{lm}^{T} + B_{ij,lm}^{E}(\phi) a_{lm}^{E} + B_{ij,lm}^{B}(\phi) a_{lm}^{B} \right] + n_{ij}(\phi). \tag{4.41}$$

Then, we can define an “m-mode" as a Fourier mode of $V_{ij}(\phi)$

$$V_{ij,m} = \int V_{ij}(\phi) e^{-im\phi} \frac{d\phi}{2\pi}. \tag{4.42}$$

For one of our beam coefficients, it should be the case that $B_{ij,lm}(\phi) = B_{ij,lm}(0)e^{im\phi}$. As such, we may write

$$V_m = \sum_{\ell} \left[B^T_{ij,lm} a^T_{lm} + B^E_{ij,lm} a^E_{lm} + B^B_{ij,lm} a^B_{lm} + n_{ij,lm} \right]. \tag{4.43}$$

As will be seen next, having this m-mode formalism will make our job easier.

4.7 Map Making

Now that I’ve reviewed the measurement process, I will begin describing how we can relate our measured visibilities to what is going on in the sky. One way to do this is through the map making process. In this section, I’ll mostly be following along with the discussion provided in Shaw et al. (2015). The standard way to produce a map from measured data is through maximizing the likelihood. In this case, we say that

$$\mathcal{L}(\mathbf{a};\mathbf{V}) \propto exp(-\frac{1}{2}(\mathbf{V} - \mathbf{Ba})^{\dagger}\mathbf{N}^{-1}(\mathbf{V} - \mathbf{Ba})). \tag{4.44}$$

In this formula, **N** is the covariance of the noise, **V** is a collection of all the visibilities we have measured and **B** is a matrix containing all of the beam spherical harmonic coefficients. This will include all baselines and all frequency channels.

Next, suppose that we whiten our data before attempting to generate a map. In this case, we scale our visibilities by a factor of $\mathbf{N}^{-1/2}$. We will denote these whitened variables as

$$\tilde{\mathbf{V}} = \mathbf{N}^{-1/2}\mathbf{V} \tag{4.45}$$

and

$$\tilde{\mathbf{B}} = \mathbf{N}^{-1/2}\mathbf{B}. \tag{4.46}$$

When using our whitened visibilities, this likelihood would become

$$\mathcal{L}(\mathbf{a};\tilde{\mathbf{V}}) \propto exp(-\frac{1}{2}|\tilde{\mathbf{V}} - \tilde{\mathbf{B}}\mathbf{a}|^2). \tag{4.47}$$

In this case, we find that the likelihood is maximized by the estimate

$$\hat{\mathbf{a}} = \tilde{\mathbf{B}}^{+}\tilde{\mathbf{V}}. \tag{4.48}$$

In this formula, the + denotes a Moore-Penrose inverse.

In practice, the approach described in equation 4.48 is not the smartest to use. For one, since we are using all of the data available, the matrices involved become quite large. In the work of Shaw, this computational burden is reduced by generating maps one m-mode at a time. This can be done without any loss of information since $\langle \mathbf{V}_m \mathbf{V}_{m'} \rangle \neq 0$ only when $m = m'$. Another issue with this Moore-Penrose aproach is the fact that the presence of small amounts of noise in the visibilities will lead to enormous amounts of noise in the resulting maps. This is due to the fact certain baselines will have very poor signal to noise for certain m-modes. However, this noise may be reduced to reasonable levels through the use of a Wiener filter. In the Wiener filter approach, one includes a prior on what the sky signal should look like. In particular, the likelihood function will now take the form

$$\mathcal{L}(\mathbf{a};\mathbf{V}) \propto exp(-\frac{1}{2}(\mathbf{V}-\mathbf{Ba})^{\dagger}\mathbf{N}^{-1}(\mathbf{V}-\mathbf{Ba}))exp(-\frac{1}{2}\mathbf{a}^{\dagger}\mathbf{C}^{-1}\mathbf{a}). \tag{4.49}$$

In this formula, $\mathbf{C}$ is an estimate for the covariance of the spherical harmonic coefficients of the sky signal. One then obtains a Wiener-filtered estimate

$$\hat{\mathbf{a}}_{wiener} = (\mathbf{C}^{-1} + \mathbf{B}^{\dagger}\mathbf{N}^{-1}\mathbf{B})^{-1}\mathbf{B}^{\dagger}\mathbf{N}^{-1}\mathbf{V}. \tag{4.50}$$

Roughly speaking, the use of the Wiener-filter will suppress modes in the data for which there is poor signal to noise. This in turn solves the noise-related instability mentioned earlier.

4.8 Power Spectrum Estimation

Now, I've provided some idea for how visibilities are generated and how they can be converted into maps of the sky. I'd like to now devote some time to discussing how one might estimate the power spectrum from either visibilities or maps.

Estimating Power Spectra from Delay-Transformed Visibilities

In Parsons et al. (2012b), the authors provide a nice method for estimating the power spectrum from visibilities, one baseline at a time. It should be noted that this method is used by the HERA collaboration when estimating power spectra. Suppose from measurements that one obtains the visibility from baseline $\mathbf{b}$ $V(\mathbf{b}, \nu)$. They then take what is referred to as a "delay transform" of the data by taking a Fourier transform with respect to frequency

$$\tilde{V}(\mathbf{b}, \tau) \equiv \int \gamma(\nu) V(\mathbf{b}, \nu) e^{-2\pi i \nu \tau} d\nu. \tag{4.51}$$

In this formula, τ represents “delay", which has units of time and is the Fourier dual to spectral frequency. We also have a frequency space window function $\gamma(\nu)$. Such a window function is usually used to smoothly taper the visibility to 0 at the edges of the band. This prevents the presence of ringing in delay space. In my work, I usually use the square of a Blackman-Harris filter for γ.

I will spare the details of how it is derived, but the claim of the delay spectrum approach is that the power spectrum should be proportional to the absolute value squared of the delay spectrum. The k-mode probed is described via the equations

$$\mathbf{k}_\perp = \frac{2\pi\nu_0}{r_0 c}\mathbf{b} \tag{4.52}$$

and

$$k_{||} = \frac{2\pi\tau}{\alpha}. \tag{4.53}$$

In the first formula, $\mathbf{k}_\perp$ is the component of the k-mode that is perpendicular to the line of sight. This makes some intuitive sense, as a longer baseline will probe smaller scales and thus larger $\mathbf{k}_\perp$. ν_0 is the median frequency of observation and r_0 is the median comoving radius of the band.

Equation 4.53 gives us a way of relating the component of $\mathbf{k}$ parallel to the line of sight with the delay. Once again, this makes some intuitive sense. ν is related to redshift, telling us how far away the source is. Therefore, the Fourier dual to ν should tell us about modes in the radial direction. The constant α is estimated based on z_0, the median redshift of the survey along with $H(z_0)$.

Assuming there is negligible contribution at delay τ from polarized emission, we can say that

$$\hat{P}(\mathbf{k}_\perp, k_{||}) = \frac{r_0^2 \alpha |\tilde{V}(\mathbf{b}, \tau)|^2}{\int \gamma(\nu) B^T(\hat{n}, \nu) d\nu d\Omega}. \tag{4.54}$$

In this formula, B^T (rather than E, B or V) is the unpolarized beam of the baseline.

This delay spectrum approach carries with it a few advantages. For one, we can estimate the power spectrum one pair of antennas at a time. This reduces the accuracy with which we need to estimate the direction-independent gains g_i. Supposing we were to estimate the power spectrum in a way that required correlations between antenna pairs, then we would have to accurately understand the phases of the g_i. However, estimating the power spectrum one pair of antennas at time means that we only need to concern ourselves with the magnitudes of the g_i, since phases will be eliminated when we take the absolute value squared. Moreover, this method is easy to do. All we have to do is take the absolute value and square it. This requires very little effort and is not computationally taxing.

However, this delay spectrum approach comes with a few drawbacks. For one, the fact that we perform our estimations one baseline at a time means that we may be missing information about the 21 *cm* power spectrum that could be found in the correlations between different baselines. This would in turn mean that our errors are larger than what would be optimal. Moreover, this method requires that we use the flat-sky approximation, which would not provide accurate results for telescopes with large FOV such as CHIME.

Lastly, in order to achieve the robustness against calibration errors mentioned earlier, we must form a power spectrum estimate at each individual time step. The reason for this is that the $g_i(\nu_b, t)$ may drift with time, so averaging visibilities before forming a power spectrum estimate will once again make one vulnerable to phase drifts in the g_i. Forming power spectra before averaging comes with the cost of poorer scaling of noise with respect to time. What we described here can be called “squaring then averaging". If we square then average, we will find the noise in our power spectrum estimate scaling like $1/\sqrt{N_t}$, where N_t is the number of time samples. On the other hand, if we had “averaged then squared", we would find noise scaling like $1/N_t$. Methods have been proposed for how to keep the $1/N_t$ scaling while doing a square then average approach, but it is not clear how well they will work in practice.

A Simple Power Spectrum Estimate From Maps

Next, I’ll present another simple method one can use for estimating the power spectrum, this time in map-space. This method is commonly seen in papers where data analysis methods are tested on small patches of the sky, such as Hothi et al. (2020) and Chapman et al. (2012a). In this approach, we recall the definition of the 21 *cm* power spectrum

$$\langle \tilde{T}(\mathbf{k})\tilde{T}^*(\mathbf{k}')\rangle = 2\pi\delta(\mathbf{k}-\mathbf{k}')P_{21}(\mathbf{k}). \tag{4.55}$$

In a 21 *cm* experiment we do not observe the entire universe, but rather some subvolume of it. In this case, the $2\pi\delta$ gets replaced by a factor of observing volume V. We can then make the approximation

$$P_{21}(\mathbf{k}) \approx \frac{\langle|\tilde{T}(\mathbf{k})|^2\rangle}{V}. \tag{4.56}$$

In Hothi et al. (2020), the authors estimate the power spectrum simply by dividing $|\tilde{T}(\mathbf{k})|^2$ by V. In Chapman et al. (2012a), they seem to take an average over all modes within some window in **k**-space.

In any case, this sort of approach is easiest to use when one is considering a small patch of the sky. However, this becomes trickier when dealing with maps covering large solid angles. This is because pixels in this case will not be uniformly spaced in position-space and thus the results of the DFT from real space to k-space will be more difficult to interpret. The paper Liu et al. (2016) provides a method that does not require the flat-sky approximation. However, this method requires the assumption that the 21 *cm* signal is isotropic, which is not true due to redshift space distortions.

Quadratic Estimators

In the last two sections, I described some simple approaches to power spectrum estimation that one sees in the literature. These approaches are handy both since they are simple and don't require any apriori information about the HI signal.

However, we can make this discussion more general by introducing the quadratic estimator formalism. In this section, I will provide a summary of this approach. The quadratic estimator point of view can be used to describe the previously mentioned techniques, as well as others. The quadratic estimator approach hinges on the fact that any estimate of the 21 *cm* power spectrum will come from squaring the data.

In this approach we suppose that the 21 *cm* power spectrum can be described as a sum of functions

$$P(\mathbf{k}) = \sum_\alpha p_\alpha P_\alpha(\mathbf{k}). \tag{4.57}$$

In this formula, the p_α are referred to as "bandpowers" and the P_α are referred to as "band functions". In Datta et al. (2007), it is shown that the angular power spectrum of 21 *cm* radiation is linearly related (through an integral) to the rectilinear power spectrum P_{21}. Moreover, we showed in the previous discussion that our visibilities (or maps) are themselves linearly related (through matrix multiplication) to the spherical harmonics of the sky. As such, we can relate the covariance of our data (whether in map space or visibility space) to the power spectrum via

$$\mathbf{C} = \mathbf{C}_0 + \sum_\alpha p_\alpha \mathbf{C}_\alpha. \tag{4.58}$$

In this equation, p_α are the same bandpowers shown in equation 4.57 and $\mathbf{C}$ is the covariance of the data, which is related to the square of the data. I also introduce

$$\mathbf{C}_\alpha = \frac{\partial C}{\partial p_\alpha}. \tag{4.59}$$

Lastly, equation 4.58 contains $\mathbf{C}_0$, which is the portion of the covariance not due to the 21 *cm* signal. This can include non-HI contributions from the sky and also thermal noise from the receiver. What is important, however, is that we have found a useful way to relate the 21 *cm* power spectrum to the square of our data.

Another important ingredient in this discussion is the Fisher information matrix. Abstractly, this is defined by taking derivatives of the likelihood with respect to our parameters

$$F_{\alpha\beta} \equiv -\langle \frac{\partial}{\partial p_\alpha} \frac{\partial}{\partial p_\beta} log(\mathcal{L}(\mathbf{p};\mathbf{x})) \rangle_{\mathbf{x}}. \tag{4.60}$$

In this equation, **x** is a vector containing the data, which can be in map space, visibility space or whatever form the analyst likes as long as equation 4.58 holds. For this particular case, we really will have elements of the Fisher matrix being given by

$$F_{\alpha\beta} = Tr\left[\mathbf{C}^{-1}\mathbf{C}_\alpha \mathbf{C}^{-1}\mathbf{C}_\beta\right]. \tag{4.61}$$

I will spare details of the derivation, but we will arrive at an estimate of the power spectrum through an intermediate variable q. This variable q is introduced for the sake of notational convenience. In particular, we have

$$\hat{q}_\alpha = \frac{1}{2}\mathbf{x}^\dagger \mathbf{C}^{-1}\mathbf{C}_\alpha \mathbf{C}^{-1}\mathbf{x}. \tag{4.62}$$

This form of q_α will result in an estimator with minimum variance. Let's now package our p_α and q_α into vectors where the elements are various bandpowers. We then form an estimate of **p** as some mixture of the $\hat{q}_\alpha$:

$$\hat{\mathbf{p}} = \mathbf{M}(\hat{\mathbf{q}} - \hat{\mathbf{b}}). \tag{4.63}$$

In this formula, we have introduced the variable $\hat{\mathbf{b}}$. This term is an estimate of contributions from, among other things, receiver noise and non-HI components of the sky. By subtracting $\hat{\mathbf{b}}$ we are attempting to remove these contributions from our power spectrum estimate. We also introduce the mixing matrix **M** which determines the linear combination of q_α used to form an estimate of band power p_β. There are different choices one may make for **M**. If one chooses the mixing matrix to be the inverse of the Fisher matrix ($\mathbf{M} = \mathbf{F}^{-1}$), they will obtain an unbiased estimator with $\langle\hat{\mathbf{p}}\rangle = \mathbf{p}$. It should be noted that this estimator also leads to correlated errors between the bandpower estimates. We could also choose **M** to be diagonal, which would minimize the risk of encountering numerical issues in taking any sort of inverse. Moreover, a diagonal **M** created by summing over rows of the Fisher matrix provides a minimum variance estimator. We could also choose $\mathbf{M} = \mathbf{F}^{-1/2}$, which would provide us with uncorrelated errors in the power spectrum estimates.

This method also conveniently allows us obtain error estimates as

$$Cov(\hat{p}_\alpha, \hat{p}_\beta) = \sum_{cd} M_{ac}M_{bd}F_{ab}. \tag{4.64}$$

I'd like to stress the following features of the quadratic estimator formalism:

1. The quadratic estimator formalism requires one to model the covariance of their data, along with the relationship between the power spectrum and the covariance. This relationship can be estimated by first relating $\mathbf{C}$ to the angular power spectrum C_l. One can then relate C_l to the power spectrum $P(\mathbf{k})$ (Shaw et al. (2015)).
2. Our ability to pick a mixing matrix $\mathbf{M}$ gives us freedom in how the estimate is made.
3. We must have some model for non-signal components in the data in order to remove the bias term $\mathbf{b}$.
4. This approach may not work if the linear relationship between $\mathbf{C}$ and the power spectrum breaks down. This sort of thing can happen with certain data analysis techniques as we will see later.

5 SYSTEMATICS IN HI INTENSITY MAPPING AND ATTEMPTS AT SOLUTIONS

These last few chapters have supplied some idea of how HI intensity mapping works and the scientific promise of it. Now, it is time to face reality.

The HIM technique comes with it several systematics. For one, there of course exist other sources of radiation in the sky at the wavelengths of interest. Such non-HI sources are referred to as “foregrounds". There also exist emissions from radio sources on earth, which we refer to as “Radio Frequency Interference" (RFI).

One must also contend with instrumental effects. One challenging issue occurs in interferometers, where one must contend with interactions between the antenna elements. We refer to such interactions as “mutual coupling". There also exist other possible sources of systematics such as reflections in cabling reported by the HERA collaboration and radio emission from the sun. Independently, these systematics each cause problems for HIM instruments. Moreover, there will also exist some interaction between mutual coupling and foregrounds. In particular, the presence of mutual coupling would make foreground contamination more severe.

In this thesis, I will focus mainly on foreground contamination and on mutual coupling. I intend to focus on these issues since I believe they are the most challenging and important in the HIM context.

This chapter is laid out as follows. In section 5.1, I will describe the sources and properties of the foregrounds. This will be followed by section 5.2, where I describe attempted solutions to the foreground issue. I then give a similar treatment to coupling, describing the problem in section 5.3 followed by a summary of attempted solutions in section 5.4. I then conclude this chapter by discussing the scientific consequences of these effects in section 5.5.

5.1 Foregrounds

Source and Properties of the Foregrounds

The foreground emission seen in HIM instruments can be classified into several groups. We have

1. Galactic and extra-Galactic synchrotron radiation
2. Free-free emission, both Galactic and extra-Galactic
3. Point sources, which can also be Galactic or extra-Galactic.

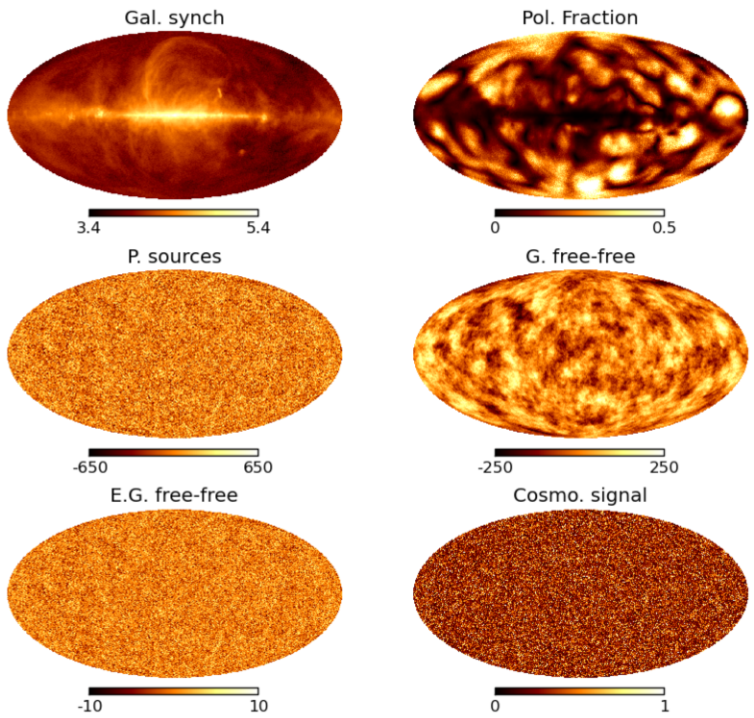

Figure 5.1: Simulated maps using the CRIME simulation package (Alonso et al. (2014)). The color bars are in units of mK, except for the Galactic synchrotron maps which shows $log_{10}(T)$ and the polarization fraction, respectively. These maps correspond to $\nu = 565MHz$.

Galactic synchrotron is caused by cosmic-ray particles undergoing acceleration by the magnetic field of the Milky Way's interstellar medium. Free-free emission refers to bremsstrahlung radiation emitted when charged particles scatter off each other. The point sources term is radiation coming from distant bright objects. One example of this is the Cygnus A radio galaxy, which is clearly visible when performing radio observations of the northern sky.

In figure 5.1, I show several simulated maps presented in Alonso et al. (2014). Note in this image that the Galactic synchrotron temperatures are displayed on a log scale. As one can see, the foreground sources are all significantly brighter than the HI signal, with the Galactic synchrotron emission being the brightest of the group. Such synchrotron radiation is brightest in the Galactic plane, being up to five orders of magnitude brighter than the 21 cm signal.

Luckily, these foreground contributions tend to be spectrally smooth compared to the HI emission. As I will discuss in more detail later, this smoothness is crucial for the avoidance or removal of foregrounds. I'll illustrate why these emissions are spectrally smooth by considering Galactic synchrotron emission. First, consider a charged particle moving in a circle with period T. If we pick some point in space away from the particle's orbit, we will find the electric field varying with period T. As such, we can see that the frequency of synchrotron radiation is tied to the frequency of the

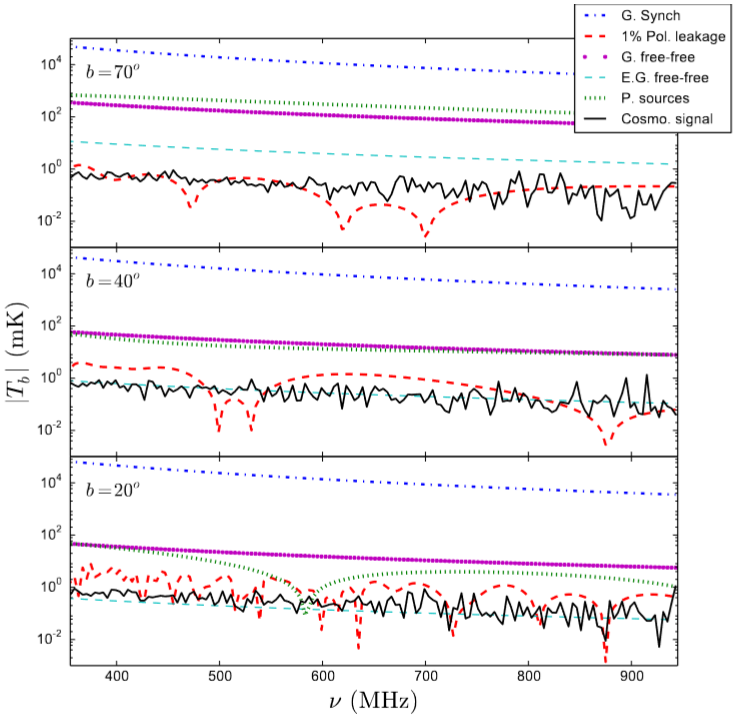

Figure 5.2: Simulated values of brightness temperature as a function of frequency. The variable b represents galactic latitude. This plot is borrowed from Alonso et al. (2014).

particle's orbit, and thus the velocity of the particle. In the case of Galactic synchrotron radiation, the cosmic-ray particles causing the radiation have a smooth velocity distribution, implying a smooth spectrum for the radiation. In figure 5.2, I show a plot from Alonso et al. (2014) displaying simulated values of temperature as a function of frequency for various foreground contributions. The angles given in this figure denote the Galactic latitude, which is the angle between the line of sight and the Galactic plane. In this plot, one can see that all contributions are spectrally smooth, aside from the polarized Galactic synchrotron emission. Moreover, this chromaticity becomes worse as one approaches the Galactic plane. I'll now discuss what causes this additional chromaticity.

The additional chromaticity seen in the polarized Galactic foregrounds is caused by the presence of Faraday rotation in the Milky Way. As a wave travels through the Milky Way, we will find that its polarization gets rotated by an amount $\beta = \psi\lambda^2$. The variable ψ is referred to as the "rotation measure" and described via

$$\psi \propto \int_0^d n_e(s) B_{||}(s) ds. \tag{5.1}$$

In this formula, n_e is the number density of free electrons and $B_{||}$ is the component of the Milky Way's magnetic field that is in the direction of propagation. The variable s gives the radial distance from Earth. We can then model the polarized part of the Galactic foregrounds as an integral over regions of various Faraday depths

$$I_p(\nu, \hat{n}) = \int d\psi k(\psi, \hat{n}, \nu) e^{i\psi(c/\nu)^2}. \tag{5.2}$$

In this formula, $k(\psi, \hat{n}, \nu)$ represents the total emission from all regions with rotation measure ψ (phase of polarization rotation) along line of sight $\hat{n}$. What is important to note here is that the amount of Faraday rotation a wave undergoes depends both on the rotation measure and the frequency of the wave. Thus, the effect of the Faraday rotation may look quite different depending on the spectral frequency channel considered. As of now, $\psi(\hat{n}, s)$ is not well understood. Typically, models of polarized Galactic foregrounds assume that the number of regions with rotation measure ψ follow a Gaussian distribution

$$N(\psi(\hat{n})) \propto exp(-\frac{1}{2}(\frac{\psi}{\sigma(\hat{n})})^2). \tag{5.3}$$

The chromaticity observed will thus increase as $\sigma(\hat{n})$ increases. One would expect $\sigma(\hat{n})$ to be largest for lines of sight close to the Galactic plane, since the Milky Way's magnetic field is strongest there.

One can also glean some information from this about how chromaticity will vary with frequency. In particular, the fact that the Faraday-rotation terms depends on $(c/\nu)^2$ means that the brightness will vary more quickly as a function of frequency at higher redshift.

In summary, we know the following about the foregrounds present in a HIM experiment:

1. Foregrounds in the HIM context can come from many sources, with Galactic synchrotron emission being the brightest contributor, barring the small number of bright point sources. These emissions are orders of magnitude brighter than the HI signal.

2. The unpolarized component of these foregrounds is expected to be spectrally smooth, as opposed to the HI which varies quickly with respect to frequency.

3. Polarized Galactic foregrounds will have a more complicated frequency dependence due to Faraday rotation in the Milkyway. This chromaticity will become more severe as one approaches the Galactic plane and at lower frequencies.

4. As of now, polarized foreground emission is not well understood. However, there is research being done on this topic (Hu and Lazarian (2023)).

Foregrounds in an Interferometer - Introducing "The Wedge"

It should be noted that there is an additional complication when doing HIM using an interferometer. In particular, we will find that the interferometer imparts its own chromaticity on the measured foregrounds. Let's consider this now.

To keep things simple, I will ignore polarization and noise. Keep in mind, however, that the effect I discuss here will also apply to the polarized foregrounds. In this zero noise unpolarized case, we find that our visibility is

$$\mathcal{V}_{ij}(\nu) = \int B(\hat{n},\nu)T(\hat{n},\nu)e^{-2\pi i\nu c^{-1}\mathbf{r}_{ij}\cdot\hat{n}}d\Omega. \tag{5.4}$$

In this equation, B is an appropriately normalized (giving antenna temperature units) beam pattern for the antennas used to form the visibility.

If we take the delay transform of our visibility, we find

$$\tilde{\mathcal{V}}_{ij}(\tau) = \int \gamma(\nu)B(\hat{n},\nu)T(\hat{n},\nu)e^{-2\pi i\nu(\tau+\frac{\mathbf{r}_{ij}}{c}\cdot\hat{n})}d\nu d\Omega. \tag{5.5}$$

Let's say $B_\gamma = \gamma(\nu)B(\nu)$. We can apply the shift and convolution theorems to say that

$$\tilde{\mathcal{V}}_{ij}(\tau) = \int \tilde{B}_\gamma(\hat{n},\tau')\tilde{T}(\hat{n},\tau+\frac{\mathbf{r}_{ij}}{c}\cdot\hat{n}-\tau')d\tau' d\Omega. \tag{5.6}$$

Let's assume that the sky emission is constant with respect to frequency, giving $\tilde{T}(\hat{n},\tau) = T_0(\hat{n})\delta(\tau)$. This is a good approximation for the smooth and simple frequency dependence expected from unpolarized foregrounds. In such a case, we would find

$$\tilde{\mathcal{V}}_{ij}(\tau) = \int \tilde{B}_\gamma(\hat{n},\tau+\frac{\mathbf{r}_{ij}}{c}\cdot\hat{n})T_0(\hat{n})d\Omega. \tag{5.7}$$

This already gives us an interesting result; a spectrum of sky emission with zero width in delay-space can lead to a visibility with significant delay-space width. We can see why this is the case and learn a bit more by examining the equation further. Let's imagine that $\tilde{B}_\gamma$ has a small width in delay-space. In such a case, we would find the integrand of equation 5.7 sharply peaked when $\frac{\mathbf{r}_{ij}}{c}\cdot\hat{n} = -\tau$. This result tells us that as τ increases, $\tilde{V}_{ij}(\tau)$ will be telling more about the sky at lines of sight closer to the horizon, where $\mathbf{r}_{ij}$ becomes parallel to $\hat{n}$. Note that this also implies that our visibility will have a non-zero width in delay-space. Moreover, we find that τ satisfying $\tau c > |\mathbf{r}_{ij}|$ should not be contaminated by foregrounds! This result is the much discussed "foreground wedge" one finds in the literature (Datta et al. (2010)) (Parsons et al. (2012b)) (Thyagarajan et al. (2013)). I provide an illustration of the wedge in figure 5.3. In this figure, delays below the blue line will be dominated by foregrounds. Since $k_{||} \propto \tau$, this means that the foreground contamination will

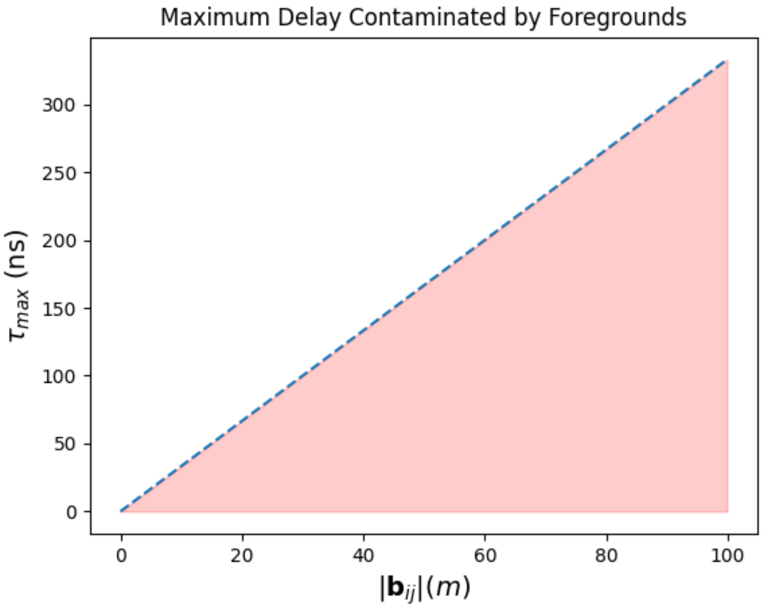

Figure 5.3: An illustration of the wedge for baselines up to $100\,m$.

be dominant for power spectrum modes with low $k_{||}$, unless one is able to clean them effectively. Moreover, this contamination should become worse as baselines get longer.

Next, let's consider the case where $\tilde{B}_\gamma$ has considerable width in delay-space. In such a case, equation 5.7 would imply that the foregrounds "leak" beyond the wedge. One rule of thumb for characterizing this leakage is to look at how long it takes $\tilde{B}_\gamma$ evaluated at zenith to drop by a factor of 10^5 (about the ratio of foreground temperature to signal temperature) from its value at $\tau = 0$. Supposing it takes $30\,ns$ for this to happen. Then, we would say that modes within $30\,ns$ of the wedge will also be significantly corrupted by the foregrounds. Due to this, HIM experiments usually try to ensure that their beam patterns are spectrally smooth, which is to say narrow in delay space, minimizing leakage.

Now, let's consider a case where the foregrounds also have significant chromaticity. In this case, we need to step back to equation 5.6. Really, this equation shows that at each line of sight, we convolve $\tilde{B}_\gamma$ and $\tilde{T}$. Thus, as one would expect, the delay-space width of foregrounds in visibilities depends both on the foregrounds themselves and on the frequency behavior of the beams.

To illustrate the conclusions of this discussion, I show a simulated delay space visibility in figure 5.4. This visibility was generated using maps simulated using foreground maps from CRIME, HI maps from the Cosmology in the Radio Band (CORA) package (Shaw et al. (2015)) and beams modeled after the CHIME experiment. These beams have constant FWHM as a function of frequency, meaning there is very little chromaticity added by the instrument. As one can see, unpolarized foregrounds are contained at low delays while the HI signal is much more spread out. We also have

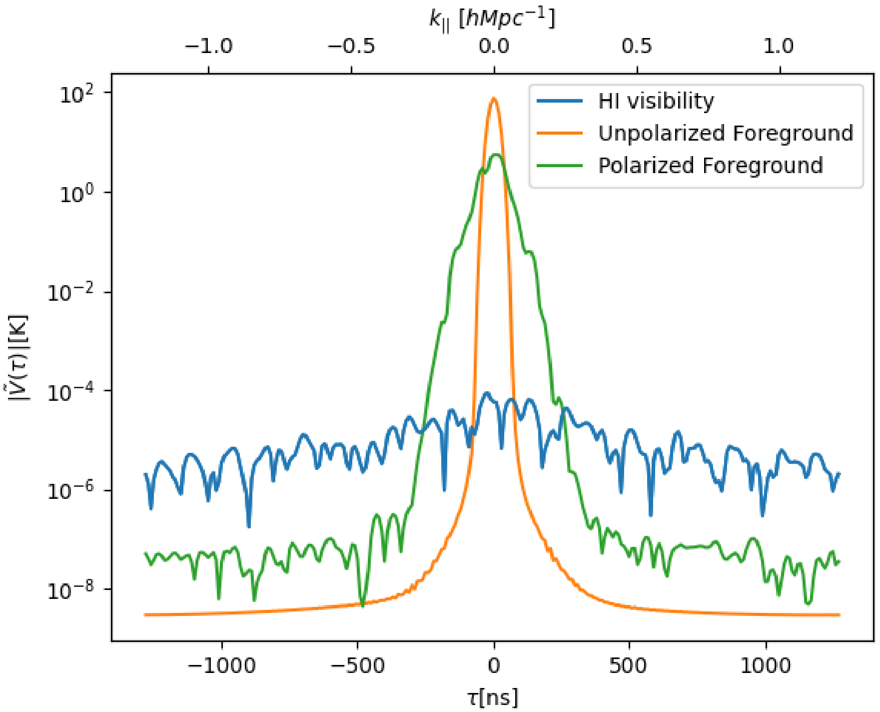

Figure 5.4: Simulated visibilities shown in delay space. The top axis shows the corresponding values of $k_{||}$. Note that the relationship between τ and $k_{||}$ will change as a function of frequency. The simulated foreground maps came from CRIME while the simulated HI map came from CORA. The visibilities shown here were calculated at 256 evenly spaced frequencies from 400 to 500 MHz. These visibilities assume CHIME-like beams with constant FWHM.

polarized foregrounds, which have a larger delay-space width than the unpolarized foregrounds. In this set of simulations, polarized foregrounds are dominant at scales $k_{||} < 0.2hMpc^{-1}$ for lines of sight away from the Galactic plane. This significantly worsens close to the Galactic plane.

5.2 Foreground Handling, Both Analysis and Hardware

Hardware Approach - Antenna Design With Foregrounds in Mind

The first line of defence against foregrounds can come in the design of the antenna elements used in one's interferometer. The first and most obvious thing to do is to design antennas with beam patterns $B(\hat{n}, \nu)$ that are spectrally smooth. This is a consequence of equation 5.6. Since the delay-space visibility is the convolution of beam $\tilde{B}$ and sky signal $\tilde{T}$, we can minimize the footprint of the foregrounds by using smooth beams. Aside from my own paper (presented in chapter 6), one can take a look at the HERA Vivaldi feed antenna (Fagnoni et al. (2021)) as an example of an antenna optimized to minimize chromaticity.

The second aspect to consider is, as mentioned in the foreground section at the beginning of this chapter, we are expecting for the polarized foregrounds to have more complicated spectral-structure, and thus a larger footprint in delay-space. As such, it would be beneficial to design the antenna elements in such a way that we can minimize the leakage of polarized foregrounds into the data. There are two factors we must consider in this process.

First, let's suppose that each antenna element in the array is dual-polarized. Let's label these polarizations X and Y. We will now try to remove polarized foreground leakage from the data by introducing the “Stokes I" visibility. We define the Stokes I visibility for antennas i and j as

$$V_{ij}^{I} \equiv \frac{1}{2}(V_{ij}^{XX} + V_{ij}^{YY}). \tag{5.8}$$

Let's make the assumption that antennas i and j have identical beam patterns and that the X and Y polarizations have orthogonal effective lengths at all directions in the sky. In other words, $\mathbf{l}^{X} \cdot \mathbf{l}^{Y} = 0$ for all possible $\hat{n}$. In such a case, we would find that our stokes I visibility schematically looks like

$$V_{ij} \sim \int exp(...) * ((|l_x|^2 + |l_y|^2)T + (|l_x|^2 - |l_y|^2)Q)d\Omega \tag{5.9}$$

In order to remove the contribution of the Q polarization, we would need for $|l_x|^2 = |l_y|^2$ at all locations. Thus, in order to make the Q polarized contribution go away we'll need $|l_x|^2 = |l_y|^2$ everywhere. In a typical dual-polarized design such as the Tianlai feed antenna, it should be the case that the y polarized part of the antenna is just the x polarized part rotated by $\pi/2$. As such, the effective lengths look something like $|l_x(\theta, \phi)| = |l_y((\theta, \phi + \pi/2)|$. Thus, we can minimize polarization leakage by minimizing $|l_x(\theta, \phi)| - |l_x(\theta, \phi - \pi/2)|$.

Thus, one can minimize polarization leakage in their data by optimizing for the following features in their beams:

- Minimize $|\mathbf{l}_x \cdot \mathbf{l}_y|$. This is often referred to as “cross polarization".
- Minimize variation in beam patterns from element to element.
- Minimize $|l(\theta, \phi)| - |l(\theta, \phi - \pi/2)|$.

Typically, antenna designers will make sure that the cross-polarization term is low. However, I'm not sure of any examples in the context of HI that were designed with the azimuthal symmetry in mind. I will say that some HERA papers do estimate this parameter for their antennas. In any case, the works of Moore et al. (2013) and Moore et al. (2017) estimate that polarization leakage will be on the percent level for HIM instruments. This value will likely be greater for cylindrical reflectors, which have very little azimuthal symmetry by design.

Thus, we find that designing antenna elements with smooth spectral properties, low cross-polarization and high azimuthal symmetry should help lessen the presence of foregrounds in the data.

Avoiding and Removing Foregrounds During the Analysis Stage

Let's now consider the removal or avoidance of foregrounds in one's data. I will begin by describing how these various techniques work and illustrating the common features found amongst them. I will also include a note on estimating the power spectrum once the data has been cleaned. Then, I will provide a review of literature that shows tests of these various methods.

When cleaning foregrounds from data, one usually begins by expressing their data using the equation

$$\mathbf{X} = \mathbf{h} + \mathbf{f} + \mathbf{n}. \tag{5.10}$$

In this formula, the terms on the right hand side of the equation represent the HI signal, foregrounds and noise, respectively. The terms in this equation are all matrices and can have different shapes depending on the way the analyst would like to package things. For this section, let's assume that **X** represents a map and has shape $n_{ch} \times n_{pix}$ where n_{ch} is the number of frequency channels and n_{pix} is the number of pixels. It should be noted that one could construct the matrix **X** out of visibility data, though removing foregrounds in visibility-space is not seen as commonly in the literature. The standard approach one finds is to assume that the foregrounds can be described via

$$\mathbf{f} = \mathbf{A}\mathbf{S}. \tag{5.11}$$

In this formula, **S** is a $n_t \times n_{pix}$ matrix of "templates". We use n_t to denote the number of templates used in the model. We multiply this by a "mixing matrix" **A**, which has shape $n_{ch} \times n_t$. Let's consider column j of the map, which we'll denote by a vector $\mathbf{x}_j$. This vector will contain all frequency channels for line of sight j. Let $\mathbf{a}_k$ be column k of **A**. We will find that

$$\mathbf{x}_j = \sum_k \mathbf{a}_k S_{kj} \tag{5.12}$$

where the S_{kj} are matrix elements of **S**. Thus, we are modelling our foregrounds at each pixel as some linear combination of frequency space functions. These frequency-space functions are provided by the mixing matrix **A** while the coefficients are provided by the template matrix **S**. Naturally, one can then remove the foregrounds by subtracting **AS** from their data.

It should be noted that interpretation of $\mathbf{a}_k$ will change depending on the method used. For example, in the case of NKL, $\mathbf{a}_k$ could be several functions of frequency concatenated, one for each partition of the map used in the analysis. In the case of the KL approach in Shaw et al. (2015), the $\mathbf{a}_k$ will be functions living in the space of the image of the beam matrix.

This template plus mixing matrix type of approach describes most of the foreground cleaning methods that one finds in the literature. Essentially, each method fits **A** and **S** to the data, subject to some regularization. The term “regularization” is used to refer to any process intended to prevent overfitting. As we will see, the main difference between many of these methods is the regularization chosen.

I would like to take a moment to note that the number of templates to use is up to the discretion of the analyst. Oftentimes, one reads papers where tests are conducted on simulated data. Usually, the authors of these papers will use different values of n_t until they get the best results possible. This approach would be unrealistic when considering real data. One option for picking n_t is the Akaike Information Criterion (Olivari et al. (2015)) (Akaike (1974)). Another is to continue increasing the number of templates used until the results begin to converge (Zhang et al. (2016)). Note that for the Karhunen-Loève transform, one obtains n_t automatically by picking a desired minimum signal to foreground ratio.

Now that I have provided the necessary background, I will begin reviewing some of the most common approaches for dealing with foregrounds in the HIM context.

Avoidance

This is the simplest approach one can take when dealing with foregrounds. Recall that when taking the delay transform of a visibility ($\tilde{V}$), we should find that the foregrounds are confined to low τ. As such, assuming a reasonably achromatic beam, one can simply avoid their foregrounds by filtering out smaller values of τ. Of course, taking this approach means that one misses large scale line-of-sight information, which is important for observing the BAO.

Principal Component Analysis

The simplest approach to cleaning (and not avoiding) foregrounds is Principal Component Analysis (PCA) (Liu and Shaw (2020)). Let’s suppose that our data **X** has zero mean (just remove the mean if it doesn’t). In PCA, one forms **A** and **S** from the eigenvectors of the covariance matrix of their data. Supposing that **C** (dimensions of $n_{ch} \times n_{ch}$ for our map example) is the covariance, we can say

$$\mathbf{S} = \mathbf{P}_S \mathbf{X} \tag{5.13}$$

and

$$\mathbf{A} = \mathbf{P}_S^{\dagger}. \tag{5.14}$$

In these equations, $\mathbf{P}_S$ is a $n_s \times n_{ch}$ matrix containing the first n_s eigenvectors of $\mathbf{C}$. The assumption here is that the first n_s eigenmodes will contain all the foregrounds and hopefully very little signal.

Viewed differently, it can be shown (Ghojogh et al. (2023)) that PCA is solving the optimization problem

$$max(Tr\left[\mathbf{P}_S\mathbf{C}\mathbf{P}_S^\dagger\right]) \tag{5.15}$$

subject to

$$\mathbf{P}_s\mathbf{P}_s^\dagger = \mathbf{I}. \tag{5.16}$$

In other words, PCA seeks a mixing matrix plus template combination that captures as much variation in the data as possible.

In practice, one does not really know $\mathbf{C}$ and instead forms an estimate from the data as $\hat{\mathbf{C}} \propto \mathbf{X}\mathbf{X}^\dagger$. We then clean the data as

$$\mathbf{X}_{clean} = (\mathbf{I} - \mathbf{P}_s^\dagger\mathbf{P}_s)\mathbf{X}. \tag{5.17}$$

fastICA

Let's consider another possibility, the fast Independent Component Analysis (fastICA) method. In this approach, we ignore the presence of noise or HI in the data, giving us

$$\mathbf{X} = \mathbf{A}\mathbf{S}. \tag{5.18}$$

Moreover, note that in fastICA one begins by whitening their data, providing an identity matrix as the covariance in that basis. For the rest of this fastICA section, anytime I mention $\mathbf{X}$ I will be referring to the whitened version of $\mathbf{X}$. In fastICA, we try to figure out the mixing matrix and templates first by assuming that there exists a matrix $\mathbf{W}$ such that

$$\mathbf{W}\mathbf{X} = \mathbf{S}. \tag{5.19}$$

The idea behind fastICA is to assume that each template, which is to say each row of $\mathbf{S}$, is statistically independent. This is taken into account when solving for the $\mathbf{W}$ matrix. Suppose that $\hat{\mathbf{w}}_i$ is a guess for row i of the $\mathbf{W}$ matrix. From this guess we calculate the variable

$$\mathbf{y} = \hat{\mathbf{w}}_i\mathbf{X}. \tag{5.20}$$

If our $\hat{\mathbf{w}}_i$ is a correct guess, then $\mathbf{y}$ will be one of the templates. If the guess is incorrect, then $\mathbf{y}$ will be a combination of templates. By the central limit theorem, we know that the sum of random variables will converge to a Gaussian distribution. Thus, we assume that the correct guess is the one that maximizes the non-Gaussianity of $\mathbf{y}$.

In ICA, one uses negentropy as a proxy for non-Gaussianity. We define the negentropy of $\mathbf{y}$ as

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}). \tag{5.21}$$

In this formula, $\mathbf{y}_{gauss}$ is a Gaussian variable with the same covariance as $\mathbf{y}$ and the entropy H is given by

$$H(\mathbf{y}) = -\sum_j P(\mathbf{y} = \mathbf{a}_j) log(P(\mathbf{y} = \mathbf{a}_j)) \tag{5.22}$$

where the sum over j is covering all possible values that the vector $\mathbf{y}$ could have. Thus, we can see how a larger value of $|J|$ would indicate that the statistics of $\mathbf{y}$ are further from Gaussian.

The trick here is to approximate the negentropy of $\mathbf{y}$ as

$$J(\mathbf{y}) \approx \left[E\left[g(\mathbf{y})\right] - E\left[g(\nu_G)\right]\right]^2. \tag{5.23}$$

In this formula, ν_G is a Gaussian variable with unit variance, g is some function that is not quadratic and $E\left[\right]$ is used to denote an expectation value.

The fastICA algorithm can be described nicely as an optimization problem as was shown in Marins et al. (2022). With the constraint that $\mathbf{W}$ is orthogonal, the optimization problem becomes

$$\mathbf{W} = \underset{\mathbf{W}}{argmax}\{J(\mathbf{WX}) - \lambda||\mathbf{W}^T\mathbf{W} - \mathbf{I}||\}. \tag{5.24}$$

In this formula, λ is some non-negative number and $|| \cdot ||$ represents the norm of the matrix. The authors of Marins et al. (2022) do not state the norm they chose, but I see no reason why one couldn't choose to use the Frobenius norm, which is defined as

$$||\mathbf{D}||_F \equiv \sqrt{\sum_{ij} |D_{ij}|^2}. \tag{5.25}$$

After solving this optimization problem, we can say

$$\mathbf{S} = \mathbf{WX} \tag{5.26}$$

and

$$\mathbf{A} = \mathbf{W}^T. \tag{5.27}$$

Thus, ICA seeks to find templates with maximal negentropy, with negentropy being a proxy for statistical independence. This is different than PCA, where the goal is simply to capture as much variation in the data as possible.

GMCA

Next, let's consider Generalized Morphological Component Analysis (GMCA). Put simply, GMCA fits templates to the data under the assumption that the templates should be sparse in some wavelet representation. I'll now discuss what this means in more detail. To begin, let's consider a single template $\mathbf{s}_j$ which is a vector of length n_{pix}. We can then describe this template as a linear combination of "wavelets"

$$\mathbf{s}_i = \sum_{j=1}^{T} \alpha_{ij} \boldsymbol{\phi}_j. \tag{5.28}$$

In this formula, the α_{ij} are coefficients and the $\boldsymbol{\phi}_i$ are the wavelet functions. Roughly speaking, wavelets are wave-like functions that have finite width both in the frequency and time domains (or whatever Fourier dual pair one is working with). There exist many possible wavelets, but the most common form one sees being used for GMCA in the HIM context is the "starlet" (Murtagh et al. (2011)). Having finite width in both domains sets wavelets apart from the Fourier transform, where the basis functions will have infinite spread in one domain and a Dirac delta function in the other.

Wavelets are often of interest because they often allow for "sparse" representations of functions. In the case of equation 5.28, we would say that $\mathbf{s}_i$ has a sparse representation in the wavelets ϕ_j if it has few non-zero α_j.

For the full template matrix $\mathbf{S}$, we can say

$$\mathbf{S} = \boldsymbol{\alpha}\boldsymbol{\Phi}. \tag{5.29}$$

In this formula, $\boldsymbol{\alpha}$ is a $n_t \times T$ matrix containing the wavelet coefficients for each template. In this case, the variable T is the number of wavelet functions used. The variable $\boldsymbol{\Phi}$ is a $T \times n_{pix}$ matrix containing the wavelet functions to be used.

In Marins et al. (2022), they nicely cast GMCA as an optimization problem. That is

$$\{\hat{A}, \hat{\alpha}\} = \underset{A\alpha}{argmin}\{||\mathbf{X} - \mathbf{A}\boldsymbol{\alpha}\boldsymbol{\Phi}||_{F,C_N} + 2\lambda \sum_{i=1}^{n_t} ||\boldsymbol{\alpha}_i||_{\ell_1}\}. \tag{5.30}$$

In this formula, the first term tries to match templates to the data while the second term promotes sparsity of the components. In the first term, the norm we use includes the noise covariance C_N. In particular, we say

$$||\mathbf{D}||_{F,C_N} = \sqrt{Tr\left[\mathbf{D}\mathbf{C}_N^{-1}\mathbf{D}^T\right]}. \tag{5.31}$$

On the sparsity term, we use what is called the l_1 norm. This is given by

$$||\boldsymbol{\alpha}_i||_{\ell_1} = \sum_j |\alpha_{ij}|. \tag{5.32}$$

It should be noted here that ideally we would promote sparsity by counting up the number of non-zero components. However, optimizing for this is difficult to do; we instead use the ℓ_1 as it is easier to deal with.

Thus, we see how GMCA searches for templates that are sparse in the wavelet representation.

GPR

Next, let's consider Gaussian Process Regression (GPR). In GPR, one models their data as a "Gaussian process". In the previous examples, the data samples have been vectors with n_{ch} frequency channels. In such a discrete case, one has some mean vector $\boldsymbol{\mu}$ and some covariance matrix $\mathbf{C}$. In a Gaussian process, we upgrade frequency to a continuous variable. We replace the mean vector with a "mean function" $m(\nu)$ and replace the covariance with a kernel $K(\nu, \nu')$. The idea then is that the mean and covariance of our discrete data comes from evaluating K and m at the appropriate frequency values.

In the context of HIM, one assumes that the kernel has contributions from foregrounds, noise and the 21 cm signal:

$$K = K_{21} + K_{fg} + K_N. \tag{5.33}$$

Naturally, one takes K_N to be a diagonal matrix. On the other hand, one will find "Matern kernels" being used to model K_{21} and K_{fg}. I will spare the details, but Matern kernels are defined via three parameters, σ, ℓ and η. We find that σ sets the standard deviation, ℓ provides the length scale of frequency correlations and η is a parameter determining smoothness of the data. Roughly speaking, one picks K_{21} and K_{fg} by maximizing the posterior of the kernel parameters given the input data. This process is a bit more involved than I'm saying here; one can read Soares et al. (2021) for more details.

In any case, once the kernels have been chosen, one can use their model to estimate the expectation of the foregrounds. It can be shown that

$$E\left[\mathbf{f}\right] = \mathbf{K_{fg}}\left[\mathbf{K}_{21} + \mathbf{K}_{fg} + \mathbf{K}_N\right]^{-1}\mathbf{X}. \tag{5.34}$$

Interestingly, this method also provides one with an uncertainty on their foreground estimate

$$cov(\mathbf{f}) = K_{fg} - K_{fg}\left[\mathbf{K}_{21} + \mathbf{K}_{fg} + \mathbf{K}_N\right]^{-1} K_{fg}. \tag{5.35}$$

One can then clean their data simply by subtracting the foreground estimate

$$\mathbf{X}_{clean} = \mathbf{X} - E[\mathbf{f}]. \tag{5.36}$$

KL

Lastly, let's consider the idea of a Karhunen-Loève transform. To begin, let's note that since signal, noise and foregrounds are uncorrelated, we can say that the covariance of our data is given by

$$\mathbf{C} = \mathbf{C}_{21} + \mathbf{C}_{fg} + \mathbf{C}_{N}. \tag{5.37}$$

Now, suppose that we set up what is called a "generalized eigenvalue problem" using our signal and foreground covariances. In particular, we have

$$C_{fg}\mathbf{\Phi} = C_{HI}\mathbf{\Phi}\mathbf{\Lambda}. \tag{5.38}$$

In this formula, $\mathbf{\Phi}$ is a matrix of "generalized eigenvectors" and $\mathbf{\Lambda}$ is a diagonal matrix of "generalized eigenvalues".

It can be shown (Ghojogh et al. (2023)) that the solutions to this generalized eigenvalue equation also are solutions to the optimization problem

$$\underset{\mathbf{\Phi}}{max} Tr(\mathbf{\Phi}^{\dagger}\mathbf{C}_{fg}\mathbf{\Phi}) \tag{5.39}$$

subject to

$$\mathbf{\Phi}^{\dagger}C_{HI}\mathbf{\Phi} = \mathbf{I}. \tag{5.40}$$

Thus, this generalized eigenvalue problem seeks modes that maximize the ratio of foreground to signal. Moreover, the eigenvalues λ_i will tell us about the expected ratio of foreground to signal power for eigenmode ϕ_i.

Note that the act of sending our data from frequency space to the eigenmode space is what we refer to as a "Karhunen Loève Transform".

Assuming the means have already been removed from the data, we can create the template matrix $\mathbf{S}$ as

$$\mathbf{S} = \mathbf{P}_s\mathbf{X}. \tag{5.41}$$

In the case of cleaning maps, the matrix $\mathbf{P}_s$ is of shape $n_t \times n_{ch}$ and contains the n_t eigenmodes with the largest eigenvalues.

Unlike the PCA case, the eigenmodes found in this manner might not be orthogonal to each other. As such, we must return to frequency space from KL space in a different manner than we would in the PCA case. Suppose that $\mathbf{P}$ is a matrix with rows equal to the eigenvectors. We will return to frequency space using $\mathbf{P}_{-s}$, which is of shape $n_{ch} \times n_t$ and contains the columns of $\mathbf{P}^{-1}$ corresponding to our n_t eigenvectors of interest. Thus, we end up with mixing matrix

$$\mathbf{A} = P_{-s}. \tag{5.42}$$

Estimating the Power Spectrum After Cleaning

When cleaning foregrounds, one will inevitably encounter some amount of signal loss. In the language of templates and matrices, we could say that $\mathbf{AS}$ will inevitably contain some signal along with the foregrounds.

Let's consider the foreground avoidance approach of Parsons et al. (2012a). When signal loss is present, one will find that the estimated power spectrum carries with it significant bias, at least for the modes for which the signal loss is appreciable. Moreover, one may even find that the foreground removal process takes HI from higher τ and brings it over to lower τ, giving the appearance of lower signal loss than is really present.

One way to deal with such losses is to use the transfer function approach of Cunnington et al. (2022). In such an approach, one adds a simulated HI signal to their data and observes the signal loss that occurs when the data is cleaned. This then provides a multiplicative factor one can use to compensate for the signal loss. I'm somewhat skeptical of these methods however, as they may be able to adjust the overall level but probably not recover the shape of the power spectrum.

Another approach is to use the previously mentioned quadratic estimator formalism. If the quadratic estimators are appropriately used, one can account for the effect of the cleaning on their data. In this case, signal loss will not lead to bias but rather increased error bars in the power spectrum estimate. However, one needs to be careful as the use of foreground cleaning methods could break the linearity requirement of the quadratic estimator formalism.

Let's take PCA as an example. Suppose that the PCA cleaning provides one with

$$\mathbf{X}_{clean} = \mathbf{RX} \tag{5.43}$$

where $\mathbf{R} = \mathbf{I} - \mathbf{AS}$. The covariance of the cleaned data is then

$$\mathbf{C} = \langle \mathbf{RXX}^{\dagger}\mathbf{R}^{\dagger} \rangle. \tag{5.44}$$

If one's cleaning matrix $\mathbf{R}$ depends on the specifics of the data, then the linear relationship between the power spectrum and the covariance of the data is broken. Thus, we can no longer express our covariance as a linear combination of the bandpowers, which is required in the optimal quadratic estimator approach.

Thus, we see that extracting cosmological information from foreground cleaned data can be a challenge in and of itself.

A review of the literature

I'll now present a literature review that I wrote up for the paper "Needlet Karhunen-Loéve (NKL): A Method For Cleaning Foregrounds From 21 cm Intensity Maps", which is given in chapter 7. Here it is:

Over the years, many foreground removal methods have been proposed. A helpful review of many of these may be found in Liu and Shaw (2020). Methods that require no priors on the HI, noise or foregrounds are described as "blind", and those requiring priors are described as "non-blind". One will also find methods being tested both on raw visibilities or on maps synthesized from data. However, unlike blind and non-blind, there is not a clear distinction here. Some methods may be applicable both in map-space and in visibility-space.

Principal Component Analysis (PCA), Independent Component Analysis (ICA) (Chapman et al. (2012b)) and Generalized Morphological Component Analysis (GMCA) (Chapman et al. (2012a)) are the prominent blind methods that one finds being applied to maps. In the literature, one can also find non-blind methods being applied in map-space. Examples include Generalized Needlet Internal Linear Combination (GNILC) (Olivari et al. (2015)), Gaussianity Process Regression (GPR) (Mertens et al. (2018)) and the "semi-blind" Singular Value Decomposition (SVD) method proposed in Zuo et al. (2023). On the other hand, tests of cleaning methods on visibilities are less commonly found in the literature. One blind method that can be used in visibility space is "foreground avoidance". In this method, one exploits the fact that the foregrounds tend to be confined to a region of certain region k-space, referred to as "the wedge" (Ewall-Wice et al. (2016)). Power spectrum modes within this region of k-space are then excluded from the analysis. Non-blind methods have also been tested in visibility space. The beam projection plus Karhunen-Loève transform proposed in Shaw et al. (2015) and the GPR method used in Soares et al. (2021) and Mertens et al. (2018) are examples we have found in the literature.

However, for the case of post-EOR surveys, such avoidance methods come at the cost of losing valuable information about Baryon Acoustic Oscillations (BAO). The BAO are located roughly in the wavenumber range $0.03 \leq k \leq 0.4\,\mathrm{hMpc}^{-1}$ (Bull et al. (2015)). Much information about large length scales such as these would be lost when taking a foreground avoidance approach.

So far, when real data is considered, the foregrounds have usually been handled in a conservative way. For instance, in Paul et al. (2023), the first detection of HI using IM without cross-correlating with galaxies, the analysts chose to use foreground avoidance rather than removal. Foreground avoidance was also used in Amiri et al. (2023). In this work, the authors cross-correlated data from the CHIME telescope with galaxies and quasars from eBOSS. In this work, it was found that the foreground avoidance method excluded length scales important for characterizing the BAO. The authors Wolz et al. (2021) present results from cross-correlating GBT data with the eBOSS survey. In this case, the authors opted to use FastICA paired with a transfer function to compensate for signal loss. In Cunnington et al. (2022), the authors cross-correlated maps from MeerKAT with WiggleZ galaxies. The authors of this paper cleaned their data with PCA and used a transfer function to compensate for signal loss. The authors of Anderson et al. (2018) cross-correlated maps from the Parkes telescope with galaxies from the 2dF galaxy survey. These authors correlated maps from different seasons to reduce noise bias, and applied a transfer function to compensate for signal loss.

One can also find numerous papers in the literature testing these various methods on simulated data. Such tests are usually conducted at lower redshifts ($z \lesssim 0.6$) and take into account beam effects by convolving with a Gaussian profile of appropriate width. In Carucci et al. (2020), tests were conducted using simulations that assumed a telescope similar to MeerKAT operating in single dish mode at low redshifts ($0.09 \leq z \leq 0.58$) and surveying the full sky. This test included contributions from polarized foregrounds simulated using the Cosmological Realizations for Intensity Mapping Experiments (CRIME) software package (Alonso et al. (2014)). In this instance, GMCA recovered the angular power spectrum of the HI plus noise with errors of 5% down to $\ell \approx 25$. Moreover, they found that GMCA provided lower errors than ICA (Carucci et al. (2020)). In Olivari et al. (2015), GNILC was tested at low redshifts ($0.13 \leq z \leq 0.48$) in a survey covering half of the sky. These simulated maps included no polarized foregrounds. In this case, GNILC was found to recover the angular power spectrum of the HI to within 6% error down to $\ell = 30$. So, like GMCA, GNILC also seemed to struggle at large angular scales. Recently, GNILC, GMCA and ICA were all tested on simulations of the BINGO experiment (Marins et al. (2022)). In this test, all of the techniques were found to provide statistically equivalent results. In Soares et al. (2021), GPR was tested on simulated maps at low redshifts ($0.18 < z < 0.58$). The simulated maps included polarized foregrounds generated using CRIME. These tests were conducted on Stripe82 and a $3000\,\mathrm{deg}^2$ region at the South Celestial Pole (SCP). In these tests, GPR and PCA provided similar performance, recovering the power spectrum with errors around 10% for all spatial scales considered.

A notable cleaning method applied in visibility-space is the combined beam projection and Karhunen-Loève (KL) transform proposed and tested in Shaw et al. (2015). This test assumed a simplified version of the CHIME instrument and was conducted at higher redshifts ($1.84 < z < 2.55$) where polarized foregrounds are more severe. The test used simulated maps produced by the Cosmology in the Radio Band (CORA) software package. This package makes different assumptions than CRIME, resulting in less severe chromaticity in the polarized foregrounds. In this paper, instrumental effects were accounted for in a more realistic way, by generating visibilities from maps using simulated beams from cylinder telescopes. This method was found to effectively recover the HI power spectrum down to $k \approx 0.02\,\mathrm{hMpc}^{-1}$. This result appears quite promising, since the BAO would be recovered at all length scales. However, the weakness of this approach is that it requires a highly accurate beam model along with priors on the unpolarized foregrounds, HI and noise. It was found that this approach breaks down when main beam uncertainties exceed 0.1%. In addition, amplifier gains must be known to better than 1% within each minute. Such accurate beam and gain calibration pose a significant challenge and may not be possible in practice.

In summary, published map-space tests of cleaning methods on simulated polarized foregrounds have only been conducted at low redshifts. Moreover, these papers usually take beam effects into account in a simple way. The only exception we found is in the work of Hothi et al. (2020), who performed tests on maps generated from visibilities. Such tests are informative, but not the whole story since these polarized foregrounds will become more severe as redshift increases. On the other hand, Shaw's method was found to work extremely well when faced with polarized foregrounds at higher redshifts, but required precise knowledge of the beam. It is reasonable to imagine that the previously described methods used in map-space might be more robust to beam mis-calibration than Shaw's method. There are two reasons for this. One is that the beam projection part of beam projection/KL approach removes the polarized foregrounds by projecting onto the null-space of the polarized beam matrix. This step may not work when the beams are not well understood. Another reason is that the KL part of the SVD/KL cleaning requires one to have priors on how the unpolarized foregrounds will contribute to the total visibility. On the other hand, the map-space applicable methods described earlier in this paper (aside from the "semi-blind" SVD method from Zuo et al. (2023)) either do not require a foreground model, or, in the case of GPR, estimate one from the data.

Although they may end up being more robust to calibration issues than the visibility-space beam projection/KL method, the available methods usable in map-space are missing certain strengths of the beam projection/KL approach. One particular strength of beam projection/KL is that it takes full advantage of the available priors. In particular, during the KL step, one considers correlations between all baselines in the telescope. In this way, one is exploiting correlations in both frequency and spherical harmonics (different baselines are sensitive to different l-modes). This is much different from commonly used map-space methods, such as PCA, GNILC, ICA, and GPR, which, as employed so far in the literature, only consider correlations along the line of sight. Although there are some exceptions, such as the commonly used GMCA method does incorporate angular correlations by enforcing sparsity of components in the wavelet domain. Another exception to this is the "semi-blind" SVD method. This method provides excellent recovery of the HI signal, but, like Shaw's method, suffers from requiring priors on the foregrounds present in the maps.

However, one must also keep in mind that working in map-space presents certain advantages not available in visibility space. For one, map-space allows for specific pixels to be selected. As such, one can mask out pixels with particularly strong foregrounds or artifacts. Moreover, one can perform a spherical-harmonic or spherical wavelet transform of the data, allowing for precise separation by angular scale. Such operations are not possible in visibility space. The closest thing one could do is separate the visibilities into m-modes, where the m refers to the azimuthal m found in spherical harmonics (Shaw et al. (2015)). Such freedom could be useful when cleaning foregrounds from maps, due to their dependence on line of sight direction and angular scale.

A comparison of techniques

Now, I will present a comparison of various popular foreground removal approaches that I wrote up for “Needlet Karhunen-Loéve (NKL): A Method For Cleaning Foregrounds From 21 cm Intensity Maps”, which is presented in chapter 7:

Before actually testing any techniques, it will be beneficial to discuss differences between the techniques considered in this paper. In particular, we will discuss these differences and try to provide some motivation for why certain techniques may provide better performance than others.

First, one should note that GNILC requires a model for the HI and noise, making it a “non-blind" technique. This is different from GMCA, PCA and ICA, where the only free parameter provided by the user is the number of templates to use.

Another aspect of these techniques to consider is locality. In the standard version of GMCA, the matrices **A** and **S** are meant to capture the foregrounds at all pixels and angular scales (Carucci et al. (2020)). PCA and ICA are typically conducted in a similar way, where one estimates the covariance **C** using the entire dataset. On the other hand, GNILC cleans the maps in a more fine-grained way. In particular, it divides the data into needlet coefficients, and then cleans the data one coefficient at a time. This allows for the foregrounds to be cleaned differently depending on the location and angular scale in question.

One can imagine reasons why such a fine-grained treatment might provide advantages. For instance, the ratio of foreground to HI power is expected to vary with angular scale. In particular, large angular scales will suffer worse contamination than smaller ones. When working with unpolarized foregrounds, one would expect similar chromaticity at all lines of sight. However, when dealing with polarized foregrounds, we expect for the chromaticity to vary with line of sight. As such, one would expect for pixels close to the Galactic plane to require more templates to clean than ones far from the plane. Thus, it seems likely that a more fine-grained approach would work better when dealing with polarized foregrounds.

It is certainly possible to imagine changing PCA, ICA or GMCA to make them more local. For instance, one could implement a scale- and location- dependent version of PCA where the covariance matrix is estimated for some neighborhood around each needlet coefficient. In addition, a technique called L-GMCA has been proposed (Bobin (2017)) in which one uses different mixing matrices for different regions of the map. For now though, we will consider only the global versions of these techniques.

So, we see reasons why a more localized approach would likely provide better results when cleaning foregrounds from maps. So far, GNILC is the most local approach that has been found in the literature. In Olivari et al. (2015), this approach was tested on low-z simulated maps without any polarization leakage included (Olivari et al. (2015)). This study found GNILC outperforming PCA at angular scales ($\ell > 30$). However, GNILC provided worse performance than PCA at scales larger than that. Interestingly, Carucci et al. (2020) found that GMCA also struggled at large angular scales $\ell < 50$, at least in the case when polarized foregrounds were included.

In the next chapter, I introduce NKL. Similarly to GNILC, NKL is a non-blind technique which acts on needlet coefficients. However, while GNILC only considers statistics along the line of sight, NKL also considers angular correlations in needlet space.

5.3 Mutual Coupling

A Description of the Mutual Coupling

Let's now consider the issue of mutual coupling in the context of HIM. When reading the literature, one will find that different authors may define the term "mutual coupling" somewhat differently. In this thesis I will use it to refer to the effect of electromagnetic coupling between antenna elements in the array. As I will explain now, this coupling can take the form of beam distortions and a noise pedestal in the data. I will now discuss in detail each of these aspects along with the mark they leave on HIM data.

A Noise Pedestal Due to Mutual Coupling

One issue that can arise from mutual coupling comes in the form of coupled receiver noise. Let's imagine forming a visibility from antennas 1 and 2. Each of these antennas will of course be connected to a receiver system such as the one shown in figure 4.8. These receiver systems produce noises $n_{1,rcv}(\nu)$ and $n_{2,rcv}(\nu)$. In the absence of coupling, one would find that the noises are uncorrelated, i.e. $\langle n_{1,rcv} n^*_{2,rcv} \rangle = 0$. As such, their contribution would average to 0 when taking a visibility, just as discussed in the interferometry section.

However, when coupling is present, one would find that noise from the Low Noise Amplifier (LNA) of antenna 1 will radiate from antenna 1 then get picked up by antenna 2. This effect was discussed in detail in Kwak et al. (2024), a paper on which I am the second author. I will follow along with that discussion now. First, let's consider figure 5.5. Ports 1 and 2 correspond of course to the terminals of antennas 1 and 2. At each port we have forward and backward waves which are related in a linear fashion

$$\begin{bmatrix} b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} \begin{bmatrix} a_3 \\ a_4 \end{bmatrix}.$$

In this formula, s_{11} and s_{22} provide information about reflections from the ports. s_{12} and s_{21} carry information about mutual coupling between the antennas. It should be noted here that the convention is to define

$$a_i \equiv \frac{V_i + I_i Z_0}{\sqrt{Z_0}} \tag{5.45}$$

$$b_i \equiv \frac{V_i - I_i Z_0}{\sqrt{Z_0}}. \tag{5.46}$$

In this formula, Z_0 is a “reference impedance" which is usually chosen to be the impedance of the transmission line connected to the antenna. Note that these waves have units of square root power.

It turns out that the LNAs used on radio telescopes produce noise. Some of this noise travels “downstream”, leaving the output of the amplifier. Moreover, there will also be noise travelling “upstream”, from the amplifier towards the antenna. We model the noise generated by LNA 1 via traveling waves A_n (downstream noise) and B_n (upstream noise). Note here that for simplicity we will not consider any noise generated by LNA 2. In this case, we also assume that the LNAs have reflection coefficient Γ_1 and Γ_2. That is to say, $b_1 = \Gamma_1 a_1$ and $b_2 = \Gamma_2 a_2$. Based on what has been discussed so far, we can make the simple observation that

$$a_2 = b_4 = s_{21} a_3 + s_{22} a_4 = s_{21}(\Gamma_1 a_1 + B_n) + s_{22} a_4. \tag{5.47}$$

In other words, we are seeing that the incoming wave present at LNA 2 contains a contribution from the noise generated by LNA 1. I will skip some details in the derivation, but it can be shown that these noise correlations generate a noise visibility which, in units of power, is given by

$$V_{12,noise} = \left[\frac{S_{21}}{1 - \Gamma_2 S_{22}}\right] \left|\frac{1}{1 - \Gamma_1 S_{11}}\right|^2 (\Gamma_1 \langle |A_n|^2 \rangle + \Gamma_1 S_{11}^* \langle A_n B_n^* \rangle + \langle A_n^* B_n \rangle + S_{11}^* \langle |B_n|^2 \rangle). \tag{5.48}$$

In Meys (1978), the LNA noise power in a frequency bin of width Δf is described via 3 different temperatures:

$$\langle |B_n|^2 \rangle = k T_b \Delta f, \tag{5.49}$$

$$\langle |A_n|^2 \rangle = k T_a \Delta f, \tag{5.50}$$

and

$$\langle A_n B_n^* \rangle = k T_c e^{i\phi_c} \Delta f. \tag{5.51}$$

We can also separate the S_{11} into phase and magnitude $S_{11} = |S_{11}| e^{i\phi_s}$. In the case of the

Tianlai experiment's signal chain (shown in figure 4.8) it was noted that Γ_1 and S_{11} were both small quantities, allowing for second order terms in these variables to be discarded. In general, HIM receivers are designed to have good impedance matching, and so this approximation should work well in general. This observation leads to the approximate result

$$\langle |V_{12}|^2 \rangle \approx |S_{21}|^2 (k\Delta f)^2 T_c \mathfrak{Re}\left[T_c e^{2i\phi_c} + 2\Gamma_1 T_a e^{i\phi_c} + 2|S_{11}| T_b e^{i(\phi_c - \phi_s)}\right]. \tag{5.52}$$

Assuming that noise going from LNA 1 to LNA 2 works the same way, we can double our result to say that the noise pedestal in temperature units is given by

$$V_{12,T} = 2\sqrt{\langle |V_{12}|^2 \rangle / (k\Delta f)^2} = 2|S_{21}| T_{amp,eff}. \tag{5.53}$$

In the last step of this equation, I have have packed everything from equation 5.52 aside from $|S_{21}|$ into the $T_{amp,eff}$ term.

How large might this noise pedestal be? Well, let's consider the noise present in the Tianlai telescope. Both measurements and simulations show that antenna elements in the Tianlai array have $|S_{21}|^2 \sim 10^{-7}$. In the Kwak et al. (2024) paper, we include measurements of LNA noise in Tianlai. In particular, these measurements show that T_a, T_b and T_c are all on the order of tens of Kelvin. As such, I'll assume that $T_{amp,eff} \sim 50\,K$ for Tianlai. These are numbers to keep in mind when viewing figure 5.6, which is borrowed from Kwak et al. (2024). This estimate was generated assuming a baseline of length $10\,m$. In this case, we can see that $S_{21} \sim 10^{-3.5}$ leads to a noise pedestal several orders of magnitude larger than the measured HI signal. Thus, even relatively low levels of coupling still lead to a noise pedestal in the data that dwarfs the signal. As I will discuss later, this pedestal can be removed as long as it remains stable over time. However, this approach will not be effective for all baselines.

Beam Effects From Mutual Coupling

The second issue related to mutual coupling is the effect that it has on the antenna beam patterns. To illustrate this issue, I'll start by giving a toy example. Suppose we have a single antenna as is shown in figure 5.7, which is borrowed from Collin and Zucker (1969). In this figure, we have a single antenna contained in some volume with radius r_0. Collin and Zucker (1969) model the antenna using an S-parameter matrix as was described in the previous section on mutual coupling. In this case, we have a^+ and a^-, which describe traveling waves in the cable or waveguide that connect the receiver (Z_L) to the antenna. Moreover, we have b_i^+ and b_i^-, which are used to describe incoming and outgoing free-space waves respectively. In Collin and Zucker, the EM waves outside r_0 are described using Hankel functions, so the index i is used to label which basis function is being used.

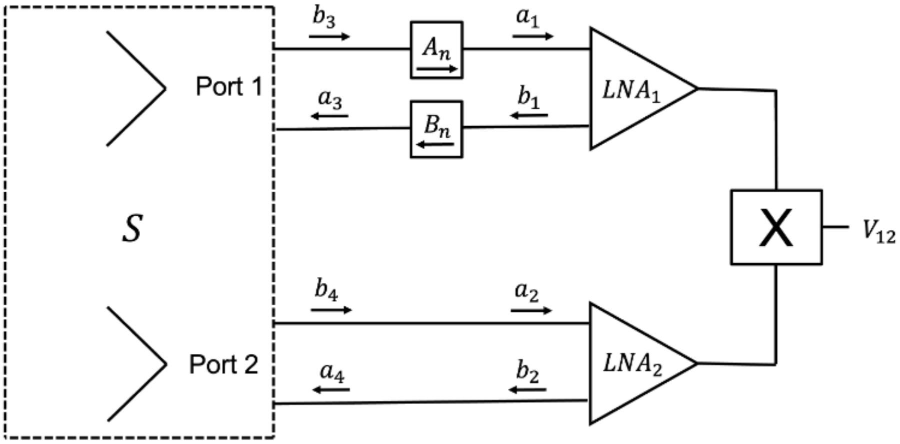

Figure 5.5: Port to Port antenna coupling.

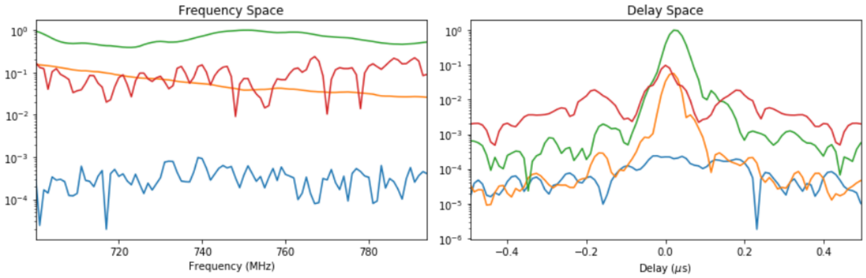

Figure 5.6: A noise pedestal estimate, for Tianlai borrowed from Kwak et al. (2024). In this figure, the vertical axis is in units of Kelvin. Moreover, blue represents the HI signal and red represents the noise pedestal. Orange and green represent contributions from polarized and unpolarized foregrounds respectively.

Let's suppose that our antenna is in receiving mode, and is excited with some incoming wave $\mathbf{b}^+$. This variable is a vector containing the coefficients of the Hankel function expansion. Such a wave could be, for instance, a planewave coming from the sky. We can get some understanding about the process that occurs by looking at the S-matrix, which I'll just include as figure 5.8. We can glean two things from this matrix. First, the index 0 is used to describe the port of the antenna, meaning that matrix elements S_{ij} for $i, j \neq 0$ tell us that the scattering will have contributions that are the same no matter the load attached to the terminals of the antenna. It is important to note that these terms are here, since some HIM mutual coupling papers do not acknowledge this contribution. The terms S_{0i} tell us about how reflections at the load will affect the scattered field.

I'll skip over the details of the derivation, but the result we get makes sense given the discussion provided. First, note that the total electric field outside the antenna can be described as

$$\mathbf{E} = \mathbf{E}_0 + \mathbf{E}_s. \tag{5.54}$$

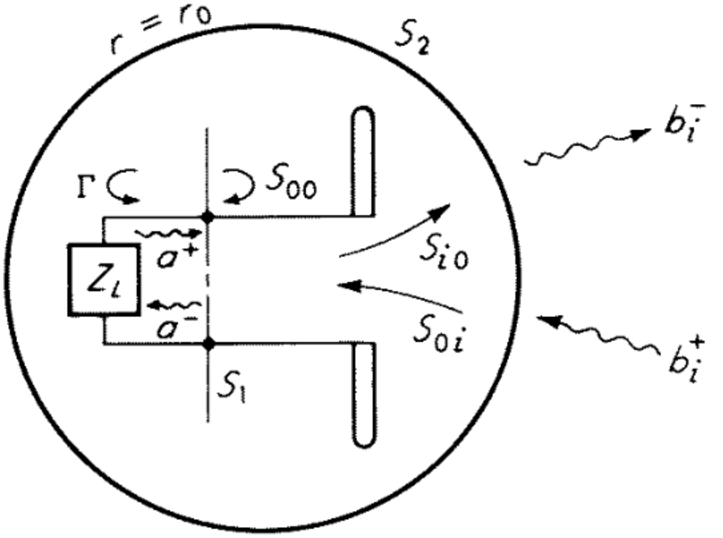

Figure 5.7: Cartoon of scattering borrowed from Collin and Zucker. S_1 represents the terminal plane in the waveguide or transmission line connecting the antenna to the load. S_2 is a reference surface with radius r_0. This cartoon doesn't seem to include the S_{ij}, but keep in mind that they do matter!

$$\begin{bmatrix} a^- \\ b_1^- \\ b_2^- \\ \cdot \\ \cdot \\ \cdot \\ b_i^- \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} S_{00} & S_{01} & S_{02} & \cdots & \\ S_{10} & S_{11} & S_{12} & \cdots & \\ S_{20} & \cdots & \cdots & \cdots & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \end{bmatrix} \begin{bmatrix} a^+ \\ b_1^+ \\ b_2^+ \\ \cdot \\ \cdot \\ \cdot \\ b_i^+ \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

Figure 5.8: Full s-matrix from Collin and Zucker.

In this formula, $\mathbf{E}_0$ is the incoming field and $\mathbf{E}_s$ is the scattered field. The scattered field can be expressed as

$$\mathbf{E}_s = \mathbf{E}_s(Z_L = 0) - \frac{I(Z_L = 0)Z_L}{Z_L + Z_A}\mathbf{E}_r. \tag{5.55}$$

In this formula, $\mathbf{E}_s(Z_L = 0)$ is the scattered field that would be present if the terminals of the antenna were short circuited. Note that this term carries all the scattering caused by the S_{ij} (for $i, j \neq 0$) parameters. As such, we know that the second term in our equation ought to contain only contributions from the S_{i0} or S_{0i} parameters. Since these parameters describe scattering from the port, we know that the second term should be proportional to the field emitted when a current is applied at the port of the antenna. This is exactly what we get, since $\mathbf{E}_r$ is defined as the field transmitted by the antenna when a unit current is applied at the terminals. It should be emphasized that $\mathbf{E}_s(Z_L = 0)$ need not have the same angular structure as the radiated field $\mathbf{E}_r$. This will complicate the issue of coupling.

Now, suppose we extrapolate this scattering idea to an array of many antennas such as Tianlai or PUMA. Suppose we take antenna i and transmit a signal from it. The electric field far away from the array will then be given by

$$\mathbf{E} = \mathbf{E}_0 + \sum_j \mathbf{E}_{s,ij}. \tag{5.56}$$

In this equation, $\mathbf{E}_{s,ij}$ is the field scattered by antenna j when an excitation is applied at antenna i. I'm using a double index here since the scattered field may be different depending on which antenna is excited. The nature of these scattered fields will be complicated, including contributions beyond first order. That is to say, scattering from antenna j will depend both upon $\mathbf{E}_0$ and the fields scattered by the other antennas. However complicated things may get, we can be sure that each of these scattered terms will include some contribution that has the same angular structure as the radiated field, plus another term that may not really look like the radiated field.

Now, let's consider why such beam effects may be an issue. For one, since scattering causes radiation to bounce between antennas, it will increase chromaticity and thus increase foreground leakage. For example, this is illustrated on simulations of the HERA array (Fagnoni et al. (2020)). In this paper, simulations were performed on a subset of the HERA array (a strip 160 m by 38 m, including 33 antennas). In this simulation, it was found that without scattering, the foregrounds should be contained within $k_{||} \approx 0.1\, hMpc^{-1}$ of the wedge. When mutual coupling effects were added, the foregrounds leaked to $k \approx 0.3\, hMpc^{-1}$ for baselines close to the center. The authors of this paper noted that the chromaticity caused by coupling becomes more severe the further the distance from the antenna of interest to the edge of the array. In the limit of a very large array, they estimate that foregrounds would leak out to $k \approx 0.5\, hMpc^{-1}$ beyond the wedge on account of the coupling. Though HERA isn't searching for the BAO, this example shows that scattering could lead to the BAO becoming inaccessible for similar arrays at lower redshift. Moreover, it was found that this chromaticity did not improve when the feeds of other antennas were removed. This result indicates that the scattering is dominated by the short circuit term rather than the term proportional to $\mathbf{E}_r$.

However, it should be noted that the HERA array was not optimized with mutual coupling in mind. On the other hand, the HIRAX collaboration has made some effort to control this coupling by using deep dishes with $f/d = 0.25$. Such deep dishes reduce coupling by having the feed antennas placed at or below the rims of the dishes, which provides shielding. In Saliwanchik et al. (2021), the authors simulated the chromaticity of their beams and found that the delay response decreases by a factor of 10^5 in about $150\,ns$. Given the fact that HIRAX operates from 400 to $800\,MHz$, this would correspond to losing modes within approximately $0.17hMpc^{-1}$ of the wedge. In this case, HIRAX would still be capable of measuring a significant portion of BAO modes, which are contained at $k \lesssim 0.4hMpc^{-1}$. Comparing this result to that of HIRAX, it seems that using a deep dish does provide one with some relief from the chromaticity caused by mutual coupling.

In addition to chromaticity, the scattering of the beams causes some additional problems. For one, scattering could lead to significant variation of antenna beam patterns. This would mean the previously mentioned redundant calibration method could be ineffective. This in turn would require one to use models of the sky and antenna beams in doing their calibration. However, this is again an issue since the Galactic foregrounds are not particularly well understood. Moreover, the inclusion of mutual coupling effects will also alter the shape of the beam. The introduction of new structure would likely make beam calibration more difficult and could also worsen polarization leakage by increasing asymmetries in the main beams. Running into trouble with calibration of beams could pose problems during the analysis stage when one needs a model of their beam in order to form maps or remove foregrounds.

In conversation, I heard some experts make another interesting point. In the presence of coupling, the effective size of the antenna becomes larger. We then run the risk of having an extremely large farfield distance for our antenna. Currently, most driftscan HI instruments measure their beams using RF transmitters mounted on drones. Such large farfield distances could become larger than the altitude limits of said drones. However, one could imagine the delay features in the beam giving one some handle on which length scales are relevant. For instance, in the case of HIRAX we see that the beam response drops by a factor of 10^5 at $150\,ns$. As such, antennas further than about $23\,m$ away probably are not relevant. Even so, this would mean the farfield distance for an array with coupling like HIRAX would be about 2 kilometers, assuming we are observing at $600\,MHz$. However, in the case of an infinite array made of HERA antennas, the relevant length scale would be more like $150\,m$. Assuming HERA observes at $150\,MHz$ this would correspond to a farfield distance of $22km$, far beyond the maximum altitude of typical drones.

5.4 Attempts at Solutions to the Coupling

Now that I've explained some of the challenges posed by mutual coupling, I think it would be beneficial to discuss some attempts that have been made at solving them.

Decreasing Coupling Through Antenna Design

The first and most obvious thing one could do is to design their antennas in such a way that the coupling is minimized. Aside from my own antenna (presented in chapter 6), this was notably tried by the HIRAX collaboration (Saliwanchik et al. (2021)). The HIRAX telescope operates over the frequency range $400 \leq \nu \leq 800\,MHz$ and consists of dipole antennas centered over $6\,m$ diameter parabolic dishes. In order to minimize coupling, the HIRAX collaboration made use of a deep dish with focal ratio $f/d = 0.25$. This quantity is the ratio of focal length to diameter for the dish. A lower value of f/d corresponds to a deeper parabola. In this case, $f/d = 0.25$ corresponds to the focus being located in a plane that contains the rim of the dish. This provides two advantages. One is that the feed is shielded by the dish, preventing direct coupling from feed to feed. Second, $f/d = 0.25$ means that the edge of the dish will be at a $90°$ angle from the peak of the feed beam pattern. As such, one will achieve a smoother and greater tapering of illumination towards the edge of the dish. Since the farfield is the Fourier transform of the dish illumination, this corresponds to lower sidelobes and thus lower coupling of dishes.

In Saliwanchik et al. (2021), the authors investigated the effects of further deepening the HIRAX dish. Their simulations showed that a deeper dish leads to larger reflections between feed and dish, but decreased reflections between dishes. These effects were characterized by looking at the delay spectra of the beams at zenith. in particular, it was found that lowering f/d from 0.25 to 0.2 led to modest improvements in the delay spectrum, but there was still noticeable chromaticity no matter what was done. On the other hand, it was found that decreasing f/d from 0.25 to 0.2 did lead to about a $20\,dB$ reduction in the S_{21} parameter. This would correspond to about a factor of 10 reduction in the noise pedestal.

The SVD Approach of HERA

One approach to solving the coupling issue is the Singular Value Decomposition (SVD) method proposed in Kern et al. (2019) in the context of HERA . To illustrate their idea, let me describe a two antenna instrument. The coupling is modeled through the antenna voltages as

$$v_1' = v_1 + \epsilon_{12} v_2 \tag{5.57}$$

and

$$v_2' = v_2 + \epsilon_{21} v_1. \tag{5.58}$$

In other words, they assume that the voltage at antenna 1, v_1', is some pristine voltage v_1 plus a copy of the voltage of antenna 2. This model is pretty good, but not quite complete. If we ignore higher order scattering, this model should account for the noise pedestal of Kwak et al. (2024) and the effects of the $\mathbf{E}_r$ term in the scattering equation. However, the previously mentioned short circuit scattering term cannot be described as simply adding a copy of the voltage from the other antenna.

In any case, this model leads to the conclusion that

$$V_{12}' = \langle v_1' v_2'^* \rangle = V_{12} + \epsilon_{12} V_{22} + \epsilon_{21}^* V_{11}^*. \tag{5.59}$$

In other words, the perturbed visibility will be the pristine visibility plus copies of the autocorrelations of antennas 1 and 2. Naturally, these autocorrelations will be shifted in delay space by an amount proportional to the distance between the antennas.

Now let's get to the actual mutual coupling removal part. This method uses an SVD and takes advantage of two facts. First, equation 5.59 tells us that the port-to-port coupling we wish to remove ought to be concentrated at values in delay space close to b_{ij}/c. Second, it ought to be the case that these port-to-port coupling contributions should vary slowly as the instrument drifts across the sky. As such, we actually describe as the visibility V_{12}' as a matrix where rows correspond to local sidereal time (LST) and columns correspond to spectral frequency. The following list of steps, quoted from Kern et al. (2019), gives the algorithm that they use:

1. Fourier transform the visibility waterfall to delay space.
2. Apply a rectangular band-stop window across delay to down-weight foregrounds at low delays.
3. Decompose the visibility via SVD.
4. Choose the first N modes to describe the systematic and truncate the rest.
5. Take the outer product of the remaining T and D modes to form N data-shaped templates.
6. Multiply each template with their corresponding singular value in S and sum them to generate the full-time and delay-dependent systematic model.

7. Fourier transform the systematic model from delay space back to frequency space and subtract it from the data.

This SVD method was found to work quite well on simulated data for east-west baselines. The port-to-port systematic was suppressed by 4 orders of magnitude for short E-W baselines and 8 orders of magnitude for longer E-W baselines. However, the SVD method unfortunately was found not to work at all for N-S baselines. The reason for this difference has to do with the rotation of the baselines. For instance, suppose we have an interferometer located at the equator. We will describe the baselines using a coordinate system (x, y, z) that remains fixed as the Earth rotates. Imagine that at $LST = 0$, an east-west baseline is along the x-axis and a north-south baseline is along the z-axis. Some time later, the north-south baseline will still be along the z-axis while the east-west baseline will have picked up a y-component. Thus, and east-west baseline will probe different k modes in the sky as the earth rotates, providing more variation with respect to LST. On the other hand, the north-south baseline will also probe the same k mode, making it harder to distinguish between the coupling systematic and the true visibility.

Modeling coupling through S-parameters.

Next, let's consider the paper by Virone et al. (2018) which investigates coupling in the the LOFAR telescope. In this paper, they seek to gain some understanding of the coupling through the S-parameters. In this formalism, the authors sought to understand the effect of receiver impedance on the antenna beams. In particular, they assume an array of N antennas and begin by saying that the electric field emitted by the array when all antennas are terminated with impedance Z_0 is given by

$$\mathbf{E} \propto \sum_{n=1}^{N} E_n^{FF}(1 + S_{nn})\sqrt{Z_0}a_n = \sum_{i=n}^{N} v_n E_n^{FF}. \tag{5.60}$$

In this equation, we have s-parameters S_{nn} and traveling waves a_n. In this context, we will use a_n to describe waves traveling towards antenna n and away from the load. We will use b_n to describe waves traveling from the antenna to the load. Lastly, we have E_n^{FF} which is defined as the electric farfield pattern produced by antenna n when supplied with a unitary voltage and when all other antennas are terminated with impedance Z_0. Phrased in the language of Collin and Zucker (1969), we are essentially saying

$$E_n^{FF} \propto E_n^0 + E_{n,s}(Z_{rcv} = Z_0), \tag{5.61}$$

Where E_n^0 is the field without coupling and $E_{n,s}$ is the scattered field.

In the case that $Z_L \neq Z_0$, one will find that scattered fields look different than in the Z_0 loaded case. Suppose now that the antennas in the array are terminated by $Z_L \neq Z_0$ and that there is an excitation a_i provided at antenna i. The authors then account for reflection off the mismatched load via

$$\mathbf{E}_n \propto \mathbf{E}_n^{FF}(1 + S_{nn})a_n + \sum_{k \neq n} \mathbf{E}_k^{FF}(1 + S_{kk})a_k^c. \tag{5.62}$$

In a nutshell, we are claiming that the electric field generated by antenna n in the Z_L loaded condition is the field as it would be in the Z_0 loaded condition plus copies of the Z_0 fields of the other antennas. In this formula, the a_k^c are related to a_i, the s-parameters, and the load reflection coefficent matrix $\mathbf{\Gamma}_L$ via

$$\mathbf{a}^c = \mathbf{\Gamma}_L(1 - \mathbf{S}^r\mathbf{\Gamma}_L)^{-1}\mathbf{S}^c a_i. \tag{5.63}$$

I won't bother with the details of this equation, but I'll note that S^r and S^c are submatrices of S. Thus, if one knows Γ_L and the S-parameters, they can undo the effects of impedance mismatches at the load.

Comparison to data showed that the model used here was fairly accurate for LOFAR, reasonably describing errors in the beam due to resonances.

Let's tie this back into the picture I gave previously. For antenna i, we know in general that the coupling is a short-circuit term plus a term proportional to the field radiated by the antenna in isolation

$$\mathbf{E}_i = \mathbf{E}_i^0 + \sum_{j \neq i} \mathbf{E}_{s,j}(Z_L = 0) + \alpha(Z_L)\mathbf{E}_{r,j} \tag{5.64}$$

In principle, we could figure out the contribution of the $\mathbf{E}_{r,j}$ by varying the load impedances and looking at the differences between the resulting $\mathbf{E}_i$. However, we will not be able to figure out the short-circuit contribution, which at least in the HERA case seems to be the most severe contribution.

One can find other papers in the literature with similar models, such as Josaitis et al. (2022) and Lui et al. (2009). However, in all cases we will not be able to obtain the short circuit scattering term.

5.5 Scientific Consequences of Mutual Coupling and Foregrounds in the Post-EOR Context

Now that I've described these systematics and attempts at solutions, it would be beneficial to discuss how they will affect the scientific promise of HIM experiments.

In essence, these systematics make it more difficult (or impossible) to extract large scale information. Although rigorous forecasting should be done, I am as of now not too concerned about these systematics in dark-ages experiments, where much of the excitement has to do with the ability to measure *small scales*. For post-EOR observations, where larger scales are important to probe the BAO, this may be a bigger problem. In particular, the BAO wiggles should be confined to $k < 0.4hMpc^{-1}$. As seen earlier, these are the sorts of scales threatened by the systematics. I will not discuss implication for EOR experiments as I am less knowledgeable about which length scales are of interest in such astrophysics contexts.

In the post-EOR context, I'd like to remind readers that the “pessimistic" assumption of Sailer et al. (2021) was that scales with $k < 0.1hMpc^{-1}$ would be unobservable. Thus, most of the BAO is still observable under this assumption. This did decrease sensitivity to cosmological parameters, although the forecasted difference between “pessimistic" and “optimistic" results became rather small once PUMA data was combined with data from other surveys.

If one includes polarized foregrounds and assumes they won't be cleanable, then this “pessmistic" picture becomes worse. As of now, it seems to me that only the work of Shaw et al. (2015) is able to effectively clean polarized foregrounds, but the calibration requirements of this method may be unrealistic.

To see the implications of uncleaned polarized foregrounds, take another look at figure 5.4, where I show a simulated visibility generated assuming a CHIME-like instrument and CRIME maps. In this case, polarized foregrounds are dominant for $k_{||} < 0.2hMpc^{-1}$. It should be noted that this visiblity is assuming the observations are not particularly close to the Galactic plane. The beams used in generating these simulated visibilities were rather achromatic. If one assumes chromaticity like what was seen in Fagnoni et al. (2020), then we will not be able to see the BAO at all. If we assume results like those of Saliwanchik et al. (2021), then we might be able to recover modes $k_{||} < 0.3hMpc^{-1}$. Thus, we'd be missing significantly more BAO than what was assumed by Sailer et al. (2021). Moreover, this visibility example covers frequencies $400 < \nu < 500\,MHz$. One can be expect these polarized foregrounds to become worse as we look at lower frequencies (higher redshifts). It is unclear to me exactly what scientific implications this will have, since I do not have the time to conduct a rigorous analysis like what was seen in Sailer et al. (2021).

I should add some caveats here. First, I should say that polarized foregrounds are as of now not well understood. As such, careful measurement may show that chromaticity of polarized foregrounds look a lot different than what is produced by CRIME simulations. Moreover, I generated these visiblities using simulated CHIME beams, which have 50 percent polarized leakage. As such, things will look somewhat better if we assume percent level leakage like what is described in the papers Moore et al. (2013) and Moore et al. (2017).

As far as the noise pedestal is concerned, one would need an extremely low level of coupling to ensure that the noise pedestal is below the HI signal. One possibility for handling them is to use the SVD method of Kern et al. (2019), but this would not work for baselines that vary slowly with LST. It may also be possible to subtract out the pedestal if one carefully calibrates to obtain values for the parameters seen in equation 5.52. Thus, the worst case scenario here is that we cannot remove the pedestal, meaning that no information cosmological information can be extracted. If the SVD of Kern et al. (2019) works, then we will be able to extract some information from a subset of our baselines. If we are able to calibrate well enough to subtract the pedestal, then the noise pedestal will not affect our cosmological results.

Summary of Mutual Coupling

In summary of these section, I believe we can say the following

- Port-to-port coupling between antennas in the array will lead to a noise pedestal and to beam distortions. For baselines oriented close to east-west, it may be possible to remove these effects using the SVD method proposed in Kern et al. (2019). However, this method does not work for baselines oriented close to the north-south axis.
- The port-to-port beam effects may also be accounted for if one understands the S-parameters well, as was shown in Virone et al. (2018), Josaitis et al. (2022) and others.
- The work of Fagnoni et al. (2020) shows that at least for HERA, the additional chromaticity provided by mutual coupling is mostly from the short circuit term.
- As of now, there hasn't been any good proposal for how to estimate or remove the effect of the short-circuit scattering term.
- The works of both Fagnoni and and Saliwanchik show additional chromaticity coming from mutual coupling. The results from Saliwanchik however are much more optimistic.
- The presence of the mutual coupling will also complicate calibration, and it is unclear how big of an issue this will be.
- In the post-EOR context, these effects will at best limit which BAO scales we can observe. At worst, we will not be able to observe the BAO at all.

6 DESIGN OF AN ULTRA-WIDEBAND ANTENNA FEED AND REFLECTOR FOR USE IN HYDROGEN INTENSITY MAPPING INTERFEROMETERS

In this chapter, I present an ultra-wideband antenna design I created with HIM systematic effects in mind. In this work, I attempted to minimize chromaticity and coupling without sacrificing too much in the way of sensitivity. Without further ado, I present an adapted version of the paper “Design of an Ultra-Wideband Antenna Feed and Reflector for use in Hydrogen Intensity Mapping Interferometers", published in volume 13 of the Journal of Astronomical Insturmentation in 2023. This paper was co-authored by Peter Timbie and I.

6.1 Abstract

This paper describes the design of a 5.5:1 bandwidth feed antenna and reflector system, intended for use in hydrogen intensity mapping experiments. The system is optimized to reduce systematic effects that can arise in these experiments from scattering within the feed/reflector and cross-coupling between antennas. The proposed feed is an ultra wideband Vivaldi style design and was optimized to have a smooth frequency response, high gain, and minimal shadowing of the reflector dish. This feed can optionally include absorptive elements which reduce systematics but degrade sensitivity. The proposed reflector is a deep parabolic dish with f/d=0.216 along with an elliptical collar to provide additional shielding. The procedure for optimizing these design choices is described.

6.2 Introduction

This paper describes the design of a 5.5:1 bandwidth feed antenna and reflector intended for applications in radio astronomy, particularly for the proposed Packed Ultra-wideband Mapping Array (PUMA) telescope (Ansari et al. (2019)). This feed will be referred to as the “proposed feed" in the paper. The proposed PUMA telescope is a radio-interferometer consisting of 32K tightly packed antennas. Each antenna will consist of a feed suspended over a 6 m diameter parabolic reflector. These antennas are planned to have a bandwidth from 200 to 1100 MHz, with extra emphasis placed on the performance from 200 to 475 MHz.

The paper also describes a version of the feed, referred to as the “minimum systematics feed". This feed is similar to the proposed PUMA feed but includes absorptive elements that decrease systematic effects but increase the noise temperature of the antenna system. It will be argued that this design may be useful for telescopes operating at lower frequencies than PUMA.

A major goal of observational cosmology is to measure the statistical properties of the spatial distribution of matter in the Universe. Such measurements are of interest as they could tell us about dark energy, inflationary physics, the growth of structure, early star formation and more. The distribution of matter in the Universe is traced by neutral hydrogen (HI). Thus, observations of the 21 cm line from HI can probe the matter distribution (up to a bias term). Using 21 cm emission or absorption to measure these density fluctuations without resolving individual galaxies is a technique referred to as 21 cm (or HI) intensity mapping (Chang et al. (2008)). Because of the expansion of the Universe, the 21 cm line will be redshifted by an amount that scales with the distance between the emitting region and Earth. For this reason, different epochs of the Universe can be probed based on the range of observing frequencies chosen. The information that can be extracted from the distribution varies with cosmological redshift, denoted by z. Measurements of the distribution at redshifts $z \leq 6$ could probe the growth of structure, the physics of dark energy, and inflationary physics. Measurements during the Epoch of Reionization (EOR) ($6 < z < 20$) would tell us about the formation of the first stars. Lastly, measurements during the cosmic dark ages ($20 < z < 1100$), an epoch preceding the first stars and relatively unaffected by astrophysical phenomena, would probe the physics of inflation. The 200 to 1100 MHz bandwidth of PUMA is chosen to capture redshifts $0.3 \leq z \leq 6$.

For reasons of sensitivity and cost, radio-interferometers are typically used for HI intensity mapping. In interferometry, an array consisting of many antenna elements is used. The signals from pairs of antennas are correlated to form “visibilities". In the absence of mutual coupling, correlating these signals eliminates the offset from system noise one sees in single dish telescopes. This in turn relaxes requirements for gain stability. Moreover, combining signals from many pairs of antennas allows the same angular scale to be measured many times, providing a distinct advantage over single dish experiments.

Such interferometers are optimized to measure the spatial power spectrum $P(k)$ of the 21 cm line. Cosmologists are particularly interested in large spatial scales, corresponding to small values of the wavenumber k. This interest in large spatial scales leads to interferometers with tightly packed antenna elements. In the case of PUMA, at the longest wavelength, dishes are planned to be about 4λ wavelengths in diameter with centers located about 5λ from each other. This spacing is small enough to probe large spatial scales at all wavelengths, while also being large enough to allow dishes to be repointed to latitudes away from zenith without beam blockage. Moreover, the HI signal being measured is small, with a brightness temperature of roughly 1 mK, while system temperatures are the order of $\sim 25 - 100$ K in the post EOR universe and up to ~ 1000 K in the case of EOR measurements. The small magnitude of the signal being measured motivates interferometers with large collecting areas. These requirements lead to ‘large N, small D’ designs, which consist of large numbers of small diameter antennas. Although this paper focuses on the proposed PUMA telescope, other designs exist as well. Table 6.1 provides summaries of several interferometers intended for measuring the EOR and post EOR Universe.

In an HI intensity mapping instrument, candidate antennas must be designed to minimize systematic effects. The primary systematic effect is contamination from Galactic synchrotron radiation. The Galactic foregrounds are of concern as they are about 5 orders of magnitude brighter than the predicted HI signal. However, the foregrounds have a smooth spectrum, while the spectrum of the HI signal is quite ‘spikey’, corresponding to clumps of hydrogen gas along the line of sight. Thus, one can confine the Galactic foregrounds into a small region of the delay domain by ensuring that the antennas have a smooth response as a function of frequency. In 21 cm intensity mapping, delay is the Fourier dual to spectral frequency (Ewall-Wice et al. (2016)). Moreover, one needs to minimize mutual coupling between antennas, which becomes challenging due to the close element spacing required in HI interferometers. Such mutual coupling can degrade delay spectrum performance, introduce correlated noise, and make certain calibration methods more difficult. Designs that minimize these systematic effects can be seen in Fagnoni et al. (2021) as well as Saliwanchik et al. (2021). Of course, one also needs to consider the sensitivity of the instrument to the HI power spectrum. Sensitivity will be given less priority during optimization than the previously mentioned systematics effects, since sensitivity can be improved by increasing the number of elements or observing for longer.

Lastly, polarization leakage also should be considered. The desired HI signal is unpolarized, while the foregrounds do have a polarized component. Leakage of polarized foregrounds into estimates of unpolarized emission can be caused by beam ellipticity and by cross-polarization in the beam. Such leakage is an additional systematic effect one could consider when designing an antenna for HI intensity mapping. This design was not optimized with polarization leakage in mind, but cross-polarization and leakage estimates are still provided in Section 6.5.

This paper is organized as follows. Section 6.3 describes the performance requirements of the instrument and the motivation for the requirements. Section 6.4 describes the optimization process for the feed antenna along with the resulting design. In addition, it shows simulations of the S-parameters and beam patterns. Section 6.5 summarizes the performance of both the proposed and minimum systematics antenna systems. Finally, Section 6.6 describes the fabrication of a $4/13$ scale version of the minimum systematics feed and provides some measured results.

6.3 Performance Requirements

The feed antennas and reflector were designed with HI intensity mapping instruments such as PUMA in mind. PUMA is a large, tightly packed antenna array. In instruments such as PUMA, mitigation of systematic effects is of primary importance. In addition, one must also consider the sensitivity of the instrument to the signal being measured.

In Sections 6.3 and 6.3, aspects of performance most crucial to minimizing systematic effects are considered. In Section 6.3, sensitivity of the instrument to the HI signal is considered.

Spectral Smoothness and the Foregrounds

Due to their smooth frequency spectrum, the foregrounds encountered in instruments such as PUMA are naturally contained in a region of 2-dimensional k-space (with k components parallel to and perpendicular to the line of sight) referred to as “the wedge". When taking measurements, the wedge is convolved with the response of the instrument, causing foregrounds to leak out into otherwise signal-dominated modes. In 21 cm intensity mapping, it is the case that delay (τ) is proportional to line of sight wavenumber $k_{||}$. It is thus desirable for the antenna response in the delay domain be concentrated at 0 delay and decay rapidly at larger delays.

In order to test the smoothness of the antenna's response, we will excite the antenna with a planewave coming from some direction $\hat{n}$. The voltage at the terminals of the antenna is then given by

$$v(\hat{n}, \nu) = \mathbf{r}(\hat{n}, \nu) \cdot \mathbf{E}(\hat{n}, \nu). \tag{6.1}$$

In this formula, **r** is referred to as the “voltage beam" of the antenna. We then quantify the delay performance of the antenna is via the Fourier transform of the power kernel. The power kernel is defined as

$$R(\hat{n}, \nu) = |\mathbf{r}(\hat{n}, \nu)|^2, \tag{6.2}$$

In particular, we will characterize the delay response of the instrument using the delay spectrum of the power kernel at zenith. This is denoted via $\tilde{R}(\theta = 0, \phi = 0, \tau)$. The goal is to have $\tilde{R}(0, 0, \tau)$ drop by 50 dB as quickly as possible as a function of τ. Refer to Ewall-Wice et al. (2016) for more details.

Mutual Coupling

Another concern in HI intensity mapping instruments is mutual coupling between the antennas in the array. It has been found that mutual coupling changes the beam patterns of antenna elements as described in Fagnoni et al. (2020) and Kern et al. (2019). This effect on the beam patterns will likely degrade spectral smoothness (Fagnoni et al. (2020)). Moreover, it causes beam patterns to vary from element to element. Such beam pattern variation is undesirable as it would prevent the use of redundant calibration techniques which have been proposed to overcome the computational challenge of calibrating large, redundant, arrays. In redundant calibration, one assumes that antennas in the array have identical beam patterns. This assumption would not be valid in the case of significant mutual coupling (Liu et al. (2010)).

Mutual coupling can also cause receiver noise radiated from one antenna to be absorbed by a different antenna. This coupling creates “correlated noise", which causes non-zero visibilities. These correlated noise contributions can masquerade as sky signals (Kwak et al. (2024)). In principle, this coupling varies slowly in time, but it sets a requirement on the gain stability of the instrument.

During the design process, mutual coupling was characterized by simulating the S_{21} coupling between two adjacent antennas. The antennas are placed 8 m apart and pointed at zenith. Such a setup can be thought of as the worst case scenario for coupling between two antennas in arrays such as PUMA. Simulations were performed with antennas co-located in each other's E-planes and H-planes. Using the model for coupled signals described in Kern et al. (2019), one should find that the “correlated noise" between two antennas follows Eq 6.3.

$$T_{coupled} \propto |S_{21}| T_{ampl}. \tag{6.3}$$

Since T_{ampl} is typically 4 to 5 orders of magnitude larger than T_{HI}, the S_{21} must be kept below $-100\,\mathrm{dB}$ to prevent correlated noise from overpowering the signal. It should be acknowledged that the model presented in Eq 6.3 is an approximation. A more sophisticated model of this phenomenon is presented in Kwak et al. (2024).

Sensitivity

In addition to minimizing systematic effects, we require this antenna (and array) to provide enough sensitivity to detect the HI signal. In such an measurement, the contribution of system noise to the power spectrum goes like (Parsons et al. (2014))

$$P_N \propto \frac{\Omega_p^2}{\Omega_{pp}} T_{sys}^2, \tag{6.4}$$

Let P be the power pattern of the antenna. We define Ω_p and Ω_{pp} in terms of P via

$$\Omega_p \equiv \int d\Omega P(\hat{\mathbf{n}}) \tag{6.5}$$

and

$$\Omega_{pp} \equiv \int d\Omega P(\hat{\mathbf{n}})^2. \tag{6.6}$$

In these formulas, the system temperature T_{sys} is estimated via (Ansari et al. (2019))

$$T_{sys} \approx \frac{T_{ampl}}{\eta_{sp}\eta_c} + \frac{1-\eta_{sp}}{\eta_{sp}} T_{ground} + T_{sky}. \tag{6.7}$$

In this formula, T_{ampl} is the noise temperature of the amplifier and η_{sp} is defined as the fraction of the antenna's beam that shines on the sky. One can compute η_{sp} using

$$\eta_{sp} = 4\pi \int_0^{\frac{\pi}{2}} \int_0^{2\pi} d\theta d\phi G(\theta, \phi). \tag{6.8}$$

The optical efficiency η_c is defined as

$$\eta_c = \eta_{S11}\eta_{rad}, \tag{6.9}$$

where η_{S11} is the fraction transmitted through the feed antenna's input terminals and η_{rad} is the radiation efficiency.

For the nominal PUMA design (Ansari et al. (2019)), the proposed values of these parameters for achieving adequate sensitivity are: $\eta_{sp} = 0.9$, aperture efficiency $\eta_A = 0.7$, $\eta_c = 0.9$ and $T_{ampl} = 50\,\mathrm{K}$. Thus, the designs proposed in this paper ought to provide comparable sensitivity to these fiducial parameters.

The performance requirements for the antennas proposed in this paper are summarized in Table 6.2.

6.4 The Proposed Antenna and its Optimization

This section describes the electrical and mechanical design of the proposed system, and motivates the choices made to achieve the performance requirements. The proposed antenna system was designed with the electromagnetic simulation software package CST Microwave Studio [1]. The work described in this paper primarily used CST's time domain solver. However, the frequency domain solver and integral equation solver were used at various times to double check results. All beam patterns presented in this work were computed using the “open (add space)" boundary condition in CST.

---

[1] `http://www.cst.com`

Dish Reflector

In this paper, 6 m diameter parabolic dishes with $f/d = 0.216$ and 0.375 are shown. These dishes are on-axis reflectors. Off-axis reflectors were also considered, but are not favored since they provide limited shielding against mutual coupling between antennas. The diameter of the dishes corresponds to 4λ at the longest wavelength. This small dish size was chosen since smaller dishes allow for a more tightly packed array, which in turn provides access to larger angular scales on the sky. This small dish size drove many design choices in the feed. In particular, it was one of the main reasons a Vivaldi-style feed was chosen, as opposed to a quad-ridge horn. The use of a horn antenna would cause too much shadowing when placed directly over a small dish.

A low f/d ratio is favored in order to provide shielding for the feeds, which reduces mutual coupling. An example of the benefits of using a deeper dish can be seen in Saliwanchik et al. (2021). In addition, outfitting the deep dish with a collar to provide extra shielding was considered. In this particular paper, a simple elliptical collar was used. Collars providing a smoother junction with the dish can be designed using the methods shown in Gupta et al. (1990). Using a curved collar, whether elliptical or through Gupta's method, is desirable for two reasons. The first is that these methods minimize the discontinuity at the boundary between the parabola and the collar. This in turn minimizes diffracted fields emanating from that boundary. The second reason is that the curvature of the collar minimizes standing waves between the feed and collar. If this were not the case, then the presence of the collar would degrade the delay spectrum significantly.

Some of the dishes considered can be seen in Figure 6.1. Decreasing f/d increases reflections between the dish and feed, which in turn degrades delay spectrum performance. Fortunately, the reduced mutual coupling provided by a deep dish improves delay spectrum performance by cutting down on inter-dish reflections (Saliwanchik et al. (2021)). It was found in the Saliwanchik *et al* that these inter-dish reflections dominate over intra-dish reflections. Thus, f/d should be chosen as low as possible while still being mechanically feasible and providing adequate sensitivity.

Some example beam patterns can be seen in Figure 6.2. These patterns were computed using an $f/d = 0.216$ dish with an elliptical collar. One will note that the value of the directivity at 90° is about 40 dB to 50 dB below its peak value. This behavior is desirable as low directivity towards the horizon corresponds to lower mutual coupling between antenna elements.

One will also notice however, that a there is a dip in gain at zenith for the 1000 MHz and 1100 MHz plots. This feature is not particularly concerning with respect to power spectrum measurements, as one is still sampling the same patch of the uv plane whether the null is present or not. So, as long as the beam is calibrated well, it should not be an issue. However, the complexity of these beams is concerning for calibration, since a complicated beam pattern will be more difficult to measure whether using drones or point source references.

It should be noted that this dip is due to phase center variation with respect to frequency. It can

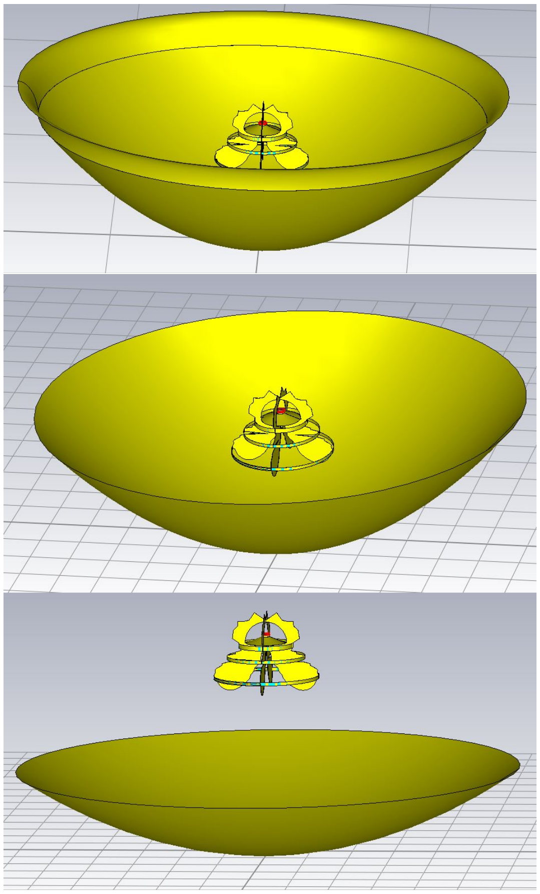

Figure 6.1: The three dishes considered in this paper. The top image shows the $f/d = 0.216$ dish with the elliptical collar included. The middle image shows the $f/d = 0.216$ dish without the collar. The bottom image shows the $f/d = 0.375$ dish.

be fixed by adjusting the location of the feed with respect to the focus of the dish. However, such adjustments will come at the expense of phase efficiency at low frequencies.

Feed

The feed designs presented in this paper are of a Vivaldi style. Such a design was chosen for its wide bandwidth, high directivity, and 'openness'. It is easier to achieve high spectral smoothness with an open design since there are no enclosed spaces in which resonances might form. Moreover, the Vivaldi geometry minimizes aperture blockage, which is especially important in this context where the dish size is only $D \approx 4\lambda$ at the longest wavelength.

Figure 6.3 presents a drawing of the minimum systematics feed. The proposed feed is identical to the minimum systematics feed, except it does not include the rings. All feed designs considered in this paper are shown in Figure 6.12. In particular, the proposed feed is shown in the bottom left

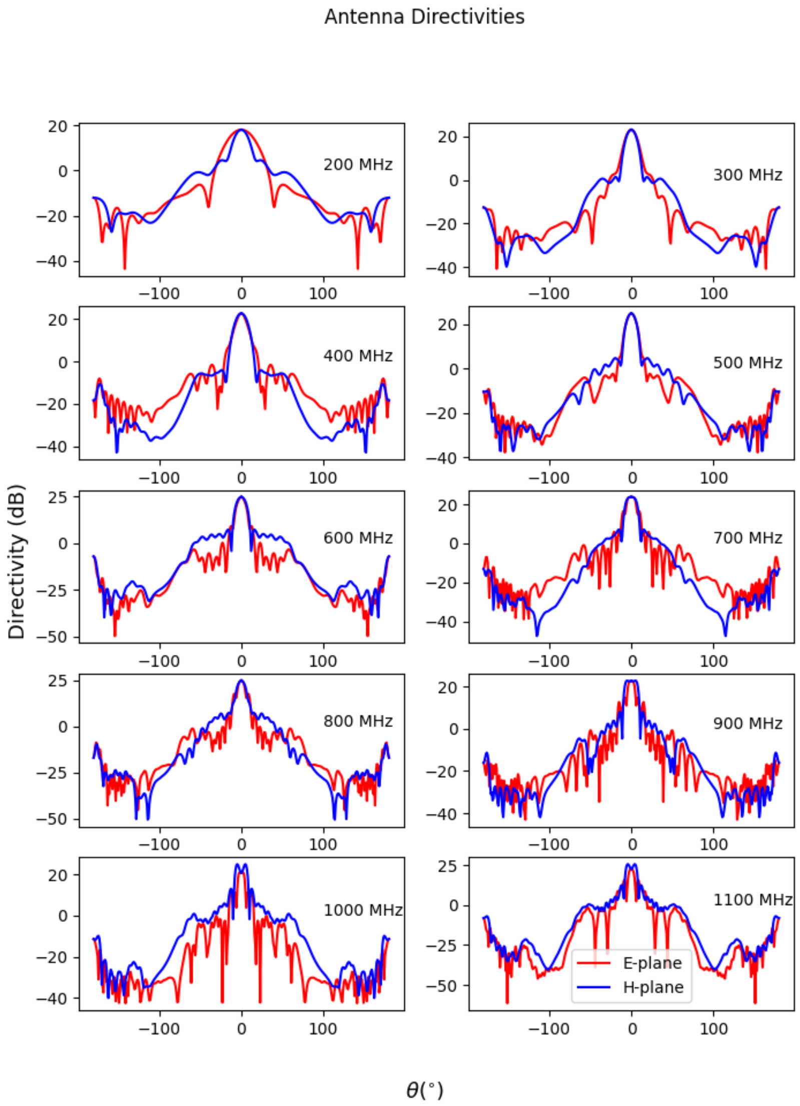

Figure 6.2: Directivities in the E and H-planes for the proposed antenna system

. These patterns were simulated using the $f/d = 0.216$ reflector with an elliptical collar. Note that the beam patterns of the minimum systematics feed look similar to these. However, the directivity at the horizon tends to be lower.

panel of the figure. Although not presented, horn antennas were also simulated but provided poor performance due to aperture blockage.

Ridges

This section describes the ‘ridges’ of the feed, which are shown in Figure 6.4.

As one can tell from the figure, the proposed feed was designed with a numerical optimizer.

In particular, the CMA-ES optimization algorithm was used (Hansen et al. (2003)). This algorithm is built-in to CST. The design of these ridges began as a more traditional Vivaldi shape, below called the ‘vanilla design’, with an exponential flare inside and flat section outside. The profile was then broken into line segments connecting points $(x_1, z_1), (x_2, z_2), \ldots$ The CMA-ES optimizer was then used to find the optimal values for the points (z_i, x_i). Note that only x_i was optimized for certain points and only z_i for others.

Allowing the ridge geometry to vary in this way was found to minimize sidelobes and backlobes, increasing spill efficiency (η_s). Changing the ridge geometry from the vanilla to jagged design was found to degrade delay spectrum performance to some extent. However, it was found that this difference is negligible compared to the non-smooth contribution from dish-feed reflections. Moreover, the addition of the absorber-lined rings described in Section 6.4 helps somewhat with spectral-smoothness. So, the final feed design performs comparably to the vanilla design (shown in the bottom right of Figure 6.12) as far as spectral smoothness is concerned.

Differential Line

The output signals are carried from the antenna to amplifiers via a differential line, which can be seen in Figure 6.5. Differential lines have been used on designs such as the horn described in Beukman et al. (2014), and the HERA Vivaldi described in Fagnoni et al. (2021). This choice eliminates the need for baluns, thus reducing losses and lowering T_{sys}. It has also been shown to improve beam patterns (Beukman et al. (2014)). Moreover, it introduces the possibility of using common modes for beam pattern shaping (Beukman et al. (2014)).

This differential line is filled with a PTFE “puck" of diameter 88 mm in order to decrease the impedance. The pins have diameter 4.2 mm. The line was chosen to be 26 mm long. These values were determined by the optimizer. The resulting line has impedance $Z_0 \approx 200\,\Omega$.

Such an impedance is quite high and not ideal for impedance matching to the amplifier. The reason the impedance is high comes from the wide bandwidth of the feed. A lower impedance requires a larger diameter differential line. Such a line would allow TE and TM modes to appear at the higher frequencies in the band. It should be noted, though, that one could lower this impedance by replacing the PTFE with a dielectric with an even higher index. For instance, using Taconic

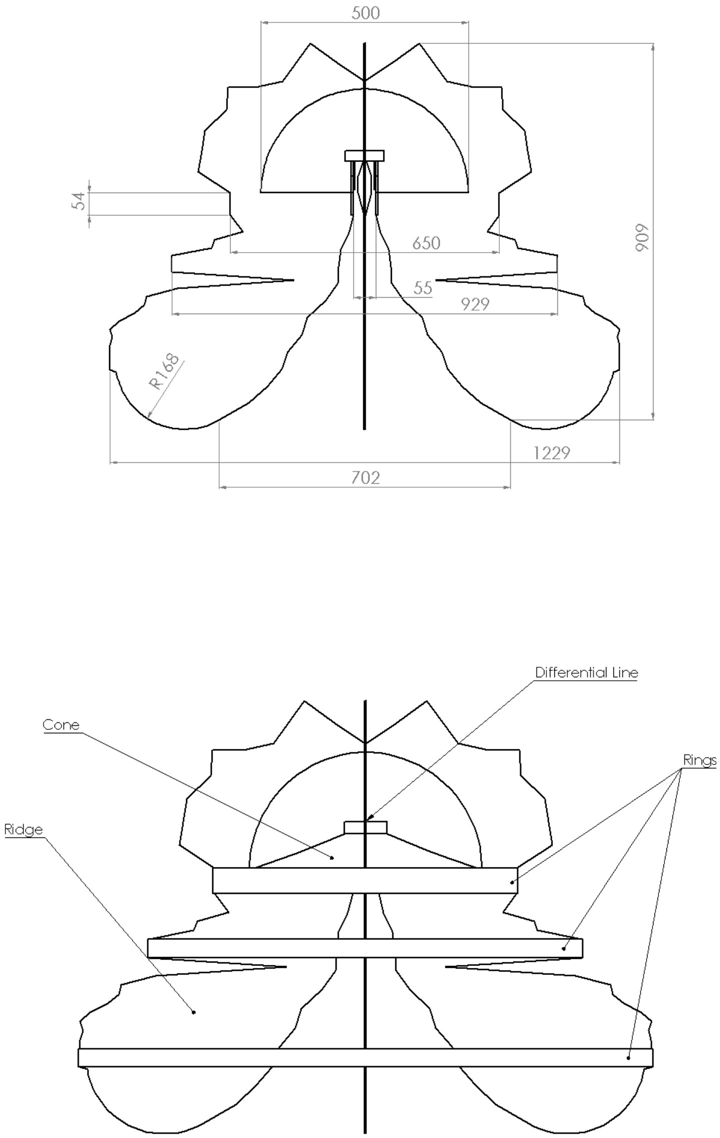

Figure 6.3: Top shows the minimum systematics feed with the cone and rings removed in order to make certain dimensions easier to see. All dimensions are in mm. Note that the ridges of the feed are 3.3 mm thick. The bottom image shows the minimum systematics feed with certain important pieces labeled. The proposed feed is identical to the minimum systematics feed, but with the rings removed.

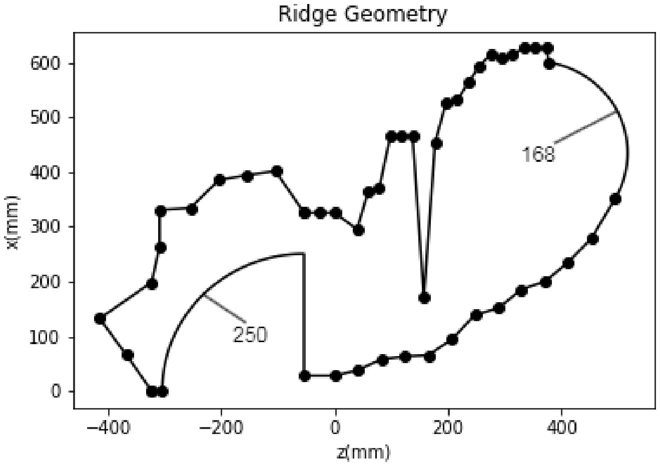

Figure 6.4: A drawing of one of the four ridges used on the proposed feed. $r = 168$ mm refers to the radius of the circular flare used at the aperture. Points denote the locations (z_i, x_i) used in the optimization.

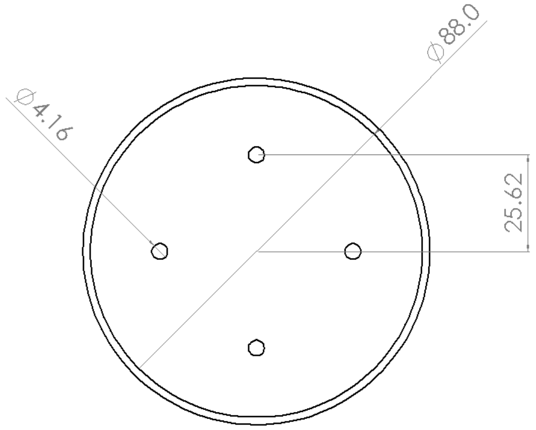

Figure 6.5: The differential line used on the feed. The line is 25.7 mm long.

RF-35, which has $\epsilon = 3.5$, would reduce the impedance to $\approx 150\,\Omega$. However, it would be best to pair this antenna with an impedance matching circuit between the feed and amplifier. This device would be similar to the HERA's front end module (FEM) (Fagnoni et al. (2021)).

Tabs

During optimization, it was found that the ideal thickness for the ridges on the feed was around 33 mm. In order to reduce weight and cost, the thickness of the ridges was reduced to about 3.3 mm. Reducing the thickness of the ridges increased the impedance of the feed, leading to a poor impedance match. In order to fix the issue, tabs were added, giving a smooth impedance transition

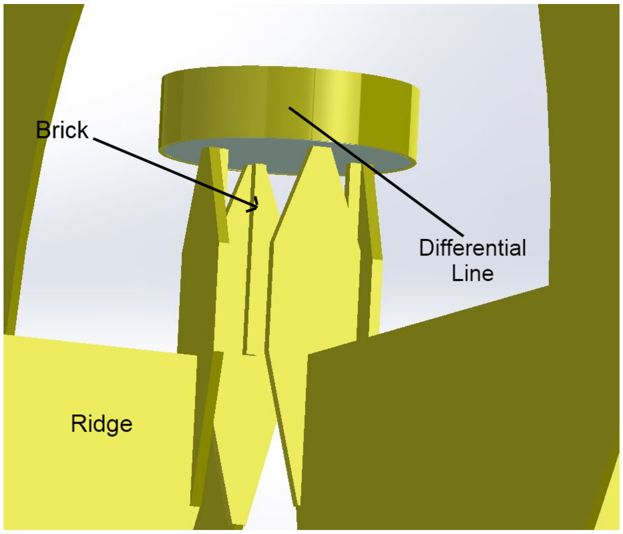

Figure 6.6: An illustration of the transition from the differential line to the ridges.

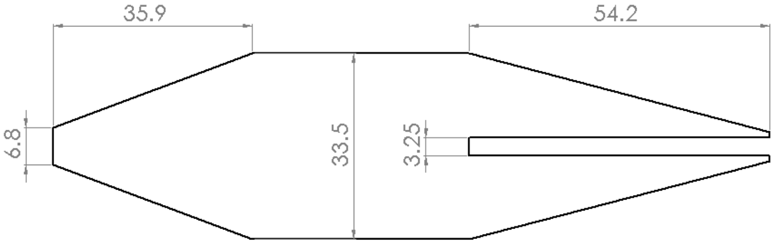

Figure 6.7: A drawing of one of the tabs used to transition from the differential line to the ridges, as seen in Figure 6.6. Dimensions are in mm and the tabs are 4 mm thick.

from the differential line to the throat of the feed. A similar approach was used in Fagnoni et al. (2021). An illustration of this transition can be seen in Figure 6.6. In this transition, the pins of the differential line are connected to rectangular “bricks". These bricks are 4.2 mm × 4.2 mm × 68 mm and are attached to the tabs. The tabs then connect to the ridges, providing a smooth impedance transition. Another view of the tabs, with dimensions, can be seen in Figure 6.7. By using the tabs, a good impedance match was achieved with only slight degradation in spillover efficiency compared to when thicker ridges were used.

Cone

To further improve beam patterns, a cone was introduced behind the throat of the ridges. This cone can be seen in Figures 6.3 and 6.9. The cone is meant to prevent energy from radiating away from the differential line before reaching the ridges. The taper of the cone is intended to give a smooth impedance transition from the line to the throat section of the ridges. As shown below in

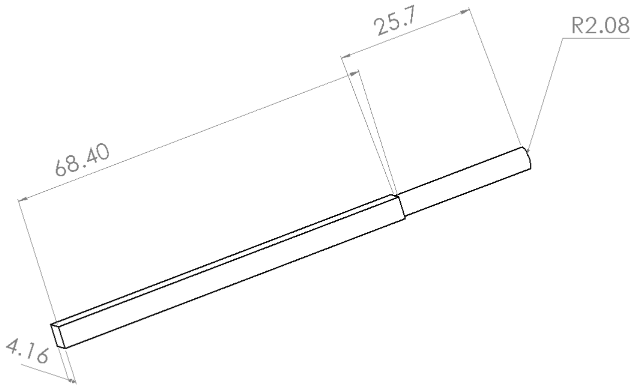

Figure 6.8: A drawing of the brick and pin used in the transition section and differential line respectively. One can see these pieces in the context of the feed in Figure 6.6. The brick fits into the slot of the tab, while the pin is used as part of the differential line.

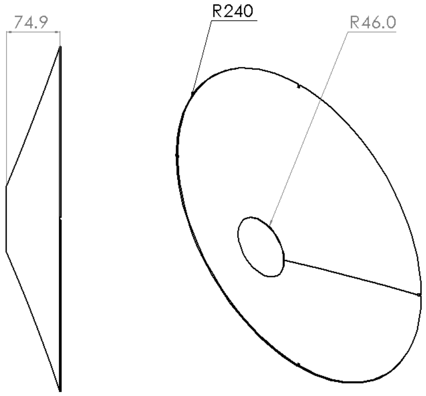

Figure 6.9: A drawing of the feed's cone, with dimensions. Our simulations used 1.7 mm as the thickness for the cone, but this dimension is not important for performance.

Section 6.4, the cone degrades the S_{11} at low frequencies by increasing the impedance. However, it also increases spill efficiency (η_s), which is a more important feature, since it corresponds to lower mutual coupling levels. Moreover, the presence of the cone was found to degrade spectral smoothness to some extent. However, as with the jagged ridge geometry, this is no longer the case once the rings are added to the design.

Rings

Lastly, rings were introduced to the design to improve its beam patterns. The incorporation of the rings came from the desire to decrease cross-coupling. In order to decrease cross-coupling, one would like minimize the beam pattern at angles toward the horizon. This sort of behavior is achieved by lowering illumination at the edge of the dish, i.e. making the feed's beam pattern narrower. In order to make the beam narrower, it was necessary to increase the size of the feed or add additional structure. It was found that the increased aperture blockage from the larger feed degraded performance. Wrapping a metal sheet around the feed to turn it into a quad-ridge horn was found to greatly improve the feed's beam patterns. However, as expected, this aperture blockage degraded performance. The compromise reached was to introduce the rings seen in Figure 6.3. These rings reduce the edge taper on the reflector while introducing minimal additional aperture blockage.

The geometry of the feed had already been optimized before the rings were introduced, so rings were added to locations on the ridges that were amenable to them. The ideal number and placement of the rings was found through trial and error.

Note, however, that an absorber was added to the rings in order to prevent a resonance from occurring. For the full-sized feed, our proposed design uses a flexible ferrite absorber. In particular, model M6 from Fair-Rite[2] was used in the simulations. The simulations described were performed using a layer of absorber 1.5 mm thick on the largest ring and 0.5 mm thick on the other two rings. The absorber does degrade radiation efficiency, as can be seen in Figure 6.10. Although not shown, simulated radiation efficiency was nearly perfect in the absence of the absorber. Note that the radiation efficiency is poorest at the lowest end of the band, but increases to -1 dB by 250 MHz.

It should also be noted here that the presence of the absorber increases the noise temperature of the antenna system. The noise contribution was measured for the fabricated 4/13 scale minimum systematics feed. In particular, the noise of the feed was measured at room temperature and submerged in liquid nitrogen. The noise temperature was then obtained by taking the difference of the two measurements while accounting for amplifier gain and impedance matching. The results of these measurements are presented in Figure 6.11. Note that only frequencies up to 1 GHz are presented due to the limitations of the low noise amplifiers available to us. Also, the lowest frequency measured was 650 MHz, as this corresponds to the lowest frequency in the band of the 4/13 scale

---

[2] `http://www.fair-rite.com/flexible-ferrite/`

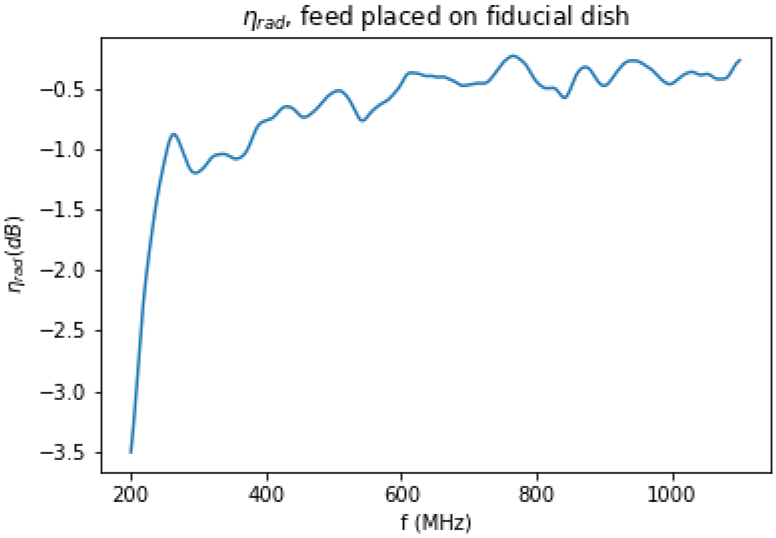

Figure 6.10: Radiation efficiency when the feed with rings is placed over a dish with $f/d = 0.216$ and an elliptical collar.

feed. From these measurements, it was found that the typical noise contribution of the absorbers was ≈ 120 K. Although these measurements were taken for the scaled down feed, it is safe to assume results would be similar for the full scale feed, as simulations showed similar beam patterns and S-parameters for the two versions.

Optimization of the Feed

Since this feed was optimized with PUMA in mind, the dimensions shown in the following sections are set to optimize performance from 200 MHz up to 1100 MHz.

The optimizer's error functions were defined as

$$\mathcal{E} = a\mathcal{E}_s + b\mathcal{E}_{S11} + c\mathcal{E}_d. \tag{6.10}$$

In this equation, $\mathcal{E}_s$ considers spill efficiency, $\mathcal{E}_{S11}$ considers impedance matching, and $\mathcal{E}_d$ is used to keep track of delay spectrum performance. The weights a, b, c were typically chosen to be ≈ 1, but their values were sometimes adjusted if some aspect of performance was lagging behind the others.

The error for the S_{11} term is defined via

$$\mathcal{E}_{S11} = \frac{1}{N}\Sigma_{n=1}^{N}\Theta(S_{11}[n] + 10)[S_{11}[n] + 10]^2. \tag{6.11}$$

In this formula, n indexes the frequency samples in the simulation, S_{11} is given in dB and Θ is the

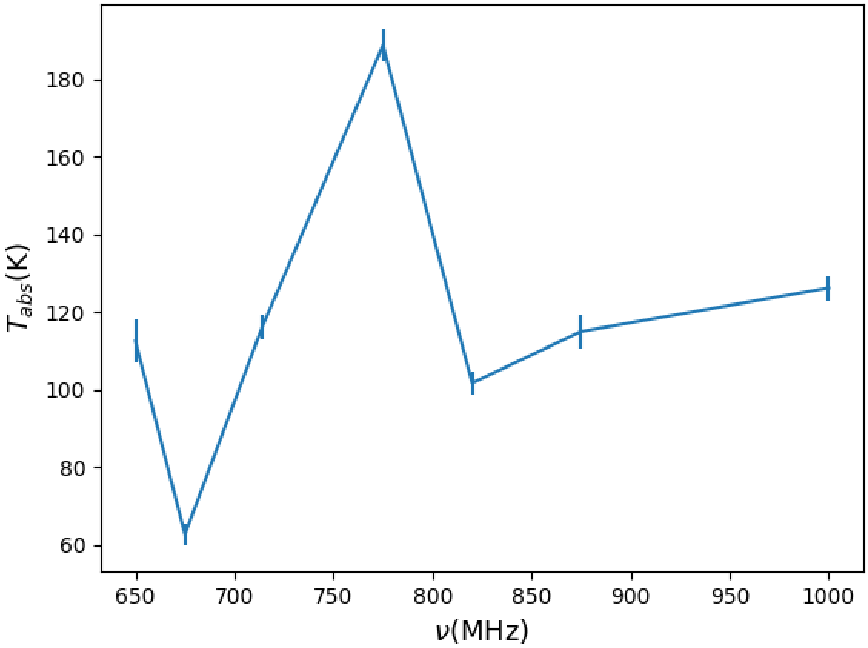

Figure 6.11: Noise contribution of the absorber in the fabricated minimum systematics feed.

Heaviside step function. In other words, the goal was to keep S_{11} below −10 dB.

The error in the spill term E_S is defined as

$$\mathcal{E}_s = \Sigma_n \Theta(0.93 - \eta_s[n])(0.93 - \eta_s[n]). \tag{6.12}$$

In this formula, η_s is defined as the portion of the feed's beam illuminating the dish. (Note that this is not the same as η_{sp}, which is defined as the portion of the dish's beam illuminating the sky.) The spill η_s is calculated assuming a dish with an f/d ratio of 0.25. Also, note that n is an index used to label the frequencies at which η_s was estimated. During optimization, it was found that the fine-tuned ridge geometry was not capable of improving η_s over the entire 200 to 1100 MHz bandwidth. As such, η_s was only considered for frequencies $200\,\text{MHz} < f < 475\,\text{MHz}$. This choice was made to maximize performance at the frequencies most important for PUMA's science goals. Finally, the delay error was defined as

$$\mathcal{E}_d = \Theta(\Delta t - 17)(\Delta t - 17), \tag{6.13}$$

where Δt is the time in nanoseconds that it takes for the energy in the system to drop by 30 dB when using the CST time domain solver. This ensures that no resonances are present in the feed. The

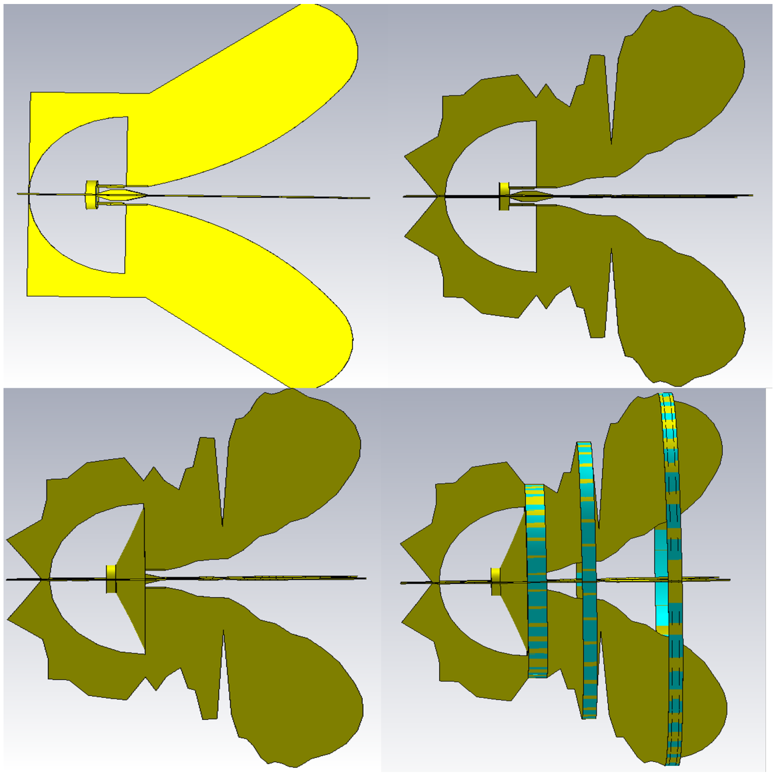

Figure 6.12: Pictures of feed versions. In reading order: vanilla feed, jagged ridges no cone, proposed feed and minimum systematics feed.

17 ns value was chosen based on a candidate design that was considered to have adequate delay spectrum performance.

A Performance Comparison

In this subsection, plots comparing the simulated performance for different feed designs are shown to illustrate the benefits of the design features added. These feed versions can be seen in Figure 6.12. In particular, η_s, which is the fraction of the feed's beam that illuminates the dish, η_{ap}, the S_{11} (of the feed alone), and some beam patterns are shown in Figures 6.13, 6.14,6.15, and 6.16, respectively.

In Figure 6.13, η_s is plotted for various iterations of the feed. “vanilla" (Figure 6.12) refers to a version of the feed with no cone and a simple exponential taper and straight lines for the outside and back part of the profile. One can see that the introduction of the jagged profile improves spillover efficiency at low frequencies but degrades it above $\approx$ 400 MHz. As mentioned in Section 6.4, this is because jagged ridges can't improve the beam patterns for all frequencies. So, the jagged ridges were optimized to improved performance in the lower part of the band where the most cosmological information is contained.

Introducing the cone (the proposed feed) reduces spillover at all frequencies, and is especially

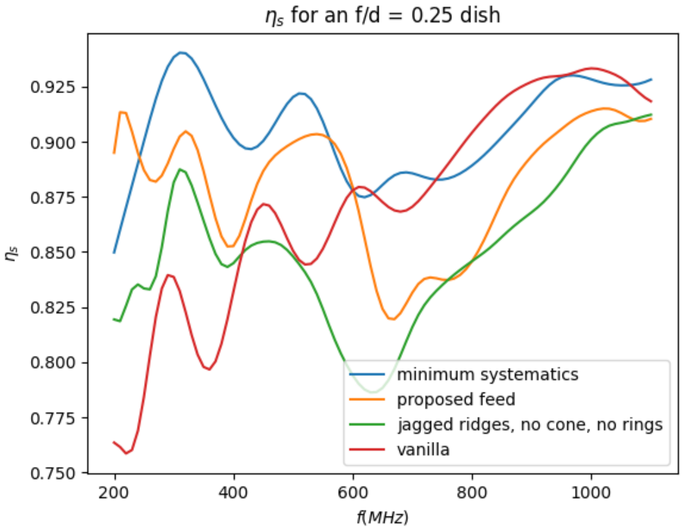

Figure 6.13: Comparisons of simulations of η_s as a function of frequency for different versions of the feed.

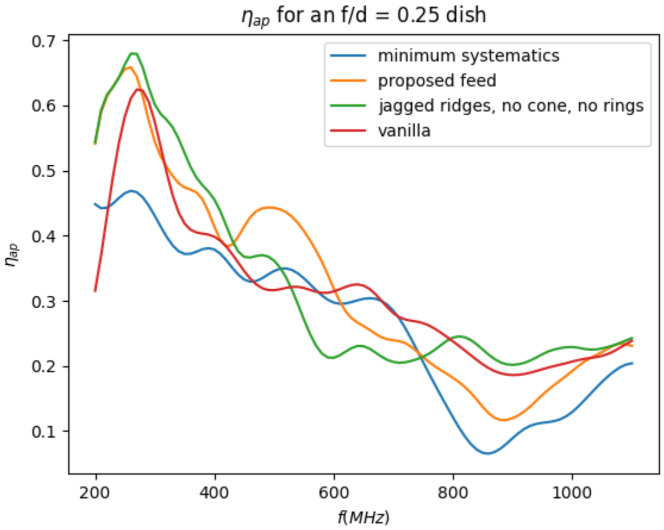

Figure 6.14: Comparison of simulated aperture efficiencies on a $f/d = 0.25$ dish as a function of frequency.

effective for $f < 300\,\text{MHz}$. Introducing the rings (minimum systematics) improves the spillover efficiency even further, and fixes the problems caused by the jagged ridges at higher frequencies.

Figure 6.14 shows simulated aperture efficiencies for the various versions of the feed design. This plot shows that the introduction of the curved ridges greatly improves the aperture efficiency at the lower end of the band. On the other hand, the introduction of the rings seems to degrade

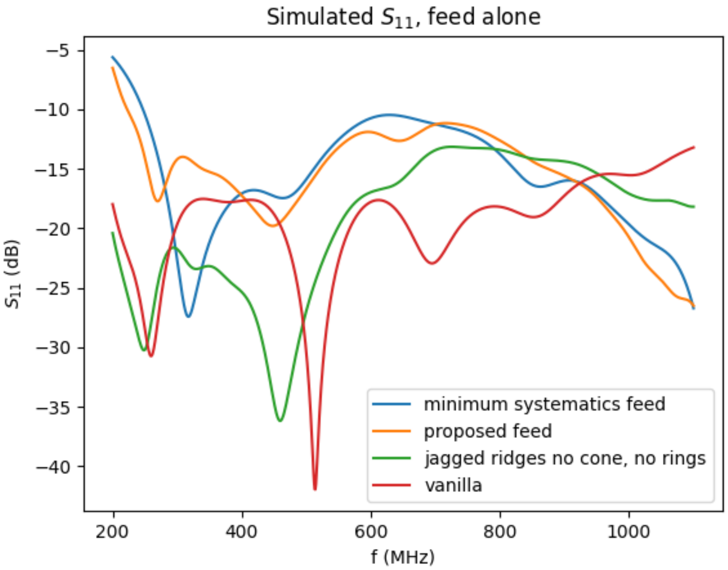

Figure 6.15: Simulated S_{11} values for different versions of the feed. The values are simulated without the dish, and assuming a reference impedance of $200\,\Omega$.

the efficiency of the minimum systematics feed. One will also notice that the aperture efficiency decreases significantly at frequency increases. This is due to the fact that the feed geometry is large compared to such short wavelengths, meaning that the dish is located in the nearfield of the feed. Although not shown, predictions of aperture efficiency based on feed farfield patterns agree with Figure 6.14 at low frequencies but begin to disagree significantly as frequency increases. Despite the low aperture efficiency, these feed designs can still achieve adequate sensitivity at high frequencies due to their high spillover efficiencies.

Figure 6.15 shows simulated S_{11} values for different versions of the feed. This plot shows that versions of the feed without the cone have $S_{11} \lessapprox -15\,\mathrm{dB}$ for the entire band. The introduction of the cone raises the S_{11} above $-10\,\mathrm{dB}$ for the frequencies below $f \approx 250\,\mathrm{MHz}$. This is due to the fact that the cone raises the impedance of the antenna at these frequencies. This slightly higher S_{11} should not cause significant problems, however, since receiver noise is subdominant to the sky temperature at these frequencies (Ansari et al. (2019)). Aside from that, the S_{11} stays below the goal of $-10\,\mathrm{dB}$ for the rest of the band. Lastly, note that introducing the rings only makes a slight difference to the S_{11}.

Figure 6.17 shows how the phase center varies with frequency for different feed designs. These phase centers were computed by CST. The “boresight" setting in CST was used, with a 90° degree angle around the z-axis. The values shown here are the average of the phase centers computed in the E-plane and H-plane.

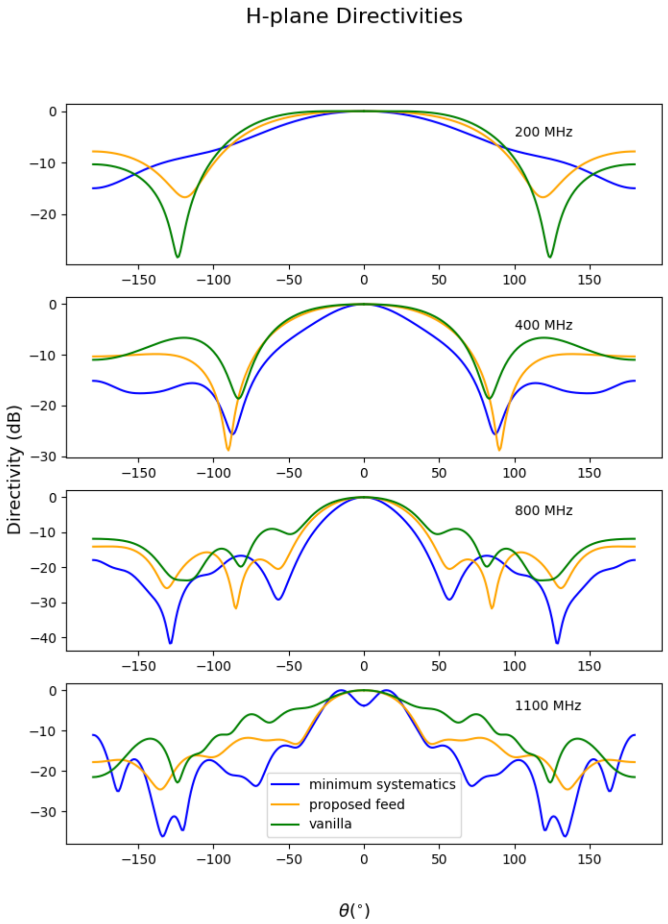

Figure 6.16: Simulated peak-normalized H-plane directivities for different versions of the feed.

First, note that the phase center (Beukman et al. (2014)) of the minimum systematics design varies by about 400 mm across the band. Such variation means that achieving good phase efficiency for the minimum systematics feed may not be possible over the whole band. Also, note that the cone helps keep the phase center stable at low frequencies. This is due to the fact that the cone prevents a resonance from occurring in the backshort, which has a perimeter close to $\lambda/4$ at the longest wavelengths of operation. Moreover, the feed with rings seems to have three different regimes of behavior. At low frequencies, the phase center is located close to the largest ring. In the middle of the band, the phase center is closer to the middle ring. At low frequencies the phase center is close to the smallest ring.

This phase center variation is responsible for the dip in gain at zenith seen in figure 6.2 at high frequencies. For these beam patterns, the feed was placed 1500 mm above the dish, meaning that the phase center is located 200 mm from the focus. At low frequencies, this means that the phase center will be quite close to the focus. However, at higher frequencies the phase center will be about

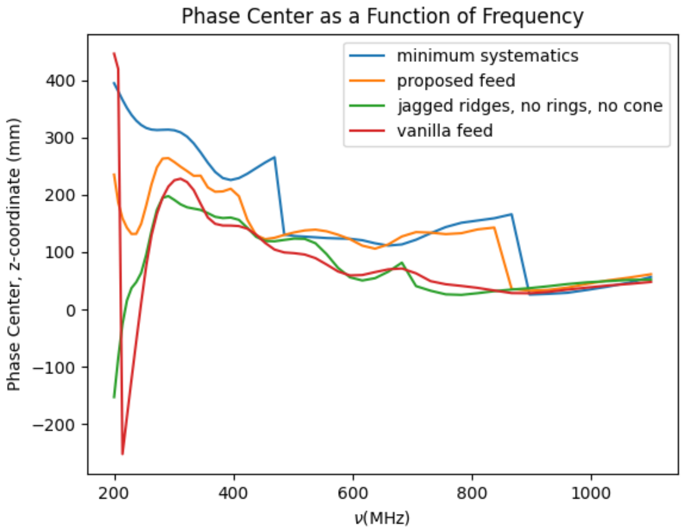

Figure 6.17: A plot showing the boresight coordinate of the phase center as a function of frequency.

150 mm from the focus, leading to higher phase errors. These errors are responsible for the dip in gain seen at zenith in Figure 6.2. In this paper, we have chosen the placement in order to optimize for performance at lower frequencies.

Finally, consider the beam patterns of the feeds. H-plane patterns for the proposed feed, the jagged feed with a cone, and the vanilla feed can be seen in Figure 6.16. Note first that the presence of the rings causes the beam pattern to taper off more quickly and smoothly. This feature is important since it will help suppress mutual coupling in the H-plane, where it tends to be most severe. Moreover, note that the proposed feed has suppressed the backlobes. The smaller backlobes are a result of the rings used in that design. E-plane beam patterns can be seen in Figure 6.18. For the E-plane, note that the red proposed feed performs the best at the very lowest end of the band. Moreover, it can be seen that the different designs perform similarly as frequency increases.

6.5 Simulated Performance and Comparison to Requirements

In this section, relevant aspects of the antennas' performance are presented and compared to the requirements described in Section 6.3. Note that these metrics were all evaluated with the feed placed on a 6 m diameter dish with $f/d = 0.216$. Since no such dishes were available for testing, estimated results based on simulations will be shown instead. Measured beam patterns and S-parameters for the proposed feed will be shown in Section 6.6.

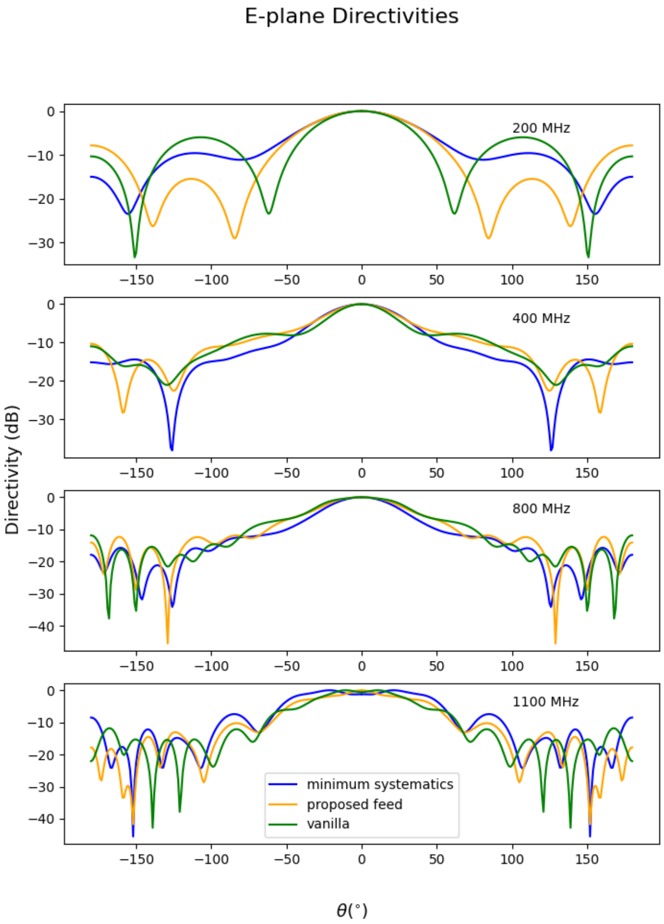

Figure 6.18: Simulated peak normalized E-plane directivities for different versions of the feed.

Delay Spectrum Performance

Recall from Section 6.3 that the delay response of the instrument can be characterized using $\tilde{R}(0, 0, \tau)$, the Fourier transform of the power kernel at zenith. In order to compute this, the antenna is excited with a plane wave propagating along the boresight of the antenna. This simulation was performed using the time domain solver in CST, with a planewave excitation coming from above, the differential line of the feed terminated by a waveguide port, and a voltage monitor placed between two pins on the differential line. The time domain solver simulated up until 400 ns after the initial excitation. This 400 ns time was chosen in order to capture delay spectrum performance at large delays. The voltage present at the antenna's terminals is recorded in order to obtain values for $R(0, 0, \nu)$. Then, two Blackman-Harris filters are applied in order to smoothly bring the power kernel to 0 at the edges of the band. These filters will be denoted as $W(\nu)$. This leads to a delay space power kernel

$$\tilde{R}(0,0,\tau) = \int d\nu R(0,0,\nu)W(\nu)^2 e^{2\pi i\nu\tau}. \tag{6.14}$$

Results for minimum systematics, proposed and vanilla feeds are shown in Figure 6.19 in terms of $k_{||}$, which is the spatial wavenumber along the line of sight of the telescope. In cosmological contexts, one finds that $k_{||} \propto \tau$.

First, note from Figure 6.19 that all three designs provide similar results in all three plots. So, it can be concluded that the presence of the cone, rings, and jagged ridges does not significantly worsen delay spectrum performance. Next, note that the top plot of the figure shows the power kernel dropping by 50 dB after a delay corresponding roughly to $k_{||} \approx 0.1\,\mathrm{hMpc}^{-1}$. Thus, roughly speaking, modes within $0.1\,\mathrm{hMpc}^{-1}$ of the wedge would be avoided in order to be sure that the foregrounds are not overpowering the HI signal. Readers should also note that lumps appear in the delay spectrum as $k_{||}$ increases. These lumps are due to reflections bouncing between the feed and the dish. In the middle plot, one can see that the lower end of the band performs a bit better, dropping by 50 dB after about $0.065\,\mathrm{hMpc}^{-1}$. This result is of interest as the band corresponds to redshifts $2 < z < 6$, which are particularly important for the PUMA experiment. Lastly, the bottom plot shows the delay spectrum for the feeds only. Readers will note that the kernel drops off quite quickly in this case, reaching −50 dB within $k_{||} \approx 0.03\,\mathrm{hMpc}^{-1}$. This is due to the absence of dish-feed reflections, which worsen performance in the middle and top plots. As mentioned in Section 6.3, the power kernel should drop 50 dB as quickly as possible. However, the presence of dish-feed reflections slows down this process, giving a −50 dB width that corresponds to $0.1\,\mathrm{hMpc}^{-1}$. The use of a higher f/d or off-axis design would help reduce these reflections. However, as seen in Saliwanchik et al. (2021), the increased mutual coupling from such a design would be worse than the dish-feed reflections one incurs with a deep dish (Saliwanchik et al. (2021)).

Mutual Coupling Performance

In order to evaluate the proposed antenna's mutual coupling performance, the S-parameters for two antennas were considered. These antennas both point at zenith and are taken to be separated by 5.3λ at the longest wavelength. Section 6.3 provides motivation on this setup. For these simulations, the vanilla, proposed and minimum sytematics feeds were considered. The results of these simulations are shown in Figure 6.20.

In the E-plane, the $f/d = 0.216$ no-collar dish has mutual coupling 20 dB to 40 dB lower than the shallower $f/d = 0.375$ dish. When the collar is used, the mutual coupling is 20 dB to 40 dB lower than the shallow dish when frequencies are between 200 MHz and 400 MHz. For the rest of the band, the S_{21} is 40 dB to 60 dB lower when the deep with-collar dish is used rather than the shallow dish.

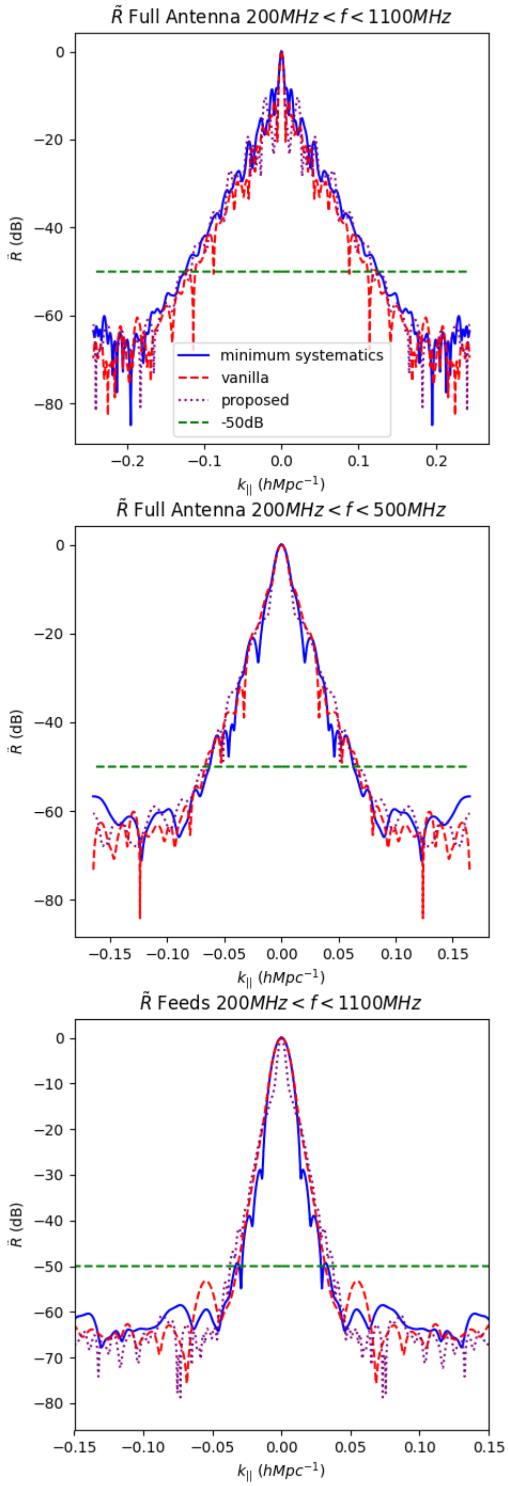

Figure 6.19: Delay spectrum: comparison of the power kernel of the full antenna for the minimum systematics, vanilla, and proposed feed. The top plot is for the full band. The middle plot was made using a bandpass that selects for frequencies 200 MHz $< \nu <$ 500 MHz. The bottom plot shows the power kernel of the feeds without the reflector present.

In the H-plane, the use of a deeper dish also helps with mutual coupling. The deep, no-collar dish provides about a 10 dB to 20 dB improvement in the lower part of the band. Note that the presence of the collar doesn't help quite as much in the H-plane, providing only about a 5 dB improvement.

Now, the various feed options will be compared. First, note that the proposed feed provides an advantage over the other two options in the E-plane at low frequencies for the deep dish with collar case. On average, we find the proposed feed providing about a 10 dB improvement over the vanilla feed for frequencies $\lesssim 300\,\mathrm{MHz}$. Moreover, the proposed feed also provides some improvement over the vanilla feed in the H-plane at low frequencies, giving about a 10 dB improvement over the vanilla feed for frequencies $\lesssim 400\,\mathrm{MHz}$. Notably, the minimum systematics feed does provide a coupling level about 20 dB better than the vanilla feed in the H-plane at the lower end of the band (from 200 MHz up to 400 MHz). Although these may sound like modest improvements at first, recall the lower end of the band is critical for PUMA's science goals. Thus, any improvement in that part of the band is especially important.

Recall also that the goal was to keep S_{21} levels below $-100\,\mathrm{dB}$. This goal is achieved in the E-plane for all designs for frequencies above approximately 600 MHz, when the deep dish with collar is used. Results do not quite meet the goal in the H-plane. The S_{21} of the minimum systematics feed stays below about $-70\,\mathrm{dB}$ for the entire band, which, although an improvement, does not meet the goal.

Recall that this simulation is meant to cover the worst case scenario, where the dishes are very close. For longer baselines, this coupling would certainly go down. However, shorter baselines cannot be ignored, since they provide information about large angular scales.

Since the mutual coupling level is still not quite low enough at PUMA's redshifts of interest, it will be necessary to remove this effect during the data analysis stage. So far, techniques have been developed for the removal of mutually coupled signals from the data (Kern et al. (2019)) (Parsons et al. (2010)). These techniques rely on the fact that coupled receiver noise varies slowly with time compared to the sky signal. However, this type of approach doesn't work so well for NS-oriented baselines, for which the sky signal also varies slowly with time. Moreover, how to handle the effects of mutual coupling on antenna beam patterns is still an open question. Thus, new techniques are likely needed to control mutual coupling.

Sensitivity Performance

In order to evaluate the sensitivity of the antennas, the PUMA noise calculator [3] was used. This calculator takes in parameters describing the properties of a fiducial PUMA design. These parameters describe beam size, observing time, and fraction of the sky observed. It then produces an estimate

---

[3] `https://puma.bnl.gov`

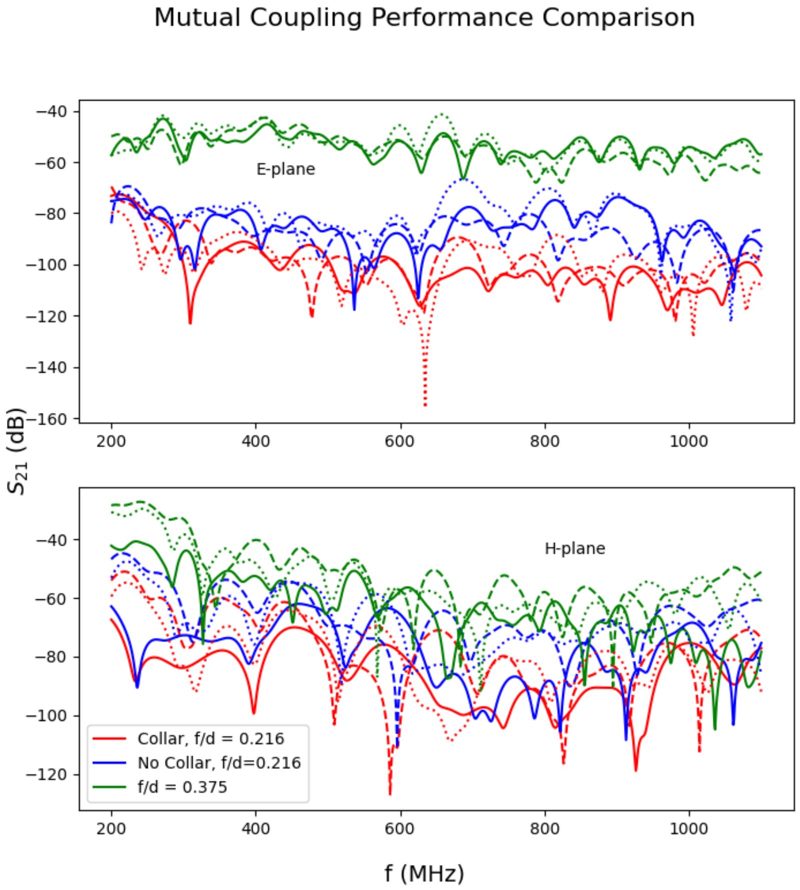

Figure 6.20: Mutual coupling: S_{21} values for various antenna options. The top plot shows S_{21} values calculated when each antenna is located in the E-plane of the other. The bottom plot shows S_{21} values when the antennas are located in each other's H-plane. Colors label different dish choices. Dotted lines correspond to the proposed feed, solid lines correspond to the minimum systematics feed, while dashed lines are used for the vanilla feed.

for the power spectrum due to noise via Equation D4 given in Ansari et al. (2019). This equation scales as Ω_p as in Eq 6.5. This contribution is computed for one $\mathbf{k}$ mode at a time. The calculator was edited, however, to use the formula for the noise power spectrum given in Eq 6.4. Sensitivity estimates are presented in Figure 6.21. Note that all of these curves assume an $f/d = 0.216$ dish with a collar included. Calculations were performed assuming a 32,000 element hexagonal array with 1.25 years of integration time. The integration time was set assuming that PUMA would run for 5 years with 6 hours per day of observation. A value of 50 K for was assumed for T_{ampl}. For the minimum systematics feed, we accounted for the noise contributed by the absorber by adding an additional term to Eq 6.7. This term is given by

$$\frac{\eta_{s_{11}} T_{abs}}{\eta_c \eta_{sp}}. \tag{6.15}$$

For these calculations, we assumed that T_{abs}, the noise temperature provided by the absorber, is 117 K. This value was obtained by averaging the data points we measured for Figure 6.11. Note that this is a rough estimate, since we were unable to accurately measure the noise contribution at higher frequencies.

The beam patterns used for the various antenna designs come from CST simulations with a single antenna pointing at zenith. Note that sensitivity will decrease slightly for pointings away from zenith due to increased noise pickup from the ground. This plot shows the performance of the proposed feed when paired with different reflector designs. The performance of the idealized PUMA antenna described in Ansari et al. (2019) is also shown. For the fiducial PUMA antenna, an Airy beam pattern and the parameters given in Section 6.3 were used.

Note in Figure 6.21 that the proposed feed and the vanilla feed do provide comparable sensitivity to the fiducial PUMA antenna, with somewhat worse performance at lower redshifts and somewhat improved performance at higher redshifts.

Lastly, one will also surely notice that the sensitivity provided by the minimum systematics feed is significantly lower than the other feed designs presented. This is due to the decrease in radiation efficiency and increase in noise caused by the absorber present on the rings. In the context of PUMA, this leads to significantly decreased sensitivity to smaller scale modes. However, this increase in noise temperature may not be so severe in the context of EOR telescopes, where the sky temperature can be up to 1000 K, which is significantly higher than the $\approx$ 100 K provided by the absorber.

Cross-Polarization

Although not considered for the performance requirements of the antenna, the cross-polarization properties will now be presented. Figure 6.22 shows the peak-normalized cross-polarized pattern of the proposed feed in the D-plane, where it is most severe. The peak-normalized patterns are given in dB, as

$$20\log_{10}\left(\frac{|E_{xp}(\theta,\phi)|}{\max(|E_{co}|)}\right), \tag{6.16}$$

where the E functions refer to the electric field patterns of the feed. For this paper, we use the Ludwig-3 convention for cross-polarization (Ludwig (1973)) (Beukman et al. (2014)). These values vary significantly with θ, with the highest levels being about 10 dB below the peak value. These peak values are comparable to those shown in Beukman et al. (2014).

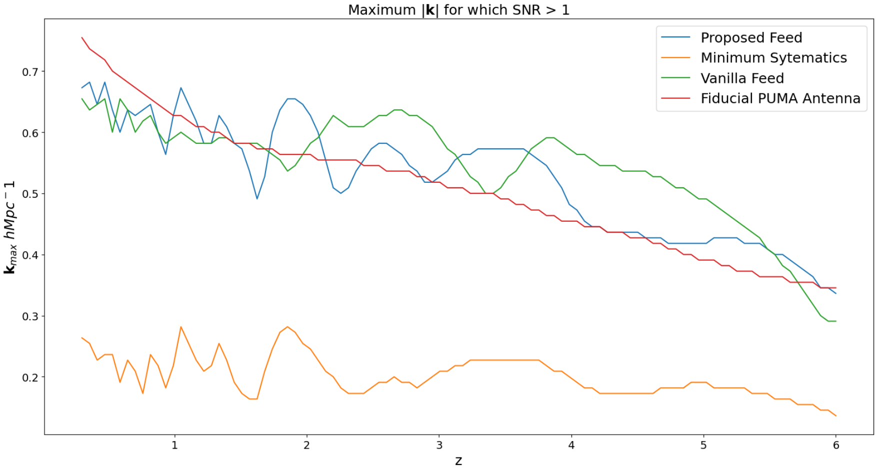

Figure 6.21: Maximum $|\mathbf{k}|$ for which SNR > 1. These calculations were performed assuming $\hat{k} \cdot \hat{n} = 0.5$.

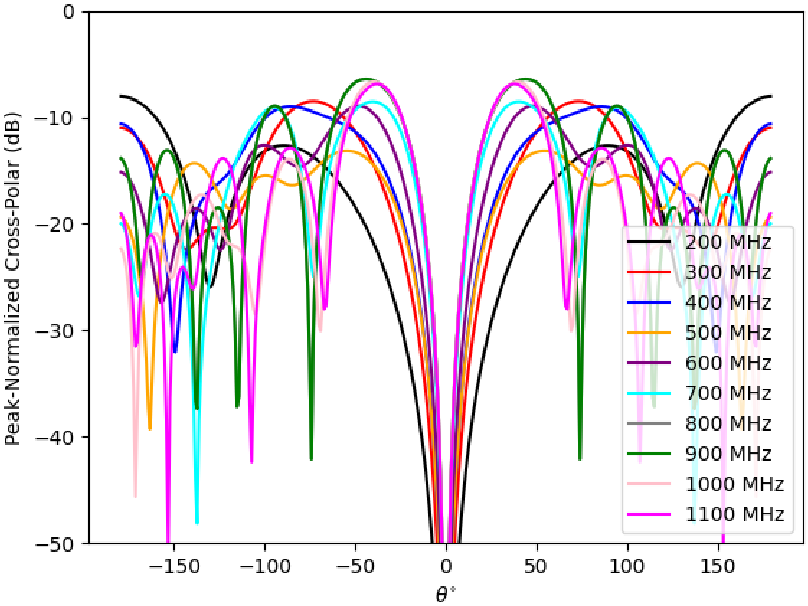

Figure 6.22: Peak-normalized cross-polar beam patterns of the proposed feed in the D-plane.

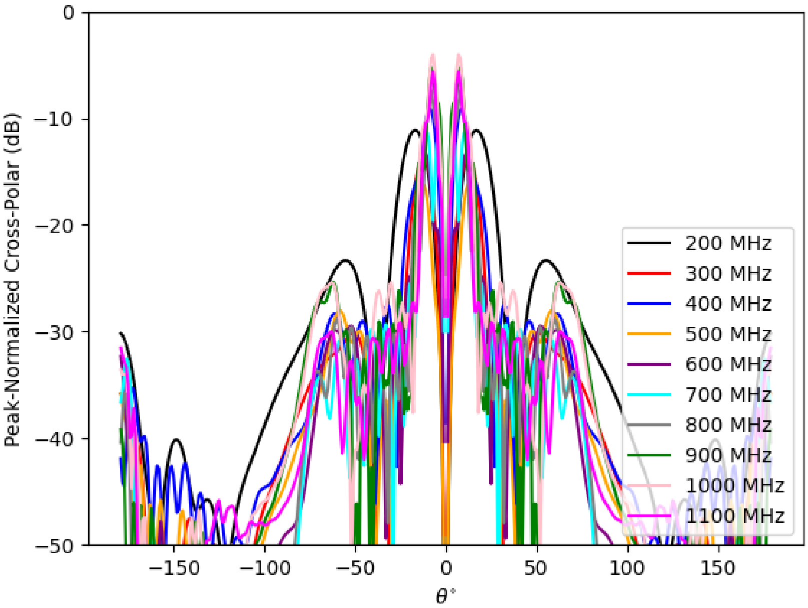

Figure 6.23: Peak-normalized cross-polar beam patterns of the proposed antenna system in the D-plane.

Figure 6.23 shows peak-normalized cross-polar beam patterns for the proposed antenna system, including both dish and feed. In this figure, one finds cross-polar beam patterns approaching 0 at angles close to the boresite These patterns peak at angles close to the edge of the main-lobe, with the sharpest peaks located at the upper end of the band.

The most important aspect to consider for this paper, however, is how this cross-polarized beam pattern may contribute to polarization leakage in the context of an intensity mapping interferometer. One can naively try to eliminate polarized emission from one's data by averaging visibilities from perpendicularly polarized feeds:

$$V_I = \frac{1}{2}(V_{XX} + V_{YY}). \tag{6.17}$$

In this formula, V refers to visibility and X, Y refer to polarizations of the feeds in the baseline. We approximate contributions of unpolarized and polarized emissions to V_I in the manner of Shaw et al. (2015)

$$R_{I\to I} = (E_x^a E_x^{b*} + E_y^a E_y^{b*})P_{ab}^I. \tag{6.18}$$

In this formula, P_{ab}^I is a matrix selecting for the I Stokes parameter. E_x^a refers to polarization component a of the electric field pattern of an x-polarized feed. Similarly, the contribution of

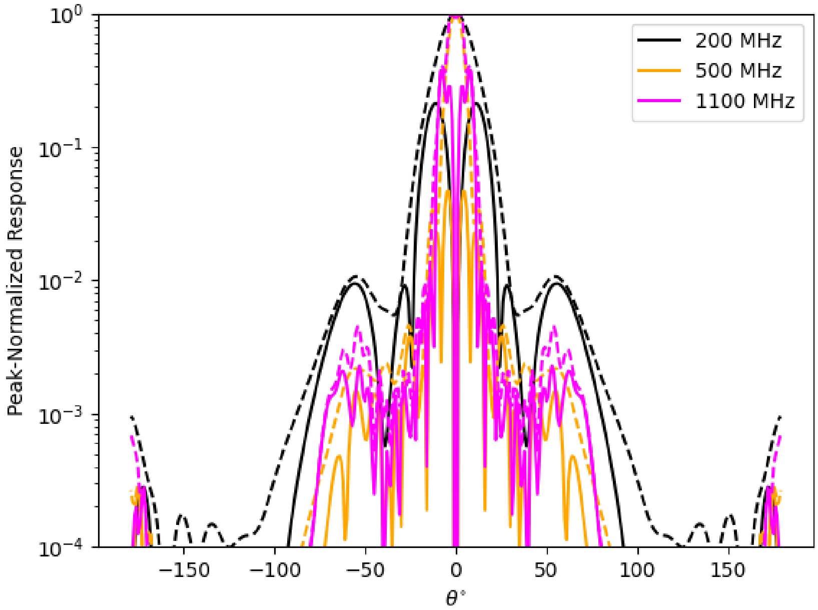

Figure 6.24: Peak-normalized R values for three frequency channels. Solid lines show $R_{P\to I}$ and dashed lines show $R_{I\to I}$. These curves show leakage in the D-plane.

polarized emission is given by

$$R^2_{P\to I} = \sum_{p\in\{Q,U,V\}} |(E^a_x E^{b*}_x + E^a_y E^{b*}_y)P^p_{ab}|^2. \tag{6.19}$$

In Figure 6.24, we present results of leakage calculations for a few frequency channels. These curves are shown in the D-plane and are peak normalized to the maximum value of $R_{I\to I}$. In all plots, one finds no polarization leakage along the boresite. However, there do exist significant levels off-axis, with the most severe case being 1100 MHz with leakage values around 0.4 of the peak copolar value. These levels of polarization leakage are comparable to those estimated for the CHIME telescope in Shaw et al. (2015).

6.6 Fabrication and Testing of the Feed Antenna

After completing the design phase of the project, we fabricated a version of the minimum systematics feed. This feed is roughly 0.9 m long and 1.2 m in diameter at its widest point. Such a large design would not fit in the near-field antenna testing chamber available to us, so a $\frac{4}{13}$ scale version was fabricated instead. In simulations, we found no significant performance difference between the $\frac{4}{13}$ scale design and the original design.

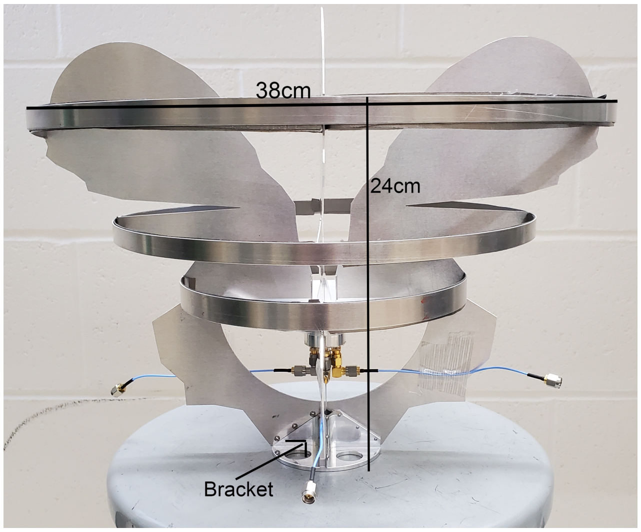

Figure 6.25: A picture of the $\frac{4}{13}$ scale model of the minimum systematics feed. The balun and coaxial cables are present just for testing. In the future, an impedance matching and amplification module would be used instead.

In this section, the fabrication of this scaled down feed is described. We also include some discussion on what would be done differently for the fabrication of the full-scale feed. Lastly, we present measurements of the feed's performance and compare these results to simulations.

Fabrication of the Scaled Feed

A photo of the fabricated feed antenna can been seen in Figure 6.25. During the fabrication of this feed, a 0.1 mm tolerance was requested on all dimensions given. This includes the positions x_i, z_i of points on the profile, differential line pin radius a_{pin}, etc.

The ridges were cut from 1 mm sheet aluminum using a laser cutter. Such an approach may work for the full-sized feed, where the ridges are $\approx 1/8$" thick. However, it may be better to use a water-jet cutter for the full-sized feed.

The cone and differential line shielding were cut from a single block of aluminum. The cone was cut from an aluminum block rather than rolled due to the difficulty of rolling such a thin sheet. Since welding was not an option, the shielding for the differential line was cut from the same block to avoid having to attach the two pieces. Such difficulties would not be present in the fabrication of the full-sized feed. In that case, the use of thicker aluminum would allow for the welding together of pieces and easier rolling of the cone.

An additional aluminum bracket was added to the feed in order to hold the ridges in place. This piece is located on the back of the ridges close to the differential line. It can be seen in Figure 6.25. Since it is located in the backlobe of the antenna, this piece makes a negligible difference to the S_{11}

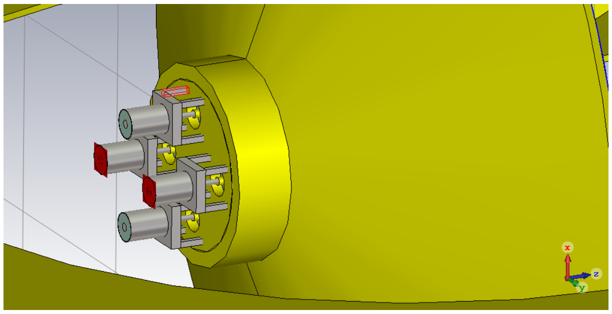

Figure 6.26: A drawing of the balun used to transition to SMA. These SMA ports then plug into the input ports of a 180 degree hybrid.

and beam patterns of the feed.

Another aspect to note here is the different absorber used. In the full scale version, Fair-Rite flexible ferrite tile was intended to be used. This material was mentioned in Subsubsection 6.4. However, data in the frequency range of interest for the scaled down feed could not be found. So, Eccosorb NS-1000 was used instead. In this scaled down design, a layer of absorber 1.2 mm thick was used on the largest ring. A 0.6 mm thick layer absorber was used on the other two rings.

Fabricating a Balun

This feed was designed with a differential line as its output. In order to test it, a balun was required to transition from the differential line to SMA connectors. This was done by using a metal disk shown in Figure 6.26. Note that this disk has holes for each pin of the differential line. The diameter of these holes is chosen to form a 50 Ω transmission line. SMA connectors were then placed on the surface of the disk. The center pins of the SMA connectors touch the pins of the differential line, forming a connection. These SMA connectors are then connected to a 180° hybrid, converting the differential signal to an unbalanced one.

This simple balun works, and impacts the beam patterns negligibly. However, it does not provide a particularly good impedance match to the feed. This is acceptable for now, since the main goal was to test whether or not the CST simulations are accurate. In the future an impedance matching circuit similar to HERA's FEM could be used (Fagnoni et al. (2021)).

Testing

In order to make sure our CST simulations were accurate, the S_{11} of the proposed feed was measured using a network analyzer. The beam patterns were then measured using a StarLab Multiprobe

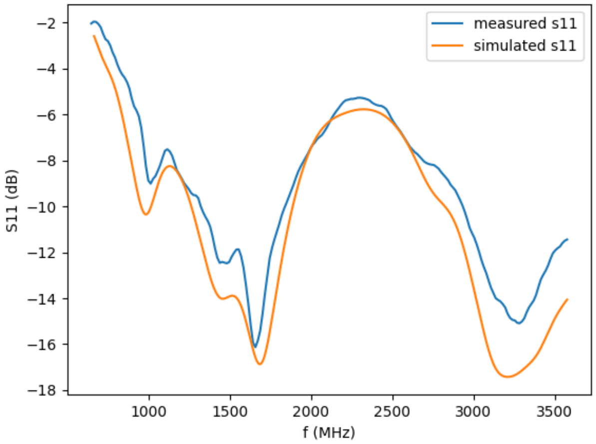

Figure 6.27: Comparison of the measured S_{11} to the simulated S_{11}

System[4] located in an anechoic chamber at UW-Madison.

The resulting S_{11} can be seen in Figure 6.27. The simulation and measurements here were both done using the $4/13$ scale version of the feed and the simple balun. This S_{11} is reasonably close to the simulated values, being within $1-2$ dB over the entire band. One will note that these values for the S_{11} are quite high. This is due to the presence of the simple balun in both the simulation and the measurement, which degrades the impedance match. This balun connects a $200\,\Omega$ differential line to a $50\,\Omega$ coaxial line without any attempts at impedance matching. These values would surely improve if one connected the antenna to an output better matched to the differential line.

Some examples of measured beam patterns can be seen in Figure 6.28 and Figure 6.29. One feature to note is that the patterns are not quite symmetric about $\theta = 0$. This is expected due to fabrication errors and the presence of the 180° hybrid inside the anechoic chamber. The only feature that is somewhat concerning is the poorer taper at low frequencies in the H-plane. Having the beam patterns taper off adequately at the rim of the dish is an important feature for reducing mutual coupling. Otherwise, the measured beams appear to be quite close to those produced by our simulations.

6.7 Conclusion

In this paper, we proposed an antenna design optimized for use in PUMA, a proposed HI intensity mapping observatory. The chief performance requirements were to achieve a smooth frequency

---

[4] Manufactured by Satimo `http://www.mvg-world.com/en/products/fieldproductfamily/antenna-measurement-2/starlab`

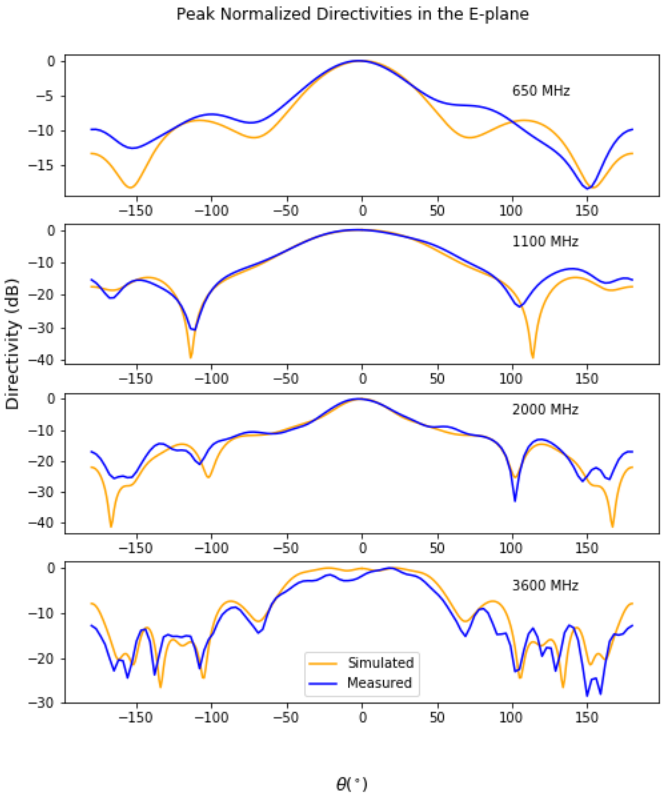

Figure 6.28: Measurements of E-plane gains for the scaled down feed.

response, low levels of mutual coupling between antennas in the array, and sensitivity comparable to that of the fiducial PUMA antenna described in Ansari et al. (2019).

An important performance goal was to have the delay-space power kernel of the instrument drop by 50 dB as quickly as possible. This goal was set in order to minimize the leakage of Galactic foregrounds beyond the foreground wedge. For the proposed design, it was found that the power kernel dropped to the −50 dB level at a delay corresponding to $k_{||} \approx 0.1\,\mathrm{hMpc}^{-1}$. This level was set mostly by feed-dish reflections, with the power kernel of the feed alone dropping by 50 dB after a delay corresponding to $k_{||} \approx 0.03\,\mathrm{hMpc}^{-1}$. Reducing these reflections would require using a shallower dish or perhaps an off-axis design. However, these design choices would worsen delay spectrum performance by increasing mutual coupling. Thus, the use of a deep dish with a collar appears to be the best option for maximizing spectral smoothness.

The next performance goal was to minimize mutual coupling. It is important to minimize this coupling due to its effects on the delay spectrum, redundant calibration, and the risk of introducing correlated noise. Based on the simple model shown in Kern et al. (2019), the goal was to keep the level of coupling below −100 dB for the entire band. This goal was not achieved, but significant progress was made through optimization. The proposed antenna has coupling at the −60 dB level at

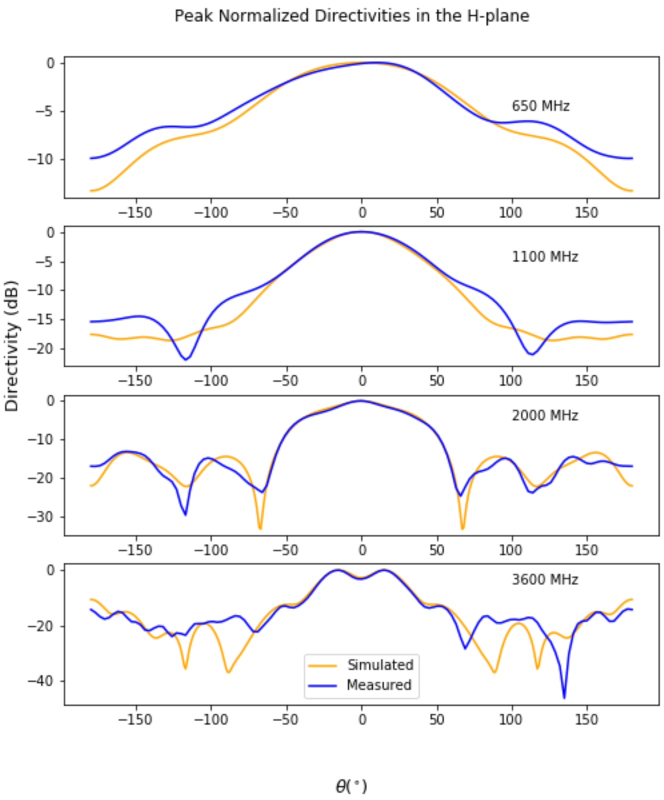

Figure 6.29: Measurements of H-plane gains for the scaled down feed.

low frequencies for closely packed antennas located in the each other's H-plane. Moreover, -100 dB coupling was achieved in the E-plane for frequencies $\nu > 600$ MHz.

We also sought to achieve sensitivity close to that provided by PUMA's fiducial antenna parameters. Our proposed feed satisified this goal, giving slightly lower sensitivity at lower redshifts and slightly higher sensitivity at higher redshifts.

The design presented here was optimized with the systematic effects of HI intensity mapping instruments in mind. This design achieved comparable delay spectrum performance to the HERA Vivaldi feed (Fagnoni et al. (2021)) and provided sensitivity comparable to the fiducial parameters provided for in PUMA telescope proposal. The optimization process for the antenna improved mutual coupling significantly, but the level achieved for the lower end of the band still did not meet the goal. It may be possible to reduce this coupling further using calibration techniques. However, it will likely be necessary to find a way to eliminate this effect in the data analysis stage.

Table 6.1: Summary of a few interferometers intended for 21 cm intensity mapping. Fiducial values for PUMA's aperture efficiency and amplifier temperature are used.

	Antenna	**B**	**Geometry**	$\mathbf{A_{eff}}/elt$
PUMA	6 m parabolic dish, feed on-axis	200-1100 MHz $6 \geq z \geq 0.3$ (Post EOR)	32K tightly packed antennas.	$\approx 20\text{m}^2$
HIRAX	6 m diameter parabolic dish, feed on-axis	400-800 MHz $2.55 \geq z \geq 0.78$ (Post EOR)	1024 elements in a 32×32 tightly packed array.	15.5m^2 at 600 MHz
CHORD	6 m diameter parabolic dish, feed on-axis. Also includes 90 m long 10 m wide cylindrical reflectors with a focal ratio of 0.25.	$300 - 1500\,MHz$ $3.7 \geq z \geq 0$ (Post EOR)	512 tightly packed dishes at the central station. Two outrigger stations placed 1500 km and 3000 km away. These outrigger stations have 64 dishes and 1 cylinder each.	
HERA	14 m diameter parabolic dish, feed on-axis	50-250 MHz $27 \geq z \geq 4.7$ (EOR)	350 dishes tightly packed in a hexagonal array. 32 outrigger antennas placed several hundred meters from the center of the array.	$\approx 100\text{m}^2$ at 150 MHz

Table 6.2: The primary performance requirements for the antenna described in this paper.

Aspect of Performance	**Goal**
Delay Response	The antenna response at zenith decays by 50 dB as quickly as possible in the delay domain. As a realistic goal, we wanted to keep the $50\,dB$ half width $\lesssim 0.10hMpc^{-1}$ when described in k-space.
Mutual Coupling	The S_{21} of two antennas separated by 8 m is at or below -100 dB when the antennas are co-located in each other's E-plane or H-plane.
Impedance Match	The feed has $S_{11} \leq -10$ dB. The final design presented will be relative to 200 Ω.
Power Spectrum Sensitivity	Power spectrum sensitivity is comparable to that achieved by the fiducial PUMA parameters (Ansari et al. (2019)). This is achieved via a low S_{11} and very high spillover efficiency ($\eta_{sp} \approx 0.99$).
Bandwidth	All of these requirements are satisfied over the entire PUMA band,200 to 1100 MHz.

7 NEEDLET KARHUNEN-LOÈVE (NKL): A METHOD FOR CLEANING FOREGROUNDS FROM 21CM INTENSITY MAPS

In this chapter, I show the paper I wrote on a foreground cleaning technique I came up with called Needlet Karhunen Loève (NKL). My goal for this paper was to come up with a Karhunen-Loève-transform based foreground cleaning technique that was robust against inaccurate calibration. I did achieve this goal, though it didn't end up providing as much of a performance advantage as I had hoped. I now present a slightly truncated version of the paper "Needlet Karhunen-Loève (NKL): A Method For Cleaning Foregrounds From 21cm Intensity Maps" published in volume 527 of Monthly Notices of the Royal Astronomical Society in 2023. This paper was co-authored by Peter Timbie and I.

7.1 Abstract

This paper introduces a technique called NKL, which cleans both polarized and unpolarized foregrounds from HI intensity maps by applying a Karhunen-Loève transform on the needlet coefficients. In NKL, one takes advantage of correlations not only along the line of sight, but also between different angular regions, referred to as "chunks". This provides a distinct advantage over many of the standard techniques applied to map-space that one finds in the literature, which do not consider such spatial correlations. Moreover, the NKL technique does not require any priors on the nature of the foregrounds, which is important when considering polarized foregrounds. We also introduce a modified version of GNILC, referred to as MGNILC, which incorporates an approximation of the foregrounds to improve performance. The NKL and MGNILC techniques are tested on simulated maps which include polarized foregrounds. Their performance is compared to the GNILC, GMCA, ICA and PCA techniques. Two separate tests were performed. One at $1.84 < z < 2.55$ and the other at $0.31 < z < 0.45$. NKL was found to provide the best performance in both tests, providing a factor of 10 to 50 improvement over GNILC at $k < 0.1hMpc^1$ in the higher redshift case and $k < 0.03hMpc^1$ in the lower redshift case. However, none of the methods were found to recover the power spectrum satisfactorily at all BAO scales.

7.2 Introduction

One of the main endeavors of observational cosmology is to measure statistical properties of the spatial distribution of matter in the Universe. Such measurements are of interest as they could provide us with information about dark energy, inflationary physics, the growth of structure, early star formation and more. The distribution of matter in the Universe is traced by neutral hydrogen (HI). Using 21 cm emission or absorption to measure these density fluctuations as a function of redshift is a technique referred to as 21 cm (or HI) intensity mapping (IM) (Bharadwaj and Sethi (2001)) (Battye et al. (2004)) (Madau et al. (1997)). Most IM instruments are radio interferometers; some notable examples are the Canadian Hydrogen Intensity Mapping Experiment (CHIME) (Bandura et al. (2014)), Tianlai (Chen (2012)), the Hydrogen Epoch of Reionization Array (HERA) (DeBoer and HERA (2015)) and the Murchison Widefield Array (MWA) (Morales (2005)). The antenna elements that make up these interferometers can take several forms, including parabolic-dishes (Tianlai, CHIME, HERA), cylindrical reflectors (CHIME, Tianlai) or phased arrays (MWA). Single-dish instruments are also used, with some examples being the Five hundred meter Aperture Spherical Telescope (FAST) (Bigot-Sazy et al. (2016)), the Greenbank Telescope (GBT) (Switzer et al. (2013)) and the More Karoo Array Telescope (MeerKAT) (Santos et al. (2016)). FAST and GBT both consist of a single, very large dish. On the other hand, MeerKAT would average signals from 64 smaller (13.5 m diameter) dishes. MeerKAT can also operate as an interferometer.

Measurement of the 21 cm line would provide valuable cosmological information. At low redshifts ($z \leq 6$), IM would serve as a complement to galaxy redshift surveys. At such redshifts, cosmologists would be particularly interested in measuring Baryon Acoustic Oscillations (BAO). The BAO would serve as a standard ruler, providing information about the expansion of the Universe and dark energy. Measurements during the epoch of reionization $6 < z < 20$ would provide information about the formation of the earliest stars. Lastly, measurements during the cosmic dark ages ($20 < z < 1100$), before the formation of the first luminous objects, would provide insights into the physics of inflation. It should be noted that no other probe aside from IM is capable of mapping the cosmic dark ages.

Although promising, IM is still a young technique, with its systematic effects and calibration requirements still being understood. In particular, foregrounds from Galactic synchrotron emission are an unsolved challenge for IM experiments. In the case of Galactic synchrotron emission, these foregrounds are up to a factor of 10^5 brighter than the HI signal.

Luckily, the unpolarized component of these foregrounds is expected to be spectrally smooth, while the HI signal is expected to have a high level of chromaticity. In other words, the foregrounds are confined to a relatively small subsection of delay-space (Datta et al. (2010)). In this context, delay refers to the Fourier dual of spectral frequency. This delay-space quarantining aids in the avoidance and removal of the foregrounds. Unluckily, this region of delay space will correspond to large spatial scales, which are important for measuring the BAO.

One must also consider the polarized component of Galactic synchrotron radiation. This component is affected by Faraday rotation in the Galaxy, introducing higher chromaticity than the unpolarized component. Moreover, the chromaticity of this component is expected to vary with line of sight direction. In particular, lines of sight closer to the Galactic plane will suffer from more severe chromaticity due to stronger Faraday rotation. Moreover, the power of the Galactic synchrotron radiation (both polarized and unpolarized components) relative to the signal is expected to vary with angular scale (Alonso et al. (2014)), with the largest scales having the worst contamination. In principle, these polarized foregrounds could be avoided altogether as long as the beams of the telescope have rotational symmetry and low cross-polar levels. However, such refined beams are not achievable in practice, and one must contend with some amount of polarization leakage. This leakage is expected to be on the percent level for IM experiments (Moore et al. (2013)) (Moore et al. (2017)).

Based on the review of the literature provided in chapter 5 of this thesis, there are several tasks that ought to be performed.

1. Create a foreground cleaning method that is usable in map space and uses both frequency and angular correlations. This method should not require a prior on the foregrounds.
2. Test the available map-based methods on polarized foregrounds at redshifts $z \gtrsim 0.6$.
3. Test the robustness of various map-based methods against beam mis-calibration.

We make progress on task (i) by introducing the Needlet Karhunen-Loève (NKL) method of foreground removal. In NKL, different sections of pixel/spherical harmonic space are separated via a needlet transform. The needlet coefficients are then cleaned using a Karhunen-Loève transform that exploits both angular and frequency correlations. This is different from most other map-based methods, which only consider frequency correlations. Moreover, this is all done without needing any priors on the foregrounds present in the map. Not requiring a foreground prior is quite useful, as polarized foregrounds are not well understood. Moreover, NKL cleans the foregrounds differently depending on location and angular scale. This is also a desirable feature, as the foreground brightness and chromaticity are expected to vary with line of sight and angular scale.

In this paper, we also make some progress on task (ii) of the list by testing GNILC, GMCA, ICA and PCA at redshifts ($1.84 < z < 2.55$). PCA, ICA and GMCA were chosen as these are all prominent in the literature. GNILC was chosen since it has similarities to the NKL method proposed in this paper.

We leave task (iii) for future work.

In Section 7.3, we describe the process of foreground cleaning in an abstract way and provide additional motivation for the NKL technique. In Section 7.4, we provide a detailed description of the NKL technique. In this Section, we also introduce Modified GNILC (MGNILC), a slightly modified version of GNILC which uses a foreground approximation acquired from the data. In Section 7.5, we describe the maps we used to perform our tests. In Section 7.6, we present results acquired by testing GNILC, ICA, PCA, GMCA and NKL on simulated maps. We then summarize our results and conclude in Section 7.7.

7.3 Foreground Removal Techniques

In this paper, we will compare the performance of the proposed NKL technique with that of other techniques commonly found in the literature. In particular, we will be considering Principal Component Analysis (PCA) (Cunnington et al. (2021)), Independent Component Analysis (ICA) (Wolz et al. (2014)) , Generalized Morphological Component Analysis (GMCA) (Carucci et al. (2020)), and Generalized Needlet Internal Linear Combination (GNILC) (Olivari et al. (2015)). Although not quite obvious at first, these techniques are in fact very similar. In particular, all techniques tested in this paper will model the foregrounds (or signal plus noise in the case of GNILC) as a mixture of template maps.

Foreground Removal Using Templates

All techniques considered in this paper begin by assuming that the maps produced by an IM experiment are given by

$$\mathbf{X} = \mathbf{f} + \mathbf{h} + \mathbf{n}. \tag{7.1}$$

In this formula, $\mathbf{X}$ is a $n_{ch} \times n_p$ matrix, where n_{ch} is the number of frequency channels and n_p is the number of pixels in each map. Moreover, $\mathbf{f}$ represents the foregrounds, $\mathbf{h}$ represents the HI signal and $\mathbf{n}$ represents the noise. In this paper we will use boldt font to denote matrices. We will also define

$$\mathbf{s} = \mathbf{h} + \mathbf{n}. \tag{7.2}$$

It is then assumed that the foregrounds can be expressed as a mixture of templates:

$$\mathbf{f} = \mathbf{AS}. \tag{7.3}$$

In this formula, $\mathbf{S}$ is a $n_t \times n_p$ matrix of templates, where n_t is the number of templates and $\mathbf{A}$ is a $n_{ch} \times n_t$ 'mixing matrix' which encodes how the templates are combined at each frequency channel. The foreground removal process then becomes a matter of fitting $\mathbf{AS}$ to $\mathbf{X}$, subject to some sort of regularization.

In PCA, one seeks to find templates that capture as much variation in the data as possible. In particular, these templates are really just the dot product of the data with the eigenvectors corresponding to the largest n_t eigenvalues of the covariance matrix. In this context, the covariance matrix is usually generated assuming that the brightness of the sky along each line of sight were independently drawn from some distribution. This assumption is not true, but is close enough to reality that PCA can still provide reasonable results. In the context of 21 cm intensity mapping, one usually estimates the covariance from the data as

$$\hat{\mathbf{C}} = \frac{1}{n_p}(\mathbf{X} - \overline{\mathbf{X}})(\mathbf{X} - \overline{\mathbf{X}})^T, \tag{7.4}$$

where row i of $\overline{\mathbf{X}}$ is the average of row i of $\mathbf{X}$. In this paper, we will use hats to denote covariance estimates. On the other hand, we will use matrices without hats to denote true covariances. In this case, $\hat{C}$ is a $n_{ch} \times n_{ch}$ matrix. In ICA, one seeks templates that are statistically independent. In GMCA, one assumes that the templates $\mathbf{S}$ ought to be sparse in some wavelet domain. ILC methods such as GNILC seek to find a filter that has a unit response to $\mathbf{h} + \mathbf{n}$ while minimizing the variance of the residuals (Olivari et al. (2015)). Readers can refer to Marins et al. (2022) for a more rigorous description of how these standard techniques are regularized.

Both Shaw's work and the Needlet Karhunen-Loève (NKL) technique described in this paper make use of the Karhunen-Loève (KL) transform (Tegmark (1997)) for foreground cleaning. The KL transform uses the covariance matrices of both the foregrounds and signal to clean the data. These covariances either come from priors or are estimated from the data itself. For NKL, we estimate the foreground covariance from the data and get the HI and noise covariances from priors. Let $\mathbf{C}_{FG}$ be the foreground covariance and $\mathbf{C}_S$ be the signal covariance. For now, let's consider only frequency correlations, resulting in matrices of size $n_{ch} \times n_{ch}$. In practice, the covariances used in NKL will have larger dimension since they will include angular correlations as well. However, this discussion will be clearer if we consider only the $n_{ch} \times n_{ch}$ case, and a mixing matrix plus template model can still be used to describe the case involving angular correlations.

The signal model $\mathbf{C}_S$ includes the statistics of whatever components the analyst would like to recover from beneath the foregrounds. In this paper, we will take $\mathbf{C}_S = \mathbf{C}_{HI}$ unless otherwise noted. It should be noted that the freedom to choose $\mathbf{C}_S = \mathbf{C}_{HI}$ will provide NKL with an advantage over the other techniques mentioned in this paper. These techniques make no distinction between the HI and noise. The KL technique works by solving the generalized eigenvalue problem

$$\mathbf{C}_{FG}\mathbf{\Phi} = \mathbf{C}_S\mathbf{\Phi}\mathbf{\Lambda}. \tag{7.5}$$

In this formula, $\mathbf{\Phi}$ is a matrix of eigenvectors and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues. It turns out that the eigenvectors $\mathbf{\Phi}$ obtained are a solution to the optimization problem (Ghojogh et al. (2023))

$$\begin{gathered}\max_{\mathbf{\Phi}} \mathbf{tr}(\mathbf{\Phi}^T\mathbf{C}_{FG}\mathbf{\Phi}) \\ \text{subject to } \mathbf{\Phi}^T\mathbf{C}_S\mathbf{\Phi} = \mathbf{I}.\end{gathered} \tag{7.6}$$

So, the KL transform finds modes that have as high a ratio of foreground to signal as possible. In particular, the eigenvalues λ_i indicate the expected ratio of foreground to signal power at that particular mode. One then cleans the data by removing modes that are foreground dominated. In the language of templates and mixing matrices, we find that the template matrix is

$$\mathbf{S} = \mathbf{P}_s(\mathbf{X} - \overline{\mathbf{X}}). \tag{7.7}$$

In this formula, $\mathbf{P}_s$ is a $n_t \times n_{ch}$ matrix whose rows are the foreground dominated eigenvectors. Moreover, let the symbol $\mathbf{P}$ to represent a $n_{ch} \times n_{ch}$ matrix whose rows are the eigenvectors generated by the generalized eigenvalue problem. The mixing matrix is then

$$\mathbf{A} = \mathbf{P}_s^{-1}, \tag{7.8}$$

where $\mathbf{P}_s^{-1}$ is a $n_{ch} \times n_t$ matrix containing only columns of $\mathbf{P}^{-1}$ that correspond to foreground dominated modes.

Discussion of Techniques

Before actually testing any techniques, it will be beneficial to discuss differences between the techniques considered in this paper. In particular, we will discuss these differences and try to provide some motivation for why certain techniques may provide better performance than others.

First, one should note that GNILC requires a model for the HI and noise, making it a “non-blind" technique. This is different from GMCA, PCA and ICA, where the only free parameter provided by the user is the number of templates to use.

Another aspect of these techniques to consider is locality. In the standard version of GMCA, the matrices **A** and **S** are meant to capture the foregrounds at all pixels and angular scales (Carucci et al. (2020)). PCA and ICA are typically conducted in a similar way, where one estimates the covariance **C** using the entire dataset. On the other hand, GNILC cleans the maps in a more fine-grained way. In particular, it divides the data into needlet coefficients, and then cleans the data one coefficient at a time. This allows for the foregrounds to be cleaned differently depending on the location and angular scale in question.

One can imagine reasons why such a fine-grained treatment might provide advantages. For instance, the ratio of foreground to HI power is expected to vary with angular scale. In particular, large angular scales will suffer worse contamination than smaller ones. When working with unpolarized foregrounds, one would expect similar chromaticity at all lines of sight. However, when dealing with polarized foregrounds, we expect for the chromaticity to vary with line of sight. As such, one would expect for pixels close to the Galactic plane to require more templates to clean than ones far from the plane. Thus, it seems likely that a more fine-grained approach would work better when dealing with polarized foregrounds.

It is certainly possible to imagine changing PCA, ICA or GMCA to make them more local. For instance, one could implement a scale- and location- dependent version of PCA where the covariance matrix is estimated for some neighborhood around each needlet coefficient. In addition, a technique called L-GMCA has been proposed (Bobin (2017)) in which one uses different mixing matrices for different regions of the map. For now though, we will consider only the global versions of these techniques.

So, we see reasons why a more localized approach would likely provide better results when cleaning foregrounds from maps. So far, GNILC is the most local approach that has been found in the literature. In Olivari et al. (2015), this approach was tested on low-z simulated maps without any polarization leakage included (Olivari et al. (2015)). This study found GNILC outperforming PCA at angular scales ($\ell > 30$). However, GNILC provided worse performance than PCA at scales larger than that. Interestingly, Carucci et al. (2020) found that GMCA also struggled at large angular scales $\ell < 50$, at least in the case when polarized foregrounds were included.

In chapter 7, I introduce NKL. Similarly to GNILC, NKL is a non-blind technique which acts on needlet coefficients. However, while GNILC only considers statistics along the line of sight, NKL also considers angular correlations in needlet space.

7.4 Implementing NKL

In this section, we introduce the NKL technique for removing foregrounds from 21 cm maps. We will begin this Section by providing background knowledge required for understanding how NKL works. In particular, the first Subsection introduces needlets. In the second, we describe the way in which needlet coefficients are partitioned before cleaning is performed. Then, we describe the ways in which we can generate an approximation of the foregrounds from the data. In the final subsection, we provide a list of steps for performing NKL.

The Needlet Transform

The NKL process begins by performing a needlet transform on each frequency slice of the 3D maps. Needlets are wavelet-like functions that have a finite width in both ℓ-space and pixel-space. These functions were first presented in Narcowich et al. (2006). For our purposes, we computed these coefficients using the `pys2let` software package, details of which can be found in Leistedt et al.

(2013). Needlets are defined via

$$\psi_{jk}(\hat{n}) = \sqrt{\lambda_{jk}} \sum_l b\left(\frac{l}{B^j}\right) \sum_{m=-l}^{l} \overline{Y}_{lm}(\hat{n}) Y_{lm}(\xi_{jk}). \tag{7.9}$$

In this formula, $b\left(\frac{l}{B^j}\right)$ is a bandpass function that is non-zero for $B^{j-1} \leq l \leq B^{j+1}$. In this paper, we choose $B = 2$ and use $4 \leq j \leq 8$. The very lowest ℓ modes are described using a “scaling function" (Leistedt et al. (2013)). The variable ξ_{jk} refers to the line of sight at which the needlet is centered. In this paper, ξ_{jk} refers to the location of pixel k in a HEALPix map with $nside = 2^{j+1}$. The variable λ_{jk} then refers to the solid angle of the HEALPix pixel in question. Needlet coefficients can then be obtained simply by computing the following integral

$$\chi_{jk} = \int T(\hat{n}) \psi_{jk}(\hat{n}) d\Omega. \tag{7.10}$$

In this formula, $T(\hat{n})$ is the function for which the needlet coefficients are being computed. In this work, $T(\hat{n})$ will be the sky temperature. One can then reconstruct their temperature map using these coefficients as follows

$$T(\hat{n}) = \sum_{jk} \chi_{jk} \psi_{jk}(\hat{n}). \tag{7.11}$$

Marinucci et al. (2007) provides more detail on needlets and their use in cosmology.

Partitioning of the Needlet Coefficients

An important part of NKL is the partitioning of the needlet coefficients. We refer to the groups of partitioned coefficients as “chunks". In Fig. 7.1, we present an example of a chunked map. For now, let’s suppose we are partitioning the coefficients for needlet scale j. Let’s package the needlet coefficients at this scale of interest in a matrix $\boldsymbol{\chi}^{(j)}$. This matrix will have dimensions $n_{ch} \times n_{co}^{(j)}$, where n_{ch} is the number of frequency channels and $n_{co}^{(j)}$ is the number of needlet coefficients per frequency channel at scale j.

Next, let $\boldsymbol{\chi}_i^{(j)}$ be the row of $\boldsymbol{\chi}^{(j)}$ corresponding to frequency channel i. The process of partitioning the row into N chunks begins by selecting the coefficients located at columns $q = a * n_{co}^{(j)}/N$ for $a = 1, 2, ..., N$. Note here that q is indexing pixels in the HEALpix map. Let’s refer to these coefficients as “anchors". Next, we assign all the other coefficients in the row to chunks according to which anchor point they have the smallest angular separation from. Adjacent neighborhoods then swap coefficients until all contain the same number of coefficients.

We then use the same chunk assignments for all rows in $\boldsymbol{\chi}^{(j)}$. This results in a new “chunked" matrix $\underline{\boldsymbol{\chi}}^{(j)}$ which has dimensions $n_{ch}N \times n_{co}^{(j)}/N$. One row of $\underline{\boldsymbol{\chi}}^{(j)}$ will contain the coefficients of

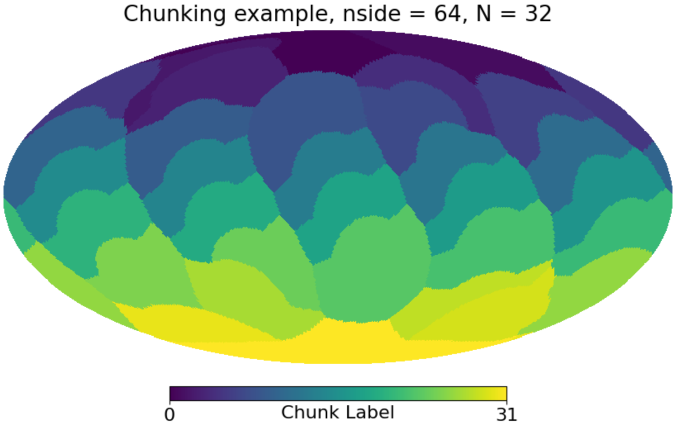

Figure 7.1: An example of dividing a needlet map into chunks.

one chunk at one frequency channel. Column q will contain the q^{th} element from each chunk at all frequency channels.

Producing a Foreground Approximation

In order to perform a Karhunen-Loève transform, we need estimates for $\mathbf{C}_{FG}$ and $\mathbf{C}_{HI}$. The estimate of $\mathbf{C}_{HI}$ is provided as a prior chosen by the analyst. For the purposes of this paper, we estimate $\mathbf{C}_{HI}$ from maps simulated by CORA (Shaw et al. (2015)). One may also want to estimate $\mathbf{C}_N$ to include in the KL transformation or for later debiasing. In this paper, we estimate $\mathbf{C}_N$ from the model used to simulate the noise. In a real experiment, this noise covariance could be determined from the data. On the other hand, we seek to estimate $\mathbf{C}_{FG}$ from the data itself. This approach is motivated by two facts. First, as of now, polarized Galactic synchrotron radiation is not well enough measured or modeled to use an a priori covariance. In addition, a method that estimates the foregrounds directly from data will be more robust against errors in calibration and systematics that may cause the data to deviate from what would otherwise be an accurate a priori model. It should be noted that relatively small errors in the model of $\mathbf{C}_{FG}$ can cause significant problems, due to the large dynamic range between the foregrounds and HI. In this paper, two possible approaches to foreground estimation are presented, PCA and DPSS Approximate lazY filtEriNg of foregroUnds (DAYENU) (Ewall-Wice et al. (2020)). In the DAYENU approach, one assumes that the foregrounds are confined within some region of delay space. This assumption still holds in our CRIME generated foreground maps, at least for lines of sight outside the Galactic plane. This assumption can be useful, since it means that the HI for larger delays will not be affected by the filter. This is different from PCA, where some signal loss outside the foreground-dominated region of delay space may occur. Moreover, unlike PCA, the DAYENU method treats each pixel (or needlet coefficient) uniquely. In PCA, one is using the same templates to approximate the foregrounds for all pixels/coefficients. The main weakness of DAYENU, however, is that it does not work when all delays are contaminated by foregrounds.

Estimating Foregrounds Using PCA

This process begins by estimating the frequency-frequency covariance of the needlet coefficient maps. For this paper, we performed PCA foreground estimates one needlet scale at a time and without any chunking. PCA was not performed on chunked maps, as our scheme for selecting the number of components to remove struggled in this case. It's possible that applying PCA to each chunk individually would provide a better approximation, but we leave this to future testing. Once again, we are making the assumption that each line of sight is drawn independently and identically from some distribution. Since the needlet coefficients at a particular frequency channel have 0 mean, we can estimate their covariance as

$$\hat{\mathbf{C}}^{(j)} = \frac{1}{n_{co}^{(j)}} \boldsymbol{\chi}^{(j)} (\boldsymbol{\chi}^{(j)})^T. \tag{7.12}$$

The only exception to this is for the "scaling function" of the needlet transform, which contains the $\ell = 0$ spherical harmonic mode. For the scaling function, one can just use equation (7.4). Due to their simpler frequency dependence, we will assume that the foregrounds are restricted to a subset of these modes. On the other hand, we will assume that the HI signal and noise ($\mathbf{s}^{(j)}$) will be more evenly spread out throughout all the modes. This naturally divides the eigenmodes into n_{fg} modes which are dominated by the foregrounds and $n_{ch} - n_{fg}$ modes which are dominated by the signal plus noise. Let $\mathbf{P}$ be a matrix whose rows are eigenvectors of $\hat{\mathbf{C}}^{(j)}$. Moreover, let $\mathbf{P}_f$ be a $n_{fg} \times n_{ch}$ matrix containing the rows of $\mathbf{P}$ corresponding to the largest n_{fg} modes of $\hat{\mathbf{C}}^{(j)}$. Next, let $(\mathbf{P}^{-1})_f$ be a $n_{ch} \times n_{fg}$ matrix whose columns are the columns of $\mathbf{P}^{-1}$ which act on the foreground dominated rows of $\mathbf{P}$. A foreground estimate can then be obtained via

$$\hat{\mathbf{f}}^{(j)} = (\mathbf{P}^{-1})_f \mathbf{P}_f \boldsymbol{\chi}^{(j)}. \tag{7.13}$$

The value for n_{fg} can be estimated in different ways. In Zhang et al. (2016), a likelihood ratio test was used. In such a method, one essentially increases the number of parameters used until the resulting power spectrum begins to converge (Zhang et al. (2016)). Alternatively, one can use their models of the HI and noise, along with the Akaike Information Criterion (AIC), to estimate n_{fg}. This was the approach taken in Olivari et al. (2015). The AIC is given by

$$AIC = 2k - 2ln(L), \tag{7.14}$$

where k is the number of parameters in the model and L is the maximum likelihood value of the model. The likelihood function used in this paper can be found in the appendix of Olivari et al. (2015). Thus, by minimizing the AIC, one rewards goodness of fit (via the second term) while also discouraging overfitting (via the first term). More information on the use of AIC in a context such as this can be found in Olivari et al. (2015). In this paper, the number of modes chosen in any given situation will always be selected using AIC unless otherwise noted.

Estimating Foregrounds Using DAYENU

Another possible approach to estimating the foregrounds is to use the DAYENU model from Ewall-Wice et al. (2020). DAYENU models the foreground covariance as simply as possible, using the fact that the foregrounds should be highly concentrated at delays close to $\tau = 0$. In particular, DAYENU will model the foreground covariance in delay space as

$$\tilde{C}^{\daleth}_{FG}(\tau, \tau') = \epsilon^{-1} \frac{1}{2\tau_w} \delta(\tau - \tau') \qquad |\tau| < \tau_w. \tag{7.15}$$

In this equation, $2\tau_w$ is the assumed full width of the foregrounds in delay space and ϵ^{-1} is their assumed magnitude.

At a high level, our DAYENU-based scheme approximates the foregrounds at some line of sight in needlet space using the following steps:

1. Estimate the width of the foregrounds in delay-space for the needlet coefficient of interest by comparing the delay transform to priors. The full width will correspond to $2\tau_w$.
2. Use this width to generate the foreground covariance model $C^{\daleth}_{FG}(\nu, \nu')$. We would like to emphasize here that this foreground covariance is estimated one line of sight at a time. It is also not the same as the foreground covariance that will be eventually used in the KL step of NKL.
3. Project the needlet coefficient onto the image of $C^{\daleth}_{FG}(\nu, \nu')$. This acts as a bandpass filter in delay-space, selecting the foreground-dominated region.
4. Perform steps 1 through 3 at all coefficients in the needlet map. This will provide a foreground approximation $\hat{\mathbf{f}}^{(j)}_k$ from which one can estimate $\hat{\mathbf{C}}^{(j)}_{FG}$.

This list has given a high level explanation of the DAYENU-based scheme. Let's now consider the details.

First, recall the DAYENU covariance definition provided in equation (7.15). In this formula, will use the value of τ_w acquired in the previous steps and will set $\epsilon = 1$ for now. We will account for the magnitude of the foregrounds later on in this work.

In frequency space, the DAYENU foreground covariance will be given by

$$C^{\daleth}_{FG}(\nu, \nu') = \mathrm{sinc}(2\pi\tau_w(\nu - \nu')). \tag{7.16}$$

When frequency samples are evenly spaced, $C^{\daleth}_{FG}(\nu, \nu')$ will be diagonalized by Discrete Prolate Spheroidal Sequences (DPSS) (Slepian (1978)). Suppose one is dealing with sequences of length N_d. The DPSS sequences $\{u^\alpha(N_d, W) \mid \alpha = 1, 2, 3, \ldots, N\}$ will form an orthonormal basis that maximizes energy in a region $[-W, W]$ of the DFT domain. We call this "spectral concentration" and denote it as s_c. More precisely, the spectral concentration for some sequence is

$$s_c = \frac{\int_{-W}^{W} |U(x)|^2 dx}{\int_{-1/2}^{1/2} |U(x)|^2 dx}. \tag{7.17}$$

In this formula, x is the variable of the DFT domain, which in our context is delay. The variable U represents the DFT of u, centered at $(N_d - 1)/2$. u^1 is defined as a unit norm sequence that maximizes s_c. Then, u^2 is created by finding a unit norm sequence that maximizes s_c while being orthogonal to u^1. In general, u^n is created by finding a unit norm sequence that maximizes s_c while being orthogonal to u^i for $i = 1, .., n - 1$.

In the case of DAYENU, the DPSS sequences will maximize energy at delays $-\tau_w < \tau < \tau_w$.

The DAYENU model can then be used to estimate the foregrounds for each needlet coefficient in the following way. For the following steps, let ${}_{,k}^{(j)}$ be a $n_{ch} \times 1$ vector containing needlet coefficients for all channels at scale j and pixel k. The procedure to estimate the foregrounds at $\chi_k^{(j)}$ is as follows.

1. Use priors on the HI and noise to generate mock needlet coefficient maps. These mock maps will be denoted by $\mathbf{h}^{m,(j)}$ and $\mathbf{n}^{m,(j)}$ for the HI and noise respectively. Moreover, let $\mathbf{y}^{m,(j)} = \mathbf{h}^{m,(j)} + \mathbf{n}^{m,(j)}$.

2. Randomly select N_p lines of sight from these mock coefficient maps. Then, apply two Blackman-Harris windows and an FFT to bring them into the delay domain. Moreover, the data should be shifted to place $\tau = 0$ at the center sample. We will denote such a transform with a tilde, i.e.

   $$\tilde{a} = SHIFT[FFT[aW^2]], \tag{7.18}$$

   where W denotes a four term Blackman-Harris window of appropriate length. Let $\mathcal{D}$ be a set containing the randomly chosen pixels.

3. For each $i \in \mathcal{D}$, perform a curve fit on $abs(\tilde{\mathbf{y}}_i^{m,(j)})$. This curve fit is meant to give one an estimate of the typical level of $abs(\tilde{\mathbf{y}}_i^{(j)})$ at each delay. It was found that using a third order

polynomial in log-space worked quite well. The curve produced by this fitting process at pixel i will be denoted $r_i^{m,(j)}$.

4. Average the fits obtained at each pixel to produce a typical HI plus noise curve in delay space:

$$\overline{r}^{(j)} = \frac{1}{N_p} \sum_{i \in \mathcal{D}} r_i^{m,(j)}. \tag{7.19}$$

   In this formula, N_p is the number of pixels chosen and $\mathcal{D}$ is the set of pixels used in the approximation.

5. Perform a delay transform on the needlet coefficient of interest, giving $\tilde{\chi}_k^{(j)}$. Next, compare $abs(\tilde{\chi}_k^{(j)})$ with $\overline{r}^{(j)}$. The delay at which the level of $abs(\tilde{\chi}_k^{(j)})$ becomes comparable to that $\overline{r}^{(j)}$ will correspond to the τ_w used in DAYENU. For this paper, we select the delay at which $abs(\tilde{\chi}_k^{(j)})$ is within a factor of 2 of $\overline{r}^j$.

6. Use the τ_w obtained in the previous step to generate a DAYENU covariance as in equation (7.16). This DAYENU matrix will be denoted via $\mathbf{C}^{\urcorner}$.

7. Compute the SVD of $\mathbf{C}^{\urcorner}$, giving $\mathbf{C}^{\urcorner} = \mathbf{USV}^{\dagger}$. This SVD will be used to project onto the image of $\mathbf{C}^{\urcorner}$. Let $\overline{\mathbf{U}}$ be a matrix containing columns of $\mathbf{U}$ for which the singular value is above some threshold. This threshold is determined based on the expected magnitude of the foregrounds. Since the foregrounds can be 5 orders of magnitude brighter than the signal, it is advisable to choose the cutoff to be at or below $10^{-10}s_{max}$, where s_{max} is the largest singular value of $\mathbf{C}^{\urcorner}$. In this particular work, we use the same threshold for all angular scales and all needlet locations. The foregrounds are then approximated via

$$\hat{f}_k^{(j)} = \overline{\mathbf{U}\mathbf{U}}^T \chi_k^{(j)}. \tag{7.20}$$

The NKL algorithm

In the literature, one finds standard foreground removal methods such as PCA, ICA and GMCA being used in a way where one pair of **A** and **S** is generated for the entire data set. One exception to this in the literature is GNILC, which performs the cleaning in a more fine-grained way, sending the data to the needlet domain and then cleaning each needlet coefficient individually. However, this method only takes advantage of frequency-frequency correlations in the data. Similarly to GNILC, NKL performs its cleaning in the needlet domain. The key difference, however, is that NKL also takes advantage of angular correlations in the needlet coefficients. Moreover, instead of using an ILC step, as in GNILC, NKL uses a KL transform. Below are step by step instructions for performing an NKL cleaning on some dataset.

1. Generate an approximation of the foregrounds at the needlet scale of interest. In this paper, foreground approximations are generated using either DAYENU of PCA. This approximation will be denoted via $\hat{\mathbf{f}}^{(j)}$.

2. Divide the needlet coefficients at the scale of interest into N “chunks" of equal size. These chunks must be adequately large to provide accurate covariance estimates. The appropriate chunk size will depend both on the severity of the foregrounds and on the needlet scale in question. This gives $\underline{\chi}^{(j)}$, which was defined in Section 7.4,

3. Divide the foreground approximation into chunks, providing $\underline{\hat{\mathbf{f}}}^{(j)}$. Use $\underline{\hat{\mathbf{f}}}^{(j)}$ to generate an estimate of the foreground covariance at the scale of interest. This covariance can be estimated in the manner shown in equation (7.4). This covariance will be denoted $\hat{\mathbf{C}}_{FG}^{(j)}$ and will be a square matrix of size $n_{ch}N$.

4. Use priors on the signal to create an estimate of $\mathbf{C}_S^{(j)}$. This matrix will have the same shape as the foreground covariance matrix described in the previous step.

5. Use the covariance matrices acquired in the previous two steps to perform a Karhunen-Loève transform on the needlet scale of interest. The steps for this process are described in Section 7.3.

Deviations From the Ideal Case

The NKL technique is based on the KL transform, which is described in equation (7.5). There are two deviations from this ideal case that we will consider here. The first is inaccuracies in our estimates of $\mathbf{C}_S$ and $\mathbf{C}_{FG}$. The signal covariance, $\mathbf{C}_S$, is derived from priors and thus may be sensitive to the model chosen. We tested this in one case by using priors generated by CORA on maps generated using CRIME. We do not present the results in this paper, but we found that there was a negligible difference in performance.

On the other hand, $\mathbf{C}_{FG}$ is estimated from the data. This foreground approximation will inevitably be contaminated by some signal and noise. More precisely, for scale j, we have

$$\underline{\hat{\mathbf{f}}}^{(j)} = \underline{\mathbf{f}}^{(j)} + \underline{\mathbf{h}}'^{(j)} + \underline{\mathbf{n}}'^{(j)}. \tag{7.21}$$

In this formula, $\underline{\mathbf{h}}'^{(j)}$and $\underline{\mathbf{n}}'^{(j)}$ represent the HI and noise present in the foreground approximation, respectively. We include underlines on these variables to emphasize that these maps have been partitioned as described in Section 7.4. In this section, we will consider the effect that $\underline{\mathbf{h}}'^{(j)}$ and $\underline{\mathbf{n}}'^{(j)}$ have on the statistics of the cleaned coefficients.

Let’s begin by considering the generalized eigenvalue problem of equation (7.5). For now, let’s assume perfect knowledge of the foregrounds, noise and signal. Moreover, let us consider only one eigenmode, ϕ_i, which has a corresponding eigenvalue, λ_i. Lastly, let's consider an analysis conducted only at one needlet scale, j. We will drop any j subscripts since the analysis is all done at the same scale. This becomes the generalized eigenvalue problem

$$\mathbf{C}_{FG}\text{Œ}_i = \lambda_i \mathbf{C}_S \text{Œ}_i. \tag{7.22}$$

However, when performing NKL, we will use $\hat{\mathbf{C}}_{FG}$. We use the hat symbol to denote covariance estimates derived from the data. This gives a new eigenvalue problem

$$\hat{\mathbf{C}}_{FG}\text{Œ}_i' = \lambda_i' \mathbf{C}_S \text{Œ}_i', \tag{7.23}$$

where apostrophes represent perturbed versions of the variables seen in equation (7.22). This estimate of the foreground covariance will vary from the true version as

$$\hat{\mathbf{C}}_{FG} = \mathbf{C}_{FG} + \mathbf{\Delta}_{FG} + \mathbf{C}_{h'h'} + \mathbf{C}_{fh'} + \mathbf{C}_{n'h'} + \mathbf{C}_{n'n'} + \mathbf{C}_{fn'}. \tag{7.24}$$

In this formula, $\mathbf{\Delta}_{FG}$ represents foreground errors that would be present if $\hat{\mathbf{C}}_{FG}$ were acquired from the true foregrounds. The $\mathbf{C}_{fh'}$ term is given by

$$\mathbf{C} = \frac{1}{N_p} \sum_q (h_q' f_q^T + f_q h_q'^T), \tag{7.25}$$

In this formula, f_q is the q$^{\text{th}}$ column of $\underline{\mathbf{f}} - \overline{\underline{\mathbf{f}}}$ and h_q' is defined as $\underline{\mathbf{h}}' - \overline{\underline{\mathbf{h}}}'$. Note that the overline here has the same meaning as in equation (7.4). The variable $N_p = n_{co}^{(j)}/N$ is the number of columns in $\underline{\mathbf{f}}'$ and $\underline{\mathbf{s}}'$. This is the same as the number of needlet coefficients per chunk per frequency channel. Note that the other terms in equation (7.24), e.g. $\mathbf{C}_{n'h'}$, are defined in a similar way.

Each of these terms ought to be a source of error in our resulting estimate of the HI signal. However, the most concerning of these is $\mathbf{C}_{fh'}$, which will lead to a negative bias in our estimate of the HI signal. In principle, $\mathbf{C}_{n'h'}$ should also lead to correlations between the noise and HI after cleaning has been done. However, we did not observe this leading to any noticeable bias in the resulting power spectra. Thus, we will focus on the contribution of $\mathbf{C}_{fh'}$.

Let $\mathcal{F}_i{}' = \phi_i'^T f_p$ and $\mathcal{H}_i' = \phi_i'^T h_p$. We will find that $\mathbf{C}_{fh'}$ will lead to a non-zero correlation between $\mathcal{H}_i'$ and $\mathcal{F}_i{}'$. It turns out that this bias will be more severe for modes with larger eigenvalues λ_i. This comes into the picture through the variable

$$\alpha_{ik} = E[\overline{\mathcal{H}_i'}\mathcal{H}_k'] \approx \begin{cases} 0 & \lambda_i < 1 \\ \delta_{ik} & \lambda_i > 1, \end{cases} \tag{7.26}$$

where $E[]$ refers to an expectation value. It will be the case that

$$E[\mathcal{F}_i'\mathcal{H}_k'] \approx \delta_{ik}\frac{1}{N_p}\Sigma_{j\neq i}\frac{\lambda_j\alpha_{ii} + \alpha_{jj}\lambda_i}{\lambda_i - \lambda_j}. \tag{7.27}$$

In this equation, δ_{ik} is a Kronecker delta. A derivation of equation (7.27) can be found in the Appendix. We emphasize that this is a rough estimate, but it does provide some insight into the statistics of NKL. There are four important features about equation (7.27) that we would like to emphasize.

1. Signal-dominated modes ($\lambda_i < 1$) will be negatively biased, as they only pick up contributions from terms where $\lambda_j > \lambda_i$.
2. The $\lambda_i - \lambda_j$ in the denominator ensures that the bias will be small for modes $\lambda_i << 1$. However, the bias will be worse for modes with λ_i closer to 1.
3. This bias scales like $\frac{1}{N_p}$, meaning that it will become less severe when larger chunks are used in the analysis.
4. The bias includes no dependence on the number of frequency channels used in the analysis. This is different from ILC methods in which the bias becomes more severe as more frequency channels are included.

A test of this equation will be presented in Section 7.6.

MGNILC, a corollary

One can also use the foregrounds approximation methods of Section 7.4 to modify the GNILC technique described in Olivari et al. (2015). In GNILC, one generates an estimate for the HI signal plus noise by applying a filter to the data:

$$\hat{s} = \hat{\mathbf{S}}(\hat{\mathbf{S}}^T\hat{\mathbf{R}}^{-1}\hat{\mathbf{S}})^{-1}\hat{\mathbf{S}}^T\hat{\mathbf{R}}^{-1}x. \tag{7.28}$$

In this formula, x is the data vector (all frequency channels for a particular needlet coefficient), $\hat{s}$ is an estimate of the HI plus noise for a particular line of sight, $\hat{\mathbf{R}}$ is an estimate of the covariance of the data and $\hat{\mathbf{S}}$ is given by

$$\hat{\mathbf{S}} = \hat{\mathbf{R}}^{\frac{1}{2}}_{HI+N}\mathbf{U}_s. \tag{7.29}$$

Here, we have that $\hat{\mathbf{R}}_{HI+N}$ is a prior of signal plus noise covariance, and $\mathbf{U}_s$ is a matrix containing the subset of eigenvectors of $\hat{\mathbf{R}}$ that are dominated by the signal plus noise. In this formula, raising a matrix to the $1/2$ power refers to taking the Hermitian square root. When using an ILC technique such as GNILC, one must consider errors in $\hat{\mathbf{R}}$. In particular, $\hat{\mathbf{R}}$ will have some error due to the finite sample size used in the estimation:

$$\hat{\mathbf{R}} = \mathbf{R} + \mathbf{\Delta}. \tag{7.30}$$

In this formula, R is the true covariance of the data and $\mathbf{\Delta}$ is an error term given by

$$\mathbf{\Delta} = (\hat{\mathbf{R}}_{HI+N} - \mathbf{R}_{HI+N}) + (\hat{\mathbf{R}}_{FG} - \mathbf{R}_{FG}) + \tilde{\mathbf{C}}. \tag{7.31}$$

In this formula, the first two terms correspond to errors in the foregrounds and HI + n respectively. The third term corresponds to a spurious correlation that appears between the foregrounds and signal plus noise due to the finite sample size used. This is given by the equation

$$\tilde{\mathbf{C}} = \frac{1}{N_p} \sum_q (s_q f_q^T + f_q s_q^T). \tag{7.32}$$

In this formula, N_p is the number of needlet coefficients in the neighborhood being considered and q is a variable that indexes lines of sight in the needlet coefficient neighborhood. We also introduce the variable $s_q = h_q + n_q$.

Similarly to what was seen in NKL, $\hat{\mathbf{C}}$ causes residual foregrounds to have a negative correlation with the HI plus noise. This leads to the negative “ILC bias" described in Delabrouille, J. et al. (2009). This bias scales like $\frac{n_{ch}}{N_p}$. Such a bias is troublesome at larger scales, for which N_p is restricted by the number of needlet coefficients available.

To mitigate this problem, we propose Modified GNILC (MGNILC). In this approach, we use a modified covariance estimate

$$\hat{\mathbf{R}}_m = \hat{\mathbf{R}}_{FG} + \hat{\mathbf{R}}_{HI+N}. \tag{7.33}$$

In this formula, $\hat{\mathbf{R}}_{FG}$ is an estimate of the foreground covariance in the needlet neighborhood of interest using the approach described in Section 7.4. In this case, our error term will look like

$$\hat{\mathbf{C}}_m = \frac{1}{N_p} (s'_q f_q^T + f_q s'^T_q). \tag{7.34}$$

In this formula, s'_q is the HI left over after performing the foreground approximation. Whether using PCA or DAYENU for the foreground approximation, we should find that s'_q is essentially a version of s_q that has been low-pass filtered in delay space. This reduces the ILC bias in two ways. For one, the fact that the HI plus noise has been low-pass filtered prevents the negative bias from affecting larger delays. Moreover, it also prevents any spurious correlations between high delay components of s with the low delay components of f, providing some additional mitigation of the

ILC bias. However, the ILC bias problem will not be fixed completely as there will of course still be correlations between f_q and the low delay content of s_q.

7.5 Simulated Maps For Testing

This section provides a description of the simulated maps used for testing the various foreground removal techniques described earlier in this paper.

Cosmological Signal

We generated simulated HI maps using the CORA package [1]. The CORA software assumes that the 21 cm signal is Gaussian and isotropic. For computational convenience, the covariance is estimated using the flat sky approximation. One then simulates the 21 cm fluctuations using equation C5 from Shaw et al. (2015). The mean HI temperature is then supplied using equation C4 from that same paper. The CORA software uses cosmological parameters from Planck 2018 (Planck Collaboration et al. (2020a)). Moreover, we used the default CORA setting, which is to estimate Ω_{HI} using the model given in Crighton et al. (2015). The CORA software package also incorporates redshift space distortions, assuming a constant HI bias of $b(z) = 1$ by default.

Foregrounds

In this paper, we consider only foregrounds from Galactic synchrotron radiation. For the synchrotron radiation, both unpolarized and polarized contributions are included. For the polarized foregrounds, 1 percent leakage of both the Q and U components was assumed. This value was chosen to be consistent with a typical level of polarization leakage for HI intensity mapping telescopes. These foregrounds are simulated using the CRIME software package (Alonso et al. (2014)). CRIME models unpolarized Galactic synchrotron emission via

$$T(\hat{n}, \nu) = T_{\text{Haslam}}(\hat{n})(\frac{\nu}{\nu_H})^{\beta(\hat{n})} + \delta T. \tag{7.35}$$

In this formula, T_{Haslam} is the Haslam map temperature, β is a spectral index generated from the Planck sky model (Delabrouille, J. et al. (2013)) and ν_H is 408 MHz, the frequency of the Haslam map. The δT term accounts for angular scales $l \gtrsim 200$, which are smaller than the Haslam map's resolution. This term is generated assuming that the foregrounds are Gaussian and isotropic at these scales. The power spectrum assumed can be found in Alonso et al. (2014).

---

[1] `https://github.com/radiocosmology/cora`

The polarized synchrotron emission is modelled using the measured Faraday depth through the entire Milky Way presented in Oppermann et al. (2012). CRIME then assumes, among other things, that the number of emitting regions follows a Gaussian distribution as a function of Faraday depth.

CORA also can be used to model the polarized synchrotron radiation. The model used by CORA is essentially the same as that used by CRIME. The main theoretical difference is that CORA assumes that emitting regions at the same Faraday depth for a particular line of sight are independent. This leads to a slightly different dependence on the emission at Faraday depth ψ for a particular line of sight. The Faraday depth coherence lengths are also slightly different, being $0.5\,\mathrm{rad/m^2}$ for CRIME and $1\,\mathrm{rad/m^2}$ for CORA. The value for CRIME was chosen to provide results consistent with the Hammurabi simulation package (Jaffe et al. (2012)), while CORA's value was chosen to be consistent with observations at 1.4 GHz (Wolleben et al. (2006)). In practice, we found that the foregrounds provided by CRIME appeared to be more chromatic than those from CORA. Thus, to be conservative, we used CRIME to simulate the foregrounds.

Instrumental Effects and Noise

In this paper, tests of various foreground removal techniques are performed assuming an instrument similar to the MeerKAT array operating in single dish mode. The parameters of this telescope are described in Table 7.1. One may notice that the observing time of 40000 hrs is quite large. This value is not particularly realistic, but was chosen to ensure that the noise would not overpower the HI signal. In future telescopes such as PUMA, adequate noise levels can be achieved with larger numbers of antennas and less integration time. The value $f_{sky} = 1$ was chosen since we are imagining an instrument that observes the whole sky. Any masking is then applied after observations have taken place.

This hypothetical instrument was chosen for realism and simplicity. Since it consists of an array of dishes operating in single dish mode, we can model the beams by convolving with a Gaussian. This is much easier than trying to derive maps from simulated visibilities. Moreover, the relatively low angular resolution of the instrument ($\approx 1^\circ$) ensures that large values of ℓ will not be required. This relaxes computational requirements when working with needlets. We approximate beam effects in a way that is identical to Carucci et al. (2020). In particular, we convolve the simulated maps with a frequency-dependent Gaussian beam. We then reconvolve all maps to give them the same angular resolution.

We added noise to the simulated maps in the same way as Carucci. In particular, the noise per pixel follows a Gaussian distribution with a standard deviation given by

$$\sigma_N(\nu) = T_{\mathrm{sys}}(\nu)\sqrt{\frac{4\pi f_{\mathrm{sky}}}{\Delta\nu t_{\mathrm{obs}} N_{\mathrm{dishes}} \Omega_{\mathrm{beam}}}}. \tag{7.36}$$

Parameter	Value
D	13.5 m
T_{inst}	20 K
f_{sky}	1
t_{obs}	40000 hrs
N_{dishes}	64
n_{ch}	256
$[\nu_{min}, \nu_{max}]$	[980, 1080 MHz] and [400, 500 MHz].
$\Delta\nu$	0.390625 MHz

Table 7.1: Parameters describing the hypothetical telescope used for the VKL analysis in this chapter.

In this formula, T_{sys} is the system temperature, f_{sky} is the fraction of the sky observed in the survey, $\Delta\nu$ is the frequency resolution, t_{obs} is observing time and Ω_{beam} is the beam solid angle. Similarly, we estimate the system temperature in the same way as Carucci, assuming

$$T_{sys}(\nu) = T_{instr}[\mathrm{K}] + 66\left(\frac{\nu}{300[\mathrm{MHz}]}\right)^{2.55}. \tag{7.37}$$

Once noise and Gaussian beam effects are accounted for, the maps used in testing can be described schematically via the equation

$$\mathbf{X} = (B_{low} - B) * (B * (\mathbf{f} + \mathbf{h}) + \mathbf{n}). \tag{7.38}$$

In this schematic, $B*$ represents convolution with the beam and B_{low} is the beam at lowest frequency channel in the band.

7.6 Tests of Methods

In this section, we test the performance of various foreground removal methods and compare the results. In particular, we test GMCA, PCA, ICA, GNILC, and NKL. It should be noted that the severity of polarized foregrounds is expected to vary as a function of redshift. In particular, the chromaticity of polarized foregrounds will become more severe at lower frequencies due to Faraday rotation. Therefore, we conduct tests in two different redshift regimes corresponding to low frequency and high frequency bands of MeerKAT.

Additional Map Preparation Steps

After performing the steps described in Section 7.5, we took some additional steps to aid the cleaning process. These were done both to exclude highly contaminated pixels and to lower computation

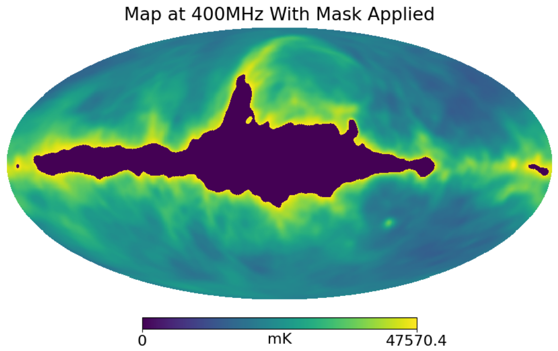

Figure 7.2: An illustration of the mask used for tests of the foreground removal methods in the high redshift case. The mask has been applied to this map, which includes simulated foregrounds, signal and noise.

times.

As expected, polarized foreground chromaticity is quite severe in the Galactic plane. We applied a mask to the data to exclude the brightest 15 percent of pixels. Such masks were created for both the low-z and high-z test cases. The mask created for the high-z test case is shown in Fig. 7.2. The chromaticity problem is illustrated in Fig. 7.3. The top plot in the figure shows delay spectra for the brightest line of sight in the high-z test case. Note here that the foregrounds dominate for almost all delays. The bottom plot shows delay spectra at the brightest unmasked line of sight. In this case, we find that the foregrounds occupy a relatively small region of delay space, allowing for more effective cleaning.

It was noted that the Gaussian half-power beam-widths in both the low-z and high-z test cases are greater than 1°. As such, we truncated maps in both test cases at $\ell = 255$ for computational simplicity. Moreover, we performed needlet transformations using $B = 2$ with scales running from $j = 4$ up to $j = 8$. We provide details of how the data is represented in the needlet domain in Table 7.2.

It is worth discussing here the impacts that B and the range of j can have on one's analysis. Increasing the value of B will decrease the spread of the needlet coefficients in real space while increasing their spread in harmonic space. So, choosing a value of B greater than 2 would in principle allow for finer chunking of the maps but would also lead to coarser partitioning in ℓ-space.

Next, let's consider the range of j chosen. Since the maps were truncated at $\ell = 255$, scales with $j > 8$ would contain no information and thus were not included. The minimum scale value $j_{min} = 4$ was chosen to ensure that the scaling function would have enough statistics to be cleanable by NKL.

The value of B and range of j chosen for this paper do work. However, other values were not tested, so it is unknown whether these are the optimal choices one could make.

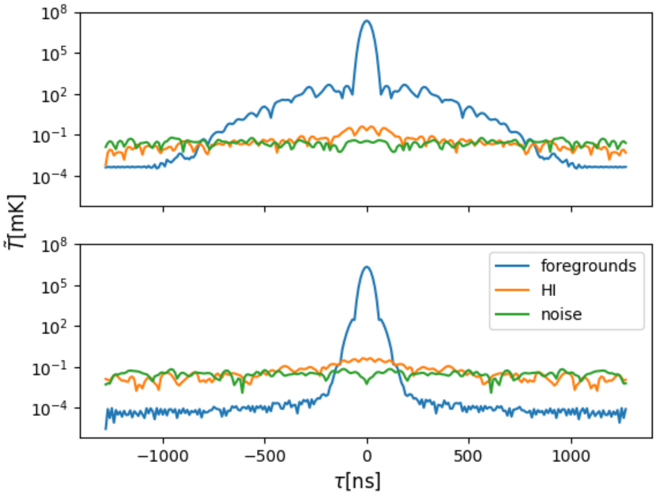

Figure 7.3: The top plot shows delay spectra at the brightest pixel in the unmasked map in the higher redshift case. The bottom plot shows delay spectra at the brightest pixel remaining after the mask has been applied.

Scale	**l range**	**nside**
Scaling Function	$[0, 16]$	16
$j = 4$	$[8, 32]$	32
$j = 5$	$[16, 64]$	64
$j = 6$	$[32, 128]$	128
$j = 7$	$[64, 256]$	256
$j = 8$	$[128, 512]$	512

Table 7.2: A summary of the needlet coefficients in these tests. We use the same scales, l ranges and nside values for both the low-z and higher-z tests.

Implementation of Cleaning Techniques

When cleaning with blind methods like PCA, ICA and GMCA, we use priors on the HI signal and noise to select the appropriate number of components to be removed. In particular, we generated additional sets of HI maps using CORA and noise maps using equation (7.36). We then applied the beam convolution described in Section 7.5 and produced covariance estimates from these processed maps using equation (7.12). In this paper, we chose to use $N_{maps} = 10$ mock maps for estimating the covariances. We then used our models of the HI and noise covariances to estimate the appropriate number of modes to remove using the AIC prescription described in Olivari (Olivari et al. (2015)).

For the cases of GNILC, MGNILC and NKL, we used the same sets of mock HI and noise maps as above to estimate the needlet coefficient covariances.

For GNILC, we selected the appropriate number of coefficients using the Olivari AIC prescription.

We also chose to use windows with a size of at least 10^5 coefficients. This seemed to be the window size required to minimize the ILC bias described in Delabrouille, J. et al. (2009). There were of course exceptions to this rule for scales up to $j = 6$, since needlet maps at lower j did not have enough coefficients. These windows were generated simply by choosing the closest N coefficients to the coefficient of interest.

Power Spectrum Estimation

In order to estimate the power spectrum, we used the method described in Liu et al. (2016). In this approach, one computes Bessel-Spherical harmonic modes of the sky

$$T_{\ell m}(k) = \sqrt{\frac{2}{\pi}} \int d\Omega dr r^2 j_\ell(kr) Y^*_{\ell m}(\hat{n}) \phi(\mathbf{r}) T(\mathbf{r}), \tag{7.39}$$

where $\phi(\mathbf{r})$ is a window function representing the survey volume. One then computes a “spherical harmonic power spectrum"

$$S_\ell(k) = 2\pi^2 \left[\int d^3r \phi(\mathbf{r})^2 j_\ell(kr)^2\right]^{-1} \frac{\Sigma_m |T_{\ell m}(k)|^2}{2\ell + 1}. \tag{7.40}$$

This spherical harmonic power spectrum is a spherical analogue to the commonly seen cylindrical power spectrum. It is also a useful tool in that it allows for different ℓ-modes to be checked individually. Then, in the case of a translation-invariant sky, one can form the power spectrum estimator

$$\hat{P}(k) = \Sigma_\ell w_\ell S_\ell(k). \tag{7.41}$$

In this formula, the w_ℓ are weights which depend on the survey volume used. These weights account for the fact that sensitivity to a certain k-mode may vary with angular scale ℓ. In this case of a translation-invariant sky, $\langle \hat{P}(k) \rangle = P(k)$. It should be mentioned that the maps presented in this paper do not strictly satisfy translation invariance. Foregrounds vary significantly with line of sight and the HI is subject to redshift-space distortions and cosmic evolution. Even the noise breaks translation invariance, as its amplitude varies with frequency. Despite this, we still present $\hat{P}(k)$ as a “power spectrum", since it is provides us with a weighted average of fluctuations at length scale k, even when translation invariance is broken.

Test at Higher Redshifts

For this test, we used the hypothetical instrument described in Table 7.1 and assumed it to have 256 evenly spaced frequency channels running from 400 MHz to 500 MHz.

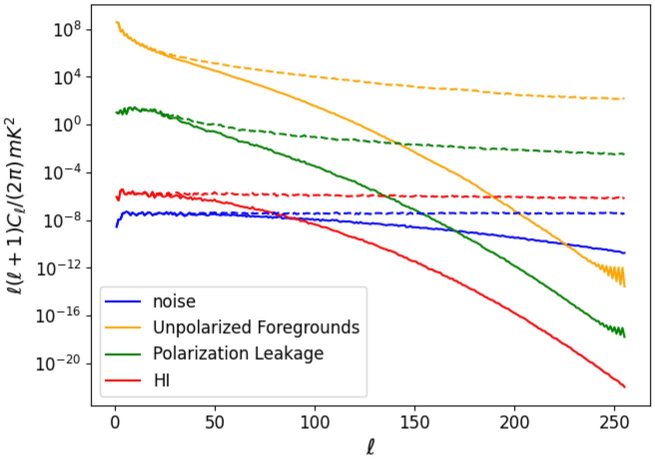

Figure 7.4: Angular power spectra estimated from unmasked maps using `healpy` at 450 MHz. Solid lines correspond to maps which have undergone beam convolution. Dashed lines correspond to maps that have not undergone any convolution steps.

In Fig. 7.4, we provide angular power spectra for the 450 MHz frequency channel used in this test. The maps used to produce these power spectra have not been masked. We chose to use unmasked maps here in order to avoid edge effects at the mask boundaries. In this case, the beam convolution causes significant loss of signal above $\ell \approx 50$. Due to this, any power spectra produced by our instrument will be missing significant amounts of small length scale information. For this reason, we will evaluate the effectiveness of the cleaning techniques based on how well they recover the power spectra of convolved maps. It should also be noted that the noise suffers less loss at $\ell > 50$ than the other components. This is due to the fact that the foreground and HI have been convolved with a beam twice while the noise has only been convolved once. This causes the noise to be the dominant component of the maps at those smaller angular scales.

For the case of NKL, the chunk size and desired SNR of the KL cleaning vary with needlet scale. We summarize this in Table 7.3. The values shown in this table were chosen through trial and error. Ideally, one would like to clean with as many chunks as possible, since this would provide the best statistics. However, we are limited by two factors. For one, making the chunks too small leads to an inaccurate $\hat{C}_{FG}$ and incomplete cleaning of the foregrounds. We also found that using small chunks also leads to inaccuracies in C_{HI}. However, this is not a fundamental issue and would be fixed by simply using more mock maps in our estimates. Trying to keep our models accurate is what motivated our choices for the number of chunks for scales up through $j = 6$. For scales $j = 7$ and $j = 8$, we used 96 chunks as this was the largest value for which our computer was able to compute eigenvectors.

In practice, an analyst using NKL would probably decide on a chunking scheme by trial and error using simulations. However, it may be worth providing some intuition on how these parameters may depend on the survey in question. We should emphasize here that this intuition is really guess work. More precise characterization will need to be determined by trying out NKL in different scenarios.

Suppose one decreased the resolution of one's instrument. This would lead to the angular power spectra like those shown in Fig. 7.4 dropping off more quickly as a function of ℓ. For adequately large (i.e. small j) needlet scales, the coefficients would not change much and our approach would stay more or less the same. For smaller needlet scales, noise contamination would become more severe and one may find that larger chunks are required to control errors in $\hat{C}_{FG}$ due to noise leakage.

Now, suppose that one decreased the survey area. Naturally, one would mask out adequately small needlet coefficients and decrease the number of chunks accordingly. It is possible that one could decrease the size of chunks used, as the small and hard to model correlations of distant chunks is now gone. However, this could still pose problems as we expect the anti-correlation between residual foregrounds and signal to scale like $1/N_p$.

In the case of an especially small survey area, one may also find that there are not enough non-zero coefficients to adequately estimate $\hat{C}_{FG}$ at lower values of j and in the scaling function. In such a case, one would likely have to increase either the minimum value of j or the value B, increasing the number of coefficients per map but providing a coarser partitioning of ℓ-space.

We found for this high-z case that approximating the foregrounds with DAYENU worked better than with PCA. We believe DAYENU's superior performance in this scenario is likely due to the fact that it treats each pixel in the needlet map individually. On the other hand, PCA gives each pixel the same treatment. This fine-grained approach is useful at higher redshifts where foreground chromaticity varies quite significantly with line of sight.

For the DAYENU foreground approximation, we chose the SVD cutoff to be a factor of 10^{-14} below the largest singular value of $\mathbf{C}^{\urcorner}$. This value ensured that all foregrounds were captured in the approximation.

The MGNILC cleaning was performed using needlet neighborhoods of size 1500 coefficients. MGNILC was also performed using the same foreground approximation as for NKL.

For all blind methods, 78 components were used to model the foregrounds. This value was selected by AIC.

In Fig. 7.5, we present the 3D power spectra recovered by the methods described above using equation (7.41). One may note that these results look more pessimistic than what is seen in other papers on this topic. This is because we are including polarized foregrounds at a higher redshift, which is uncommon. Moreover, most of the sky is kept, including areas with relatively bright foregrounds. One will also notice that all methods tested behaved similarly at small spatial scales ($k \gtrsim 0.6\,\mathrm{hMpc}^{-1}$). One exception to this is the NKL curve, which is lower than the others at such scales. This is due to the fact that NKL is the only method tested here that makes a distinction

Scale	**Number of Chunks**	SNR
Scaling Function	3	1
$j = 4$	12	4
$j = 5$	32	4
$j = 6$	64	4
$j = 7$	96	4
$j = 8$	96	4

Table 7.3: A summary of the NKL cleaning parameters used in the high-z test.

between HI and noise. These small scales carry noise comparable to or greater than the HI signal. As such, NKL looks different than the other methods in this regime. It should be noted that this feature goes away when we set $\mathbf{C}_S = \mathbf{C}_{HI} + \mathbf{C}_N$. This feature is also not a problem, as the debiased data will show that the HI is still well preserved by NKL at these small scales. At larger spatial scales, there is a clear bifurcation between the blind and non-blind methods. The blind methods clean extremely aggressively at such scales, leaving residuals up to 5 orders of magnitude below the signal plus noise power spectrum. This makes sense, as these large scales will correspond to the foreground-dominated delays such as those seen in Fig. 7.3. There is also a notable trough for certain techniques below $k \approx 0.20\,\mathrm{hMpc}^{-1}$. These scales should correspond roughly to the regime in delay space shown in Fig. 7.3 where the foregrounds and HI plus noise are approaching the same magnitude. On the other hand, the non-blind methods seem to be able to preserve some information from within this foreground-dominated region of delay space, having residuals several orders of magnitude closer to that of the true signal plus noise power spectrum. Moreover, NKL seems to provide roughly a factor of 10 to 50 improvement over GNILC at these larger scales. However, it should be noted that all methods incur significant signal loss at the large spatial scales. Even NKL is a factor of a few below the desired level. This is perhaps something that could be compensated for through the use of a transfer function, but we leave such an analysis to future work.

In Fig. 7.6, we present power spectrum curves which have been debiased to remove the noise. The power spectra are debiased according to the equation

$$\hat{P}_{\mathrm{debiased}}(k) = \hat{P}(k) - \hat{P}_N(k). \tag{7.42}$$

In this formula, $\hat{P}(k)$ is the power spectrum estimate obtained from the cleaned maps and $\hat{P}_N(k)$ is an estimate of the noise power spectrum obtained from our model of the noise. Note that NKL was debiased slightly differently, with the noise model used being one taking into account the effects of the KL transform on the noise. There are a few features in this plot worth noting. First, it should be noted that the debiased power spectra for GNILC, ICA, GMCA and PCA become negative at lower k values. This is due to the loss in both the signal and noise incurred at large scales by these

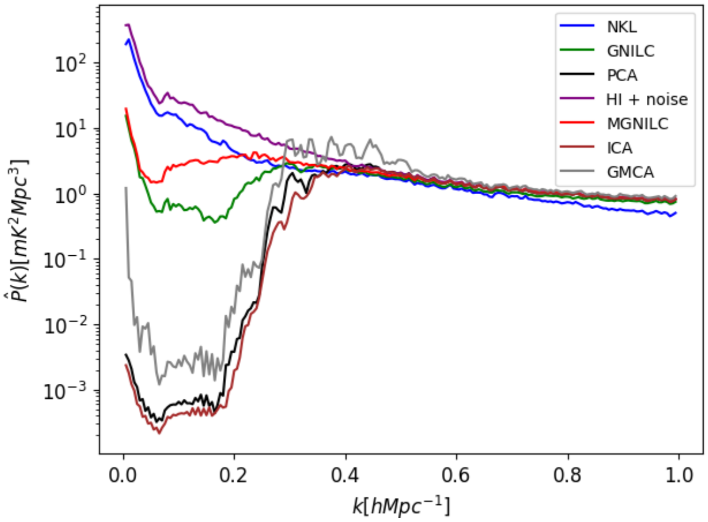

Figure 7.5: Power spectra estimated from the cleaned high redshift maps. The “HI + noise" curve shows the power spectrum of the masked, beam-convolved maps. NKL was tested on a second realization of signal plus noise, producing similar results.

methods, combined with the fact that the noise power spectrum estimate $\hat{P}_N$ does not account for this loss. It should be noted that these power spectrum values would likely not become negative had we accounted for this loss. However, even with a more sophisticated debiasing, one would still find these methods being outperformed by MGNILC and NKL. The MGNILC power spectrum estimate never becomes negative. This is because it suffers from less signal loss than the previously mentioned methods. NKL was debiased somewhat differently than the other methods, with the noise power spectrum being estimated from maps cleaned with NKL. It was found that NKL increased the power of the noise at large length scales $k \lesssim 0.35\,\mathrm{hMpc}^{-1}$, and decreased the noise power at scales $k \gtrsim 0.35\,\mathrm{hMpc}^{-1}$. The use of a modified noise power spectrum resolved these issues in the case of NKL. Moreover, GNILC seems to underestimate the HI power spectrum at higher values of k. This is due to the ILC bias described in Delabrouille, J. et al. (2009).

In Fig. 7.7, we present graphs illustrating the ratio of $|S_{l,HI}|$ to $|S_{l,cd}|$ as a function of k. The variable $S_{l,HI}$ represents the spherical harmonic power spectrum of the HI plus noise. On the other hand, $S_{l,cd}$ represents the spherical harmonic power spectrum obtained from cleaned maps. Note that we also debias $S_{l,cd}$. In this figure, one will notice that there are common features between techniques. First, one will notice larger values of $\log_{10}(|\frac{S_{l,HI}}{S_{l,cd}}|)$ in the top left corner of each plot. This region corresponds to large angular scales and low delays. Since such modes carry the most severe foreground contamination, one will find them being cleaned most aggressively. This effect is most severe for PCA and least severe for NKL. Next, one will notice the plots becoming darker at

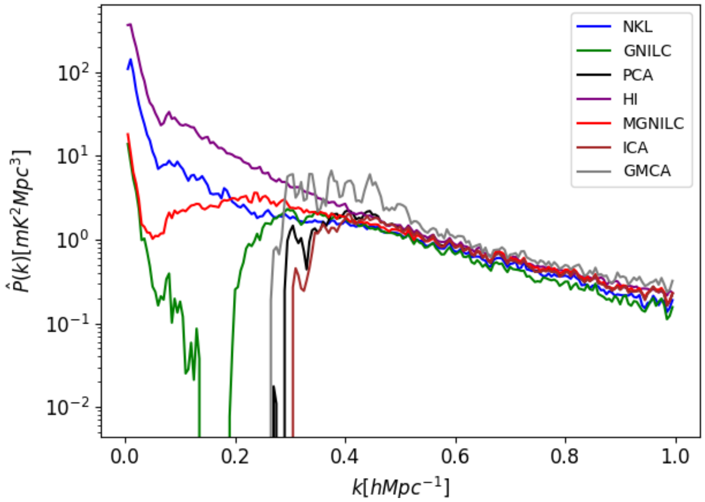

Figure 7.6: Power spectrum estimates which have been debiased. The “HI" curve provides the power spectrum of the HI maps after beam convolution and masking.

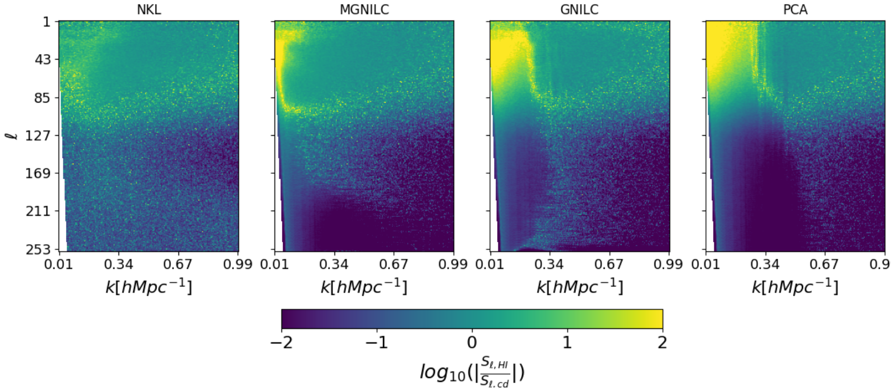

Figure 7.7: The log-scaled absolute value ratio of the true high-z spherical harmonic power spectrum ($S_{l,HI}$) over the spectrum recovered by cleaning and debiasing ($S_{l,cd}$).

higher ℓ. This is due to the fact that the beam convolution applied causes the noise to overpower the HI at small angular scales. The strength of the noise relative to the HI at these scales leads to inaccuracy in the debiasing process. Overall, it appears that NKL has the lowest errors at large spatial scales, as expected from Fig. 7.5.

It is also important to consider residual foregrounds and any bias that these residuals may add to the estimated the HI signal. We will investigate these foreground residuals for GNILC, MGNILC, NKL and PCA. Each of these methods cleans the data by first producing a matrix (or matrices) used for cleaning, followed by application of the matrix (matrices) to the data. For instance, when using PCA, one generates a cleaning matrix $\mathbf{R}$ from a covariance estimate of the data, then cleans the data as $\mathbf{d}_{cleaned} = \mathbf{R}(\mathbf{d} - \overline{\mathbf{d}})$. In order to produce foreground residual estimates, we apply these cleaning matrices to data containing only foregrounds. So, for instance, the foreground residuals for PCA will be

$$\mathbf{f}_{residual} = \mathbf{R}(\mathbf{f} - \overline{\mathbf{f}}). \tag{7.43}$$

A similar approach can be taken to obtain the foreground residuals for GNILC, MGNILC and NKL. We also produce an estimate of the HI plus noise present in the cleaned maps in a similar way. In the case of PCA, we would have

$$(\mathbf{h} + \mathbf{n})_{cleaned} = \mathbf{R}((\mathbf{h} - \overline{\mathbf{h}}) + (\mathbf{n} - \overline{\mathbf{n}})). \tag{7.44}$$

Fig. 7.8 illustrates the effect that these foreground residuals have on the resulting power spectra. First, notice that the dotted lines, representing the power spectra of the signal plus noise present in the clean maps, tend to be higher than the power spectra of the cleaned maps for the NKL, GNILC, and MGNILC cases. This is indicative of the presence of a negative bias in all three techniques. Moreover, this feature implies that the negative bias present in these techniques decreases the power spectrum estimates, worsening agreement with the true HI power spectrum. The negative bias seen in GNILC and MGNILC is predicted by the ILC bias described in Delabrouille, J. et al. (2009). The negative bias in NKL was of course predicted in the analysis performed in Section 7.4.

We also explicitly tested equation (7.27) on the results of NKL at needlet scale $j = 4$. In this case, we have $N_p = 1024$. The results of this are shown in Fig. 7.9. For this figure, we estimate $E[\mathcal{F}_i'\mathcal{H}_i']$ by averaging the KL coefficients across all columns of $\underline{\mathbf{h}}$ and $\underline{\mathbf{f}}$. So, we have

$$\hat{E}[\mathcal{F}_i'\mathcal{H}_i'] = \frac{1}{N_p}(\phi_i'^T \underline{\mathbf{h}})(\underline{\mathbf{f}}^T \phi_i'). \tag{7.45}$$

Since we are less interested in foreground-dominated modes, we only present results up to $\lambda = 2$ for clarity. In this plot, one can see that equation (7.27) underestimates the severity of the bias, but is able to roughly capture what is happening. In particular, it predicts that the bias is negative and that

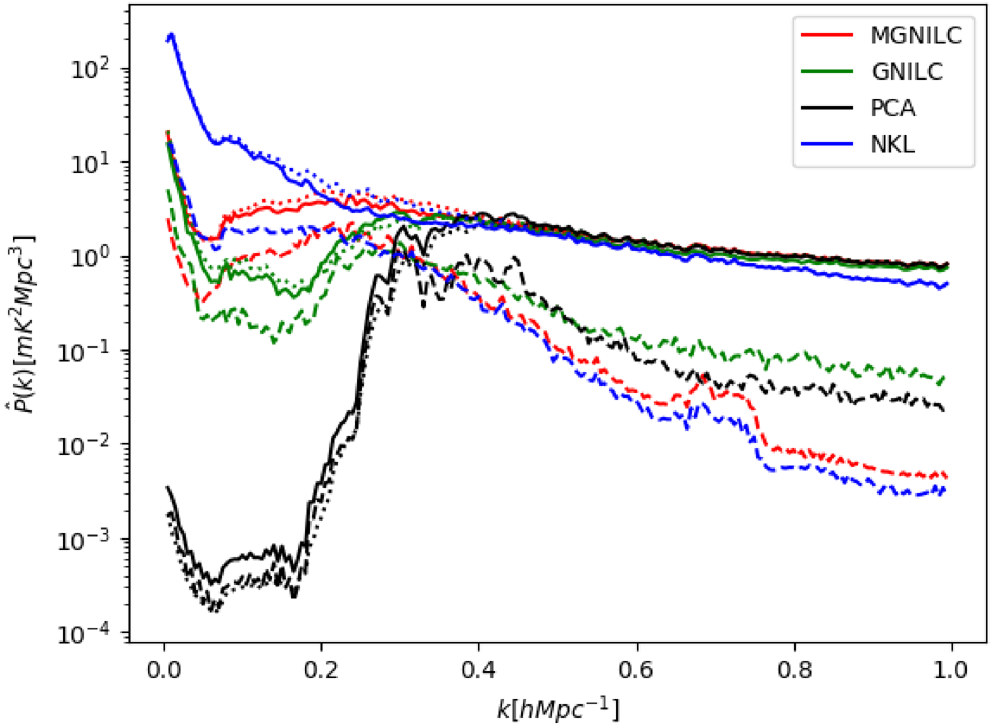

Figure 7.8: Power spectra illustrating the effects of foreground residuals in the high redshift case. Solid lines correspond to the power spectra of the cleaned maps. Dashed lines correspond to spectra obtained from foreground residuals. Dotted lines correspond to the power spectrum of the HI plus noise present in the cleaned maps.

it becomes more negative as it one approaches $\lambda' = 1$, the eigenvalue at which the foregrounds and signal plus noise are expected to have equal variance.

Another aspect of equation (7.27) to consider is the claim that $E[\mathcal{H}_i'\mathcal{F}_k'] = 0$ for $i \neq k$. We present a test of this in Fig. 7.10. In this figure, the blue curve shows estimated values of $E[\mathcal{H}_i'\mathcal{F}_k']$ for $\lambda_i' = 0.15$. We see that this correlation does not strictly come out to be 0 as predicted by equation (7.27). Rather, it seems to oscillate about 0, with 0 correlation being outside the error bars of many eigenvalues. In this case, the average of these off-diagonal terms (up to $\lambda' = 0.25$) was slightly negative (order 10^{-3} and within a standard deviation of 0. This is much smaller than the diagonal terms $E[\mathcal{H}_i'\mathcal{H}_i']$, which are of order 1. As such, these off-diagonal terms will increase errors in the power spectrum estimates, but will not contribute much of a net bias.

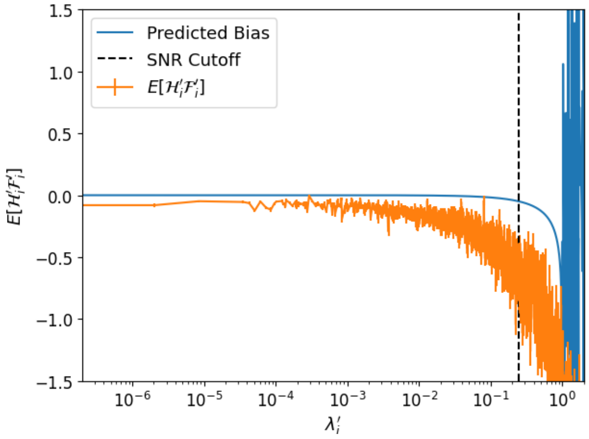

Figure 7.9: A test of equation (7.27), which modeled the negative bias incurred by NKL, performed on the $j = 4$ needlet scale.

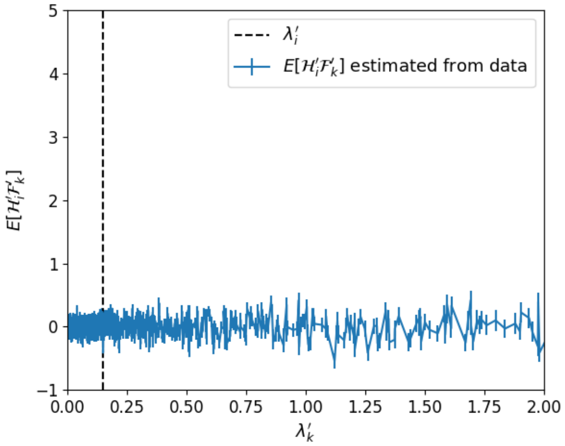

Figure 7.10: Estimated values of $E[\mathcal{H}_i'\mathcal{F}_k']$ for $\lambda_i' = 0.15$.

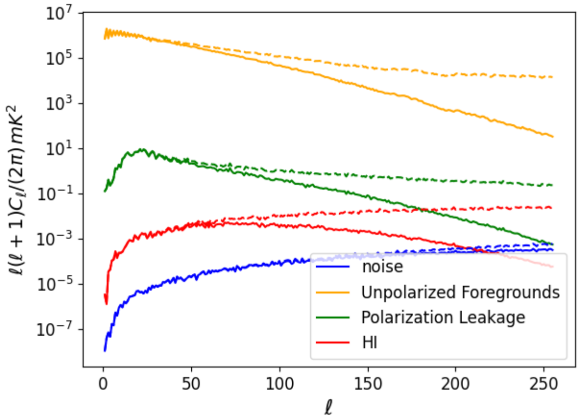

Figure 7.11: Angular power spectra estimated from unmasked maps using healpy at 1030 MHz. Solid lines correspond to maps that have undergone beam convolution. Dashed lines correspond to maps that have undergone no beam convolution.

Test at Lower Redshifts

In this subsection, we test these same foreground removal methods at lower redshifts than in Section 7.6. In particular, we used 256 evenly spaced frequency channels from 980 MHz up to 1080 MHz. This test was motivated by the fact that the characteristics of the foregrounds change with redshift. In particular, models predict that polarized foregrounds ought to be less severe in this regime. Note that we have again chosen to mask the brightest 15 percent of pixels.

In Fig. 7.11, we present the angular power spectra estimated at the 1030 MHz frequency channel. In this case, for the convolved maps, the HI has a much stronger contribution to the maps than the noise for scales up to $\ell \approx 200$. Although not as severe as in the higher redshift case, the angular resolution of the antenna beams have once again caused significant signal loss at the smaller angular scales.

For the blind methods, we tried to use AIC to estimate the optimal number of modes to remove. In particular, AIC predicted 4 modes as the best option. However, preliminary checks using the radial power spectrum estimation code from Carucci [2] showed a large spike in the radial power spectrum at low k_ν. This spike disappeared when the number of modes was increased to 5, resulting in better agreement between the radial power spectra of the HI plus noise and the cleaned maps. As a result, we chose to remove 5 modes for all blind methods in this test case.

As for GNILC, NKL and MGNILC, we once again used the needlet domain scheme described in Table 7.2. In this case, we again only compute needlets for modes up to $\ell = 255$, as the beams used in this regime have widths $\theta \approx 1^\circ$. GNILC was performed in the same way as in Section 7.6,

---

[2] `https://github.com/isab3lla/gmca4im`

Scale	Number of Chunks	SNR
Scaling Function	6	1
$j = 4$	12	1
$j = 5$	32	1
$j = 6$	64	1
$j = 7$	96	1
$j = 8$	96	1

Table 7.4: A summary of the number of chunks and SNR values chosen for each scale in the low-z test.

where we sought to use windows of size 10^5 coefficients. For MGNILC and GNILC, we used neighborhoods of size 3000 coefficients as AIC struggled to select the correct number of modes when using neighborhoods of size 1500. For NKL, we describe the cleaning parameters in Table 7.4. The number of chunks used has remained more or less the same as in Section 7.6. We have however, increased the number of chunks for the scaling function and the $j = 4$ scale. In addition, note that the same SNR is used for all needlet scales. For this particular test, we found that approximating the foregrounds using the PCA approach was more effective than using DAYENU. That is what we will present in this section. It seems likely that PCA provides an advantage over DAYENU in this scenario for two reasons. For one, the foregrounds in this scenario were quite narrow in delay space. The delay spectra of these foregrounds were particularly narrow compared to those of the higher redshift test case. The fact that the foregrounds vary so rapidly at delays close to 0 lead to our DAYENU based scheme struggling to accurately estimate the width of the foregrounds. In particular, this approach was usually overestimating the width of the foregrounds in delay space. On the other hand, the fine-grained nature of the DAYENU-based approach was less important here than in the higher redshift case. This is due to the fact that foreground chromaticity did not vary as much as function of the line of sight in the lower redshift test case. It should be noted however, that AIC struggled with selecting the correct number of modes during the foreground approximation step. We fixed this simply by choosing to remove $m + 2$ modes, where m was the number of modes selected by AIC. It should also be noted that we chose $\mathbf{C}_S = \mathbf{C}_{HI} + \mathbf{C}_N$ for this particular test case. This was done since our computer program interpreted $\mathbf{C}_{HI}$ as being singular, preventing the KL transform from being performed. Tacking on $\mathbf{C}_N$ was a convenient workaround since the noise is very small in this test and the computations for the $\mathbf{C}_S = \mathbf{C}_N + \mathbf{C}_{HI}$ test had already been performed.

In Fig. 7.12, we present power spectra estimated from cleaned maps. It should be noted that the noise is quite low in this test case. As such, the debiasing process of equation (7.42) makes an imperceptible difference. Because of this, we only present the debiased results. Note that all methods in this figure were debiased using the same noise power spectrum estimate $\hat{P}_N(k)$. In this figure, one will notice first that GMCA and PCA produce essentially identical results, and that they

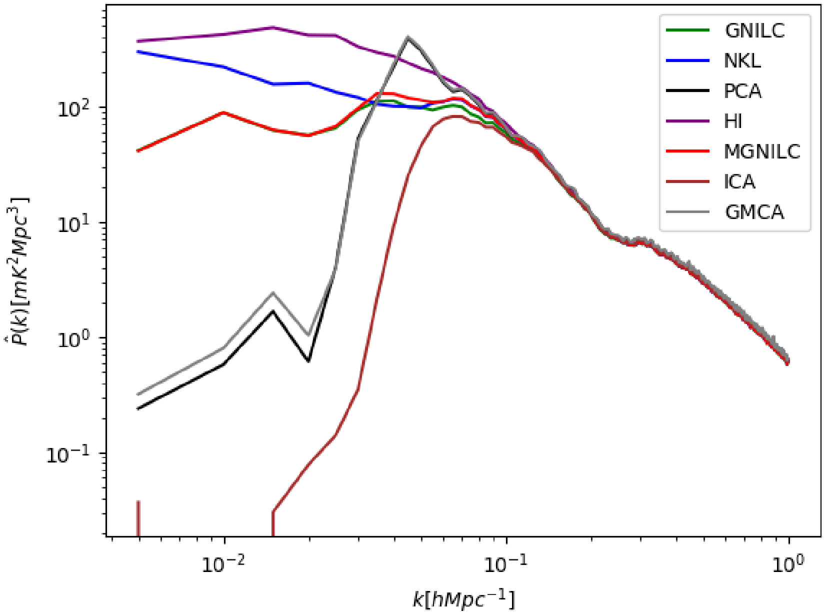

Figure 7.12: 3-dimensional power spectra computed for various techniques in the lower redshift case. The “HI" curve was estimated from maps that had undergone beam convolution and masking. We also tested NKL on a second realization of the signal plus noise, which produced similar results.

seem to provide the best match to the HI plus noise power spectrum. However, this is somewhat misleading, as the foreground residuals at low k roughly match the HI plus noise power spectrum. This is made clear in Fig. 7.13, where one can see that the power spectrum at these scales is mostly due to foreground residuals. The GNILC and MGNILC power spectra also receive a boost from their foreground residuals, but the effect is not as dramatic as in the case of GMCA and PCA. It also appears that the ILC bias in this case is not as severe as in the higher redshift case. This is likely due to the fact that the larger neighborhoods of needlet space were used. As in the higher redshift test case, we find that the presence of foreground residuals decreases the NKL power spectrum at low k. However, the effect on NKL is smaller in this case compared to the one at higher redshift.

Interestingly, ICA has performed differently from GMCA and PCA in this context. It appears that ICA has smaller foreground residuals than either GMCA or PCA in this case. Next, one will notice that ICA, MGNILC, GNILC and NKL give very similar results down to $k \approx 0.15\,\mathrm{hMpc}^{-1}$. For $k < 0.15\,\mathrm{hMpc}^{-1}$, the non-blind methods once again provide significantly improved results. For scales $0.03\,\mathrm{hMpc}^{-1} < k < 0.15\,\mathrm{hMpc}^{-1}$, we find that NKL, MGNILC and GNILC provide similar results. At the very lowest scale ($k < 0.03\,\mathrm{hMpc}^{-1}$), we find NKL provides a factor of a few improvement over GNILC and MGNILC.

Consider again Fig. 7.13, which shows the foreground residuals present in this test case. As

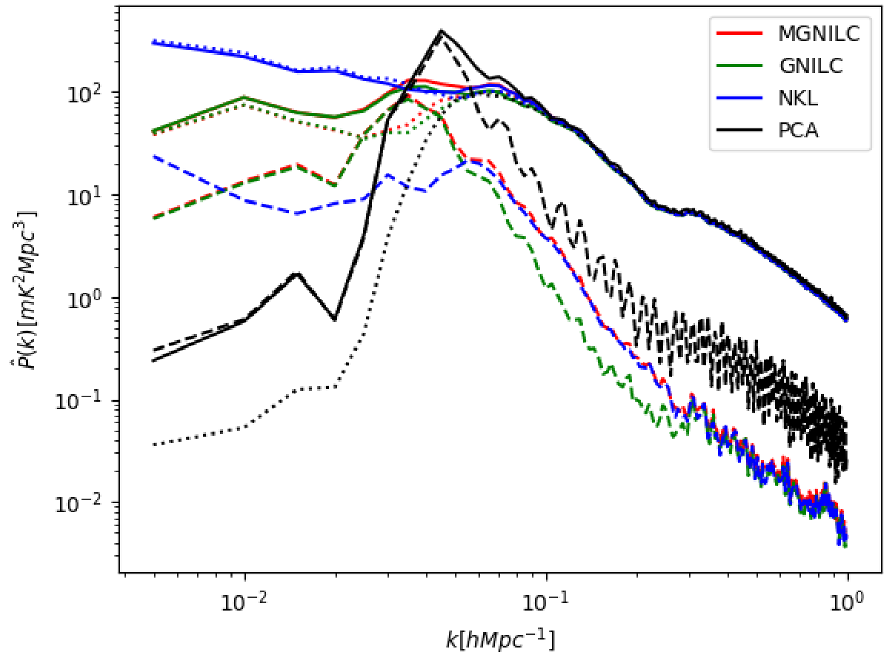

Figure 7.13: Power spectra illustrating the effects of residual foregrounds for various techniques in the low redshift case. These curves are organized in the same manner as Fig. 7.8.

mentioned earlier, PCA and GMCA are dominated by foregrounds at k values less than about 0.2 hMpc^{-1}. For the non-blind methods, we find that foreground residuals are strongest at low values of k. This is expected, as such scales will be dominated by contributions from low delays where foregrounds are strongest. However, foreground residuals are not quite as severe as in the higher redshift test case.

In Fig. 7.14, we show the spherical harmonic power spectra for various cleaning methods. This figure presents the same metric as was presented in Fig. 7.7. In this plot, notice first that PCA performs poorly at low k for the lower ℓ-modes. At these values of ℓ, the very lowest k modes were cleaned too aggressively, while the slightly higher k modes were not cleaned enough, leaving behind significant foreground residuals. PCA did not suffer from such under-cleaning at the higher ℓ-modes, however. For $\ell \gtrsim 200$, MGNILC and NKL provide very similar results, with both leaving behind significant foreground residuals. This is due to the fact that PCA was used to generate the foreground approximation in this test case. When using PCA to generate the foreground approximation, AIC chose to include too few modes. As a result, not all of the foregrounds were included in $\hat{f}$. Any foregrounds not included in $\hat{f}$ will not be cleaned by NKL or MGNILC, leading to the undercleaning observed here. At lower ℓ, however, it appears that NKL performs best, providing the least signal loss at the lowest k modes.

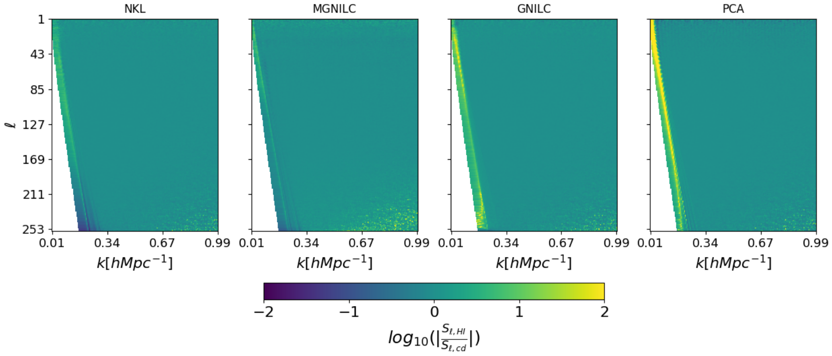

Figure 7.14: The ratio of the true spherical harmonic power spectrum ($S_{l,HI}$) to the spectrum recovered by cleaning and debiasing ($S_{l,cd}$) for the low-z test.

Performance of NKL Given Modified HI priors

One source of concern with respect to non-blind methods is how their performance may change subject to systematic effects or modified priors. We will leave systematics testing for future work. For now we will test the performance of NKL subject to modified HI priors.

In Shaw et al. (2015), it is mentioned that the CORA software package, which was used for this paper, assumes the HI power spectrum in the flat sky limit is equal to

$$P_{T_b}(\mathbf{k}, z, z') = \overline{T_b}(z)\overline{T_b}(z')(b + f\mu^2)^2 D_+(z)D_+(z')P_m(k). \tag{7.46}$$

In this formula, $\overline{T}_b$ is proportional to the HI abundance $\Omega_{HI}(z)$ and $b(z)$ is the HI bias.

Although modelling and measurement work has been done, the values of these parameters remain uncertain, particularly at higher redshifts. For the past two decades, much work has been done to derive estimates of Ω_{HI} from measurements. Figure 14 of Hu et al. (2020) conveniently provides a summary of $\Omega_{HI}(z)$ estimates out to $z = 5$. This figure shows increasing error bars and even disagreement between measurements as redshift increases. This becomes particularly severe for $z \gtrsim 4$.

When searching in the literature for information about the hydrogen bias, one will typically find constraints coming from cross-correlation studies. These cross correlation studies will present estimates of the product $\Omega_{HI} b r_{HI}$, where r_{HI} is a correlation coefficient. As a result, most of the information one finds in the literature about b by itself comes from modelling work. The model described in Castorina and Villaescusa-Navarro (2017) was presented at redshifts $0.8 < z < 5$ and showed significant uncertainties. Possible values of $b(z)$ spanned over a factor of ≈ 2 at any given

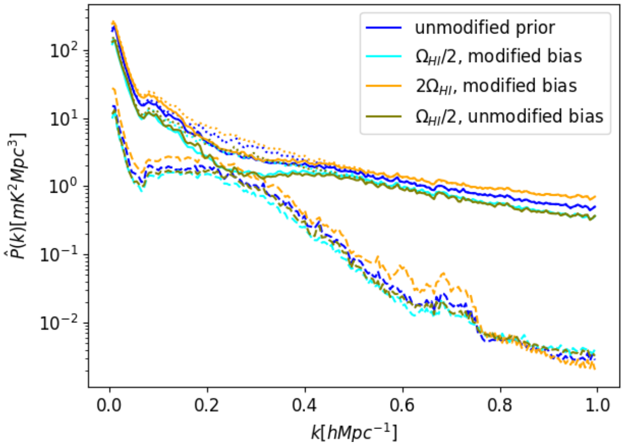

Figure 7.15: A summary of NKL performance subject to modified priors in the higher redshift scenario. The curves shown here should be interpreted in the same as Fig. 7.8.

redshift. This model however, did predict a smoothly increasing $b(z)$ as a function of redshift. As a simple way to test the effect of errors in the HI prior, we repeated NKL subject to modified HI priors. In particular, NKL was performed on the same maps as before, but provided with modified estimates of the HI covariance matrices. In these test cases we vary both the bias b and the abundance Ω_{HI} assumed by the prior. In particular, we scale Ω_{HI} by a constant factor and change the bias from a constant to a linear function. The coefficients of this linear curve are chosen to ensure $b(z_{min}) = 1$ and $b(z_{max}) = 2$. To summarize, we modify the priors in three different ways:

1. $\Omega_{HI} \to \Omega_{HI}/2$ and $b(z) = 1 \to b(z) = az + c$
2. $\Omega_{HI} \to 2\Omega_{HI}$ and $b(z) = 1 \to b(z) = az + c$
3. $\Omega_{HI} \to \Omega_{HI}/2$.

These robustness tests were performed both in the higher redshift scenario and the lower redshift scenario. The coefficients chosen for the bias were adjusted in each case to ensure that the bias varied from 1 to 2 over the redshift range.

In Fig. 7.15, we present results from the robustness check at higher redshifts. As can be seen, NKL is fairly robust to errors in the prior in this scenario. All priors have provided roughly similar levels of foreground residuals. One will also notice that the $\Omega_{HI}/2$ curves are lower and the $2\Omega_{HI}$ curve is higher. This is due to the fact that the KL step of NKL chooses the number of modes to discard based on the expected signal to ratio. This SNR estimate is of course dependent upon the assumed HI amplitude. It should also be noted here that the two $\Omega_{HI}/2$ curves look quite similar. This would imply that errors in the bias prior did not make much of a difference in this case.

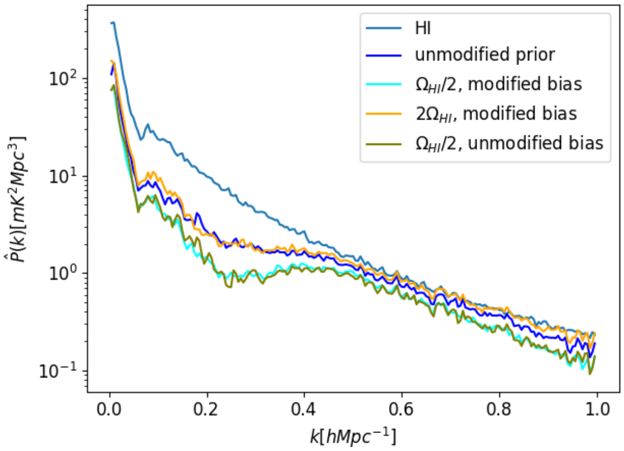

Figure 7.16: Debiased NKL power spectra in the high redshift test case, subject to modified priors.

In Fig. 7.16, we present debiased power spectra for the high redshift test case. As one could predict from Fig. 7.15, we see that the $2\Omega_{HI}$ provides roughly similar performance to the correct prior case upon debiasing. On the other hand, the cases with $\Omega_{HI}/2$ suffer from some additional signal loss.

In Fig. 7.17, we present a summary of results with varied priors in the low-z case. As in the higher redshift scenario, decreasing Ω_{HI} without adjusting the bias leads to roughly similar results. However, adjusting the bias seems to have more of an impact, with both the $\Omega_{HI}/2$ and $2\Omega_{HI}$ test cases showing larger power spectrum values at low k compared to the case with correct priors. Further testing found that NKL discarded the same number of modes in both the correct prior and $\Omega_{HI}/2$ with linear bias test cases. This would imply that the performance difference between the cyan and blue curves of Fig. 7.17 is due to differences in the eigenmodes generated for the KL transform, rather than NKL discarding fewer modes due to the change in prior.

In Fig. 7.18, we present debiased power spectra for the lower redshift scenario. In this figure, one sees that decreasing Ω_{HI} without adjusting the bias leads to similar results, with the cleaning being conducted slightly more aggressively. The test cases with modified bias assumptions oddly lead to better agreement, with some over estimation at the lowest values of k in the $2\Omega_{HI}$ case.

7.7 Discussion and Conclusion

In this paper we introduced the NKL and MGNILC techniques for cleaning foregrounds from HI intensity maps. Moreover we tested these methods, and various others, on full sky maps in two different bands of observations ([980 MHz, 1080 MHz] and [400 MHz, 500 MHz]). These tests

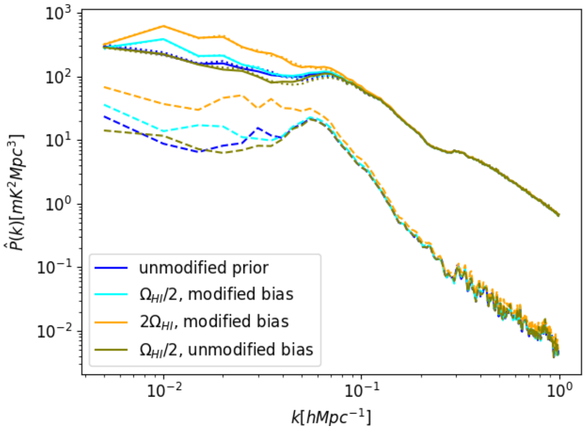

Figure 7.17: A summary of NKL performance subject to modified priors in the lower redshift scenario. The curves shown here should be interpreted in the same way as Fig. 7.8.

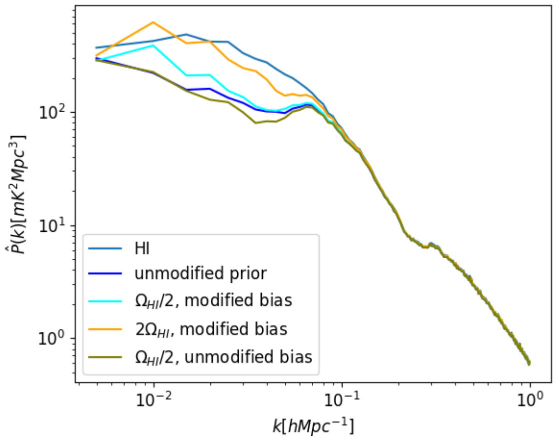

Figure 7.18: Debiased NKL power spectra in the lower redshift test case, subject to modified priors.

were conducted assuming a hypothetical telescope similar to MeerKAT operating in single-dish mode. Instrumental effects were modelled simply, with noise being drawn from a Gaussian distribution and beam effects modeled by convolution with Gaussians of frequency-dependent width. Moreover, we assumed full sky coverage and masked out the brightest 15 percent of pixels.

At higher redshifts, where foregrounds are more severe, we found that non-blind methods such as GNILC and NKL outperform blind methods by several orders of magnitude at large spatial scales. However, all methods suffered significant signal loss. The most accurate method was NKL, which was still off by an order of magnitude in the higher redshift test and a factor of 2 or 3 in the lower redshift case. It may be possible to compensate for such signal loss using a transfer function method. For instance, the work presented in Cunnington et al. (2023) seems to show promising results when signal loss is at the 50 percent level.

We also tested the robustness of NKL against modified priors. We found that the performance of NKL remains fairly stable while varying Ω_{HI} and the bias b. It was found that varying Ω_{HI} changed how aggressively NKL cleaned, while varying the bias lead to different behavior at low k in the lower redshift scenario. However, even when provided with modified priors, NKL still provided better performance at large length scales than the other map-space cleaning methods described in this paper.

We must emphasize that these tests do not tell the entire story of foreground removal. For instance, convolution with a Gaussian beam is not quite realistic. A more realistic beam would be more complicated than this, including sidelobes and asymmetries. Moreover, a realistic mapmaking process is not so simple as just convolving with a beam. One has to perform some sort of maximum likelihood estimate based on the raw data coming from the antennas. Such a process may lead to mapmaking artifacts that aren't so easily described. In addition, one must also consider inevitable imperfections in beam calibration. Such imperfections would likely provide their own artifacts and possibly add additional chromaticity to the foregrounds. We plan to address these issues in future work.

For perspective, we would like to recall the visibility-based work presented in Shaw et al. (2015). In that paper, polarized foregrounds were removed by projecting the visibilities onto the null-space of the polarized beam matrix. A KL transform was then used to clean the unpolarized foregrounds from the visibilities. That approach recovered the HI signal effectively down to $k \approx 0.02\,\mathrm{hMpc}^{-1}$ for a simplified version of the CHIME telescope operating between 400 and 500 MHz. One conclusion, however, was that the main beam widths needed to be understood to an accuracy of 0.1% for this method to work.

In this paper, our motivation for further investigation of map-based techniques was that they may be more robust to beam mis-calibration than are visibility-based methods. For instance, Shaw's methods would not work in a situation where beamwidth errors are at the 1% level. This leads to an incorrect model of C_{FG}, preventing the Karhunen-Loève transform from being effective. We mention C_{FG} in specific because it is more sensitive to errors than C_{HI}. Having 1% errors in C_{FG} can be hazardous when dealing with such a large dynamic range as in HI intensity mapping. However, NKL and MGNILC estimate C_{FG} directly from the data, hopefully providing it with some robustness against calibration issues.

As a next step, we intend to use the driftscan package, which employs the m-mode formalism of Shaw et al. (2015), to produce more realistic maps. This approach creates visibilities from a beam model. We can then use a modified version of the package to generate Wiener-filtered maps from these visibilities. We will then introduce beam-width errors to test how robust these methods are. We also intend to introduce a method similar to NKL which acts in the visibility domain. Such a method may be robust to errors in the beam widths, as the foreground covariance would be estimated from the data and not from a beam model.

7.8 Details of The Error Analysis

In this appendix, we provide derivations of the expressions provided in Section 7.4. We derive the formulas in this subsection using first order perturbation theory. To begin, let's consider the generalized Eigenvalue problem at hand. It turns out that equation (7.23) can be massaged into

$$\mathbf{C}_S^{-1/2}\hat{\mathbf{C}}_{FG}\mathbf{C}_S^{-1/2}\phi_i'' = \lambda_i\phi_i''. \tag{7.47}$$

In this formula, $\phi_i'' = \mathbf{C}_S^{1/2}\phi_i'$ and $\mathbf{C}_S^{1/2}$ is the Hermitian square root of $\mathbf{C}_S$. Note that this is the same as the equation we would get had we set up the generalized Eigenvalue equation using variables whitened with $\mathbf{C}_S^{-1/2}$.

It turns out that the math involved in this problem is easier when whitened variables are used. As such, any symbols used during the rest of this derivation will be used to represent whitened variables. So, for instance, f_p will refer to row p of $\mathbf{C}_S^{-1/2}\underline{\mathbf{f}}$. Moreover, ϕ_i' will refer to the whitened eigenvector ϕ_i''. This will not change our results, since it can be easily shown that $E[\mathcal{F}_i'\mathcal{H}_k']$ does not change upon whitening of the variables.

To first order, the perturbed eigenvector ϕ_i' will look like

$$\phi_i' = \phi_i + \sum_{j\neq i}\frac{\phi_j^T\mathbf{\Delta}\phi_i}{\lambda_i - \lambda_j}\phi_j. \tag{7.48}$$

In this formula,

$$\mathbf{\Delta} = \mathbf{\Delta}_{FG} + \mathbf{C}_{h'h'} + \mathbf{C}_{fh'} + \mathbf{C}_{n'h'} + \mathbf{C}_{n'n'} + \mathbf{C}_{fn'}. \tag{7.49}$$

Refer to Section 7.4 for definitions of the terms on the right of equation (7.49). Let's now consider the correlation between residual foregrounds and HI. We will find that

$$E[\mathcal{F}_i'\mathcal{H}_k'] = E[(\phi_i + \Delta\phi_i)^T f_p h_p^T(\phi_k + \Delta\phi_k)]. \tag{7.50}$$

Refer to Section 7.4 for definitions of $\mathcal{F}_i'$ and $\mathcal{H}_i'$. Note also that a_p refers to column p of matrix a. This expectation value ought to be equal to 0 in the unperturbed case. We will also discard the second order term, giving

$$E[\Delta\phi_i^T f_p h_p^T\phi_k] + E[\phi_i^T f_p h_p^T\Delta\phi_k]. \tag{7.51}$$

Recall that $\Delta\phi_x$ really consists of several contributions

$$\Delta\phi_x = \sum_{j\neq x} \frac{\phi_j^T(\mathbf{\Delta}_{FG} + \mathbf{C}_{h'h'} + \mathbf{C}_{fh'} + \mathbf{C}_{n'h'} + \mathbf{C}_{n'n'} + \mathbf{C}_{fn'})\phi_x}{\lambda_x - \lambda_j}\phi_j. \quad (7.52)$$

The term involving Δ_{FG} will not contribute to $E[\mathcal{F}_k'\mathcal{H}_i']$ as it will consist of the expectation of products of 3 terms depending on the foregrounds and 1 term depending on the HI. Since the HI and foregrounds are uncorrelated, this should result in an expectation value of 0. Similar arguments can be made for $\mathbf{C}_{h'h'}$, $\mathbf{C}_{n'h'}$, $\mathbf{C}_{n'n'}$ and $\mathbf{C}_{fn'}$, which will each involve odd degrees of foregrounds, noise or HI, resulting in 0 (or at least small) expectation. Small expectation is mentioned since terms like h_p' and f_p will be correlated in principle, since h_p' does depend on the foregrounds. However, we will assume that such correlation are small.

It will be the case however, that $\mathbf{C}_{fh'}$ will create a non-negligible correlation between the foreground residuals and HI. We will have

$$E[\mathcal{F}_i'\mathcal{H}_k'] = \sum_{j\neq i} \frac{1}{\lambda_i - \lambda_j}(E[\phi_j^T \mathbf{C}_{fh'}\phi_i\phi_j^T f_p h_p^T \phi_k] + \sum_{j\neq k} \frac{1}{\lambda_k - \lambda_j} E[\phi_i^T f_p h_p^T \phi_j \phi_j^T \mathbf{C}_{fh'}\phi_k]). \quad (7.53)$$

Let's start by considering the first term of the right side of equation (7.53). Applying the definition of $\mathbf{C}_{fh'}$, we obtain

$$\sum_{j\neq i} \frac{1}{\lambda_i - \lambda_j} \sum_q \frac{1}{N_p} E[\phi_j^T (h_q' f_q^T + f_q h_q'^T)\phi_i \phi_j^T f_p h_p^T \phi_k]. \quad (7.54)$$

For first term of equation (7.54), we will have

$$\Sigma_{j\neq i} \frac{1}{\lambda_i - \lambda_j} \Sigma_q \frac{1}{N_p} E[\phi_j^T h_q' f_q^T \phi_i \phi_j^T f_p h_p^T \phi_k]. \quad (7.55)$$

Let us now assume that h' and f are uncorrelated. This assumption is not strictly true, as the filter used to generate h' will be a function of the foregrounds. However, we found in numerical tests that correlations between h' and f were small. So, this approximation is safe to make. Next, we rearrange terms inside the expectation value and using our recent assumption, we can express equation (7.55) as

$$\sum_{j\neq i} \frac{1}{\lambda_i - \lambda_j} \sum_q \frac{1}{N_p} E[\phi_j^T h_q' h_p^T \phi_k] E[\phi_i^T f_q f_p^T \phi_j]. \quad (7.56)$$

Next, note that the rapid angular variation of the signal implies that terms of in the sum of equation (7.56) will be suppressed when q differs from p. As such, we will approximate equation (7.56) as

$$\Sigma_{j\neq i}\frac{1}{\lambda_i-\lambda_j}\frac{1}{N_p}E[\phi_j^T h'_p h_p^T\phi_k]E[\phi_i^T f_p f_p^T\phi_j]. \tag{7.57}$$

Assuming that $E[f_p f_p^T]$ varies little as a function of p, then we can say $E[\phi_i^T f_p f_p^T\phi_j]\approx 0$. This assumption should be a good one when chunk sizes are small enough that the foreground temperature varies little within each chunk. On other hand, consider the case where one has a single chunk covering the whole map. In this case, the columns of f will correspond to individual lines of sight. In such a case, $E[f_p f_p^T]$ will vary significantly with p and our assumption is no longer valid.

Thus, we find that the first term on the right side of equation (7.54) is approximately 0, as long as the chunk sizes are chosen to be adequately small. As for the second term of equation (7.54), we can do some rearranging and apply our previous assumptions to say that

$$\begin{aligned}\sum_{j\neq i}\frac{1}{\lambda_i-\lambda_j}\sum_q\frac{1}{N_p}E[\phi_j^T f_q h_q'^T\phi_i\phi_j^T f_p h_p^T\phi_k]&\\ \approx\frac{1}{N_p}\Sigma_{j\neq i}\frac{1}{\lambda_i-\lambda_j}&E[\phi_j^T f_p f_p^T\phi_j]E[\phi_i^T h'_p h_p^T\phi_k].\end{aligned} \tag{7.58}$$

Let's define the matrix $\alpha_{ik}=E[\phi_i^T h'_p h_p^T\phi_k]$. In a numerical test, we found that α_{ik} roughly follows the relation

$$\alpha_{ik}\approx\begin{cases}0 & \lambda_i<1\\ \delta_{ik} & \lambda_i>1.\end{cases} \tag{7.59}$$

This makes sense, as h'_p is just a version of h_p that has been smoothed using some filter (either PCA or DAYENU). We know that modes with large λ_i will couple strongly to the foregrounds, and thus to the smooth parts of h_p. As such, it should be the case that

$$E[\phi_i^T\overline{h}_p h_p^T\phi_k]\approx E[\phi_i^T h_p h_p^T\phi_k]=\delta_{ik} \tag{7.60}$$

when λ_i is large. On the other hand, for smaller λ_i, one would expect that the mode would couple most strongly to higher delay parts of h_p. Since h'_p has been smoothed, it should be the case that

$$|E[\phi_i^T h'_p h_p^T\phi_k]|<<1 \tag{7.61}$$

for modes with small λ_i. Using this result, we obtain

$$\sum_{j\neq i}\frac{1}{\lambda_i-\lambda_j}\sum_q\frac{1}{N_p}E[\phi_j^T f_q h_q'^T\phi_i\phi_j^T f_p h_p^T\phi_i]\approx\frac{\delta_{ik}}{N_p}\sum_{j\neq i}\frac{\alpha_{ii}\lambda_j}{\lambda_i-\lambda_j}. \tag{7.62}$$

We can go through the same process for the other term involved in equation (7.53). This results in

$$\frac{\delta_{ik}}{N_p}\Sigma_{j\neq i}\frac{\alpha_{jj}\lambda_i}{\lambda_i-\lambda_j}, \tag{7.63}$$

Where δ_{ik} is a Kronecker delta function. Combining these two terms, we find

$$E[\mathcal{F}_i'\mathcal{H}_k'] \approx \frac{\delta_{ik}}{N_p}\Sigma_{j\neq i}\frac{(\alpha_{ii}\lambda_j+\alpha_{jj}\lambda_i)}{\lambda_i-\lambda_j}. \tag{7.64}$$

7.9 An afterword

In this chapter, I showed a truncated version of the NKL paper co-authored by Peter and I. Everything mentioned in this paper is true, though there is an aspect to the performance I had not considered at the time the paper was written. It turns out that the seemingly superior performance of NKL at low delays is mostly artificial. The KL transform was taking signal from higher $|\tau|$ and moving it down to lower $|\tau|$. Upon closer inspection none of the methods used really give one much of any information about what happens inside the wedge at higher redshifts.

At lower redshifts, NKL provides roughly the same level of performance as a scale-dependent version of PCA.

8 VISIBILITY KARHUNEN-LOÈVE

In addition to NKL, I also considered a similar method that acts on visibilities, which I referred to as Visibility Karhunen-Loève (VKL). One may recall that the method described in (Shaw et al. (2015)) uses an SVD to remove polarized foregrounds, followed by a KL transform to remove unpolarized foregrounds. These techniques were all applied in m-mode space, and were found to require highly precise beam calibration in order to work. My hope with this project was to come up with a foreground cleaning method that acts directly on visibilities but without such stringent requirements on accurate beam calibration. After testing, I found that the VKL method didn't really perform any better than foreground avoidance. However, it may still be worth discussing this work a bit.

I was motivated to try out the VKL idea for the sake of relaxing calibration requirements. The map-making process requires one to know the beam patterns of their antennas, and as of now it is unclear exactly how sensitive this process is to errors in the beam models. As mentioned when I discussed the delay-spectrum approach, working one baseline at a time reduces calibration requirements by eliminating the need for relative phases between baselines to be understood. Thus, I thought it would be a good idea to see if I could clean foregrounds strictly in visibility-space.

For this work, I generated visibilities using the *driftscan* software package [1]. This software assumes a CHIME-like instrument and provides one with beam patterns and visibilities in m-mode space. For my VKL work, I used a miniature version of the CHIME instrument operating from 400 to 500 MHz. The details of this are given in table 8.1.

---

[1] `https://github.com/radiocosmology/driftscan`

Bandwidth	400 to 500 MHz
n_{ch}	256
$\Delta\nu$	0.390625 MHz
T_{sys}	50 K
Cylinder Width	20 m
f-ratio	0.25
Cylinders	2 cylinders, oriented N-S, placed side-by-side
Feeds	64 dual-polarized dipoles per cylinder
Feed Spacing	0.3 m
N_{days}	733
Telescope Latitude	45°
Feed H-plane FWHM	$2\pi/3$
Feed E-plane FWHM	$0.7 * 2\pi/3$

Table 8.1: Parameters describing the hypothetical telescope used in this paper.

The VKL method can be summarized simply as follows:

1. Let $V_{ij}^{m}(\nu)$ be the m-mode space visibility for baseline ij at frequency ν. Send this visibility to real space using an IDFT as

$$V_{ij}^{\phi_n}(\nu) = \sum_{m=-m_{max}}^{m_{max}} V_{ij,m}(\nu) e^{im\phi_n}. \tag{8.1}$$

   In this equation, m_{max} is the largest value of m probed by the baseline of interest at the lowest frequency channel. This typically results in one having a few hundred azimuthal samples for their visibility.

2. Use the result of the previous step to form Stokes-I visibilities (optional).

3. Form a foreground approximation using either DAYENU or PCA, as was described in chapter 7.

4. Form an estimate of $\mathbf{C}_{HI}$ for the baseline of interest. This should be a $n_{ch} \times n_{ch}$ matrix and can be estimated using one's beam model and priors on the HI signal.

5. Form an estimate of $\mathbf{C}_{FG}$ from the foreground approximation formed in step 2.

6. Use these estimates of $\mathbf{C}_{FG}$ and $\mathbf{C}_{HI}$ to clean the visibilities with a KL transform.

As one can see, this method is rather simple, and is pretty much a corollary of the work I did in the NKL paper. I believed that this method would be especially robust against beam errors since only the HI covariance estimate depends on one's understanding of their beam patterns. In my previous work, it appeared to me that the KL cleaning process is not particularly sensitive to the specifics of one's model of $\mathbf{C}_{HI}$.

After performing the cleaning, I found that VKL was not capable of recovering really any information from foreground dominated delays. In figure 8.1, I show power spectrum estimates obtained after cleaning simulated visibilities with VKL.

I obtained the power spectrum estimate of the “VKL" curve using VKL-cleaned visibilities from 10 short baselines. The cutoff SNR chosen was 10. These baselines were all oriented in the north-south direction, the longest one being $4\,m$. Note that the visibilities used were all Stokes-I visibilities. I also did not include noise in this analysis. I then estimated the power spectrum from these visibilities using the optimal quadratic estimator method. In this plot, the acronym OQE stands for Optimal Quadratic Estimator. The “PCA, OQE" curve was obtained in a similar way, where the covariance used for cleaning was the estimated covariance of the foreground approximation. In this PCA case, I removed 120 modes from each visibility. For the “avoidance, delay spec"

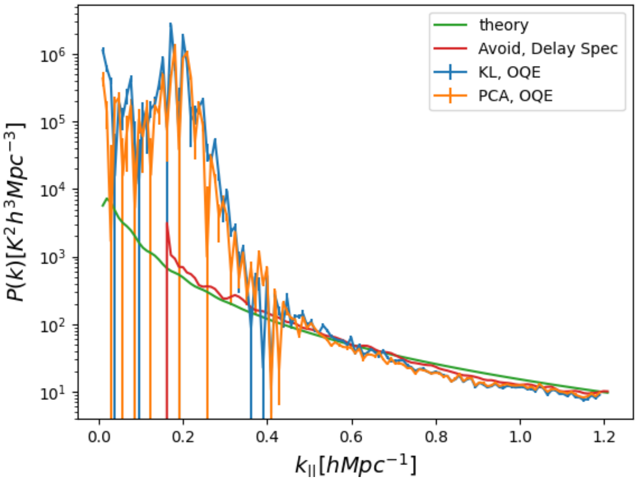

Figure 8.1: Comparison of foreground avoidance and the VKL cleaning approach combining data from 10 short baselines and ignoring noise. These baselines all probe a perpendicular k-space bin with $0 < k_\perp < 0.01hMpc^{-1}$. The blue curve represents the performance of VKL.

curve, I performed foreground avoidance on a set of visibilities (without noise) then formed power spectrum estimates using the delay spectrum approach. The “theory" curve provides the correct power spectrum result.

It should be noted that VKL and PCA give roughly the same results in this case. Moreover, they both give somewhat worse results than avoidance. In fact, avoidance was able to reasonably estimate the power spectrum down to $k_{||} \approx 0.2hMpc^{-1}$. I believe the relatively poor performance of VKL and PCA here is due to the fact that the foregrounds vary so much as a function of ϕ. As a result, VKL and PCA will aggressively clean all lines of sight, even those with relatively mild foregrounds. As such, these methods will remove signal that would be easily retrievable when taking an avoidance approach.

Thus, it seems for now that foreground avoidance is the best choice when calibration of the beam patterns and gain are not trusted. In the case of high levels of confidence in the calibration process, one would should consider using the SVD/KL method proposed in Shaw et al. (2015). That being said, it would be interesting to see how GPR would perform in a context such as this, since as of now testing of this method has pretty much all been done at low redshift.

9 SUMMARY AND IDEAS FOR THE FUTURE

9.1 Summary

In this thesis, I began by reviewing some basic cosmology and presenting some open questions in the field. Currently, some of the most interesting questions are “what is the nature of dark energy?", “what is causing the Hubble tension?" and “what is the nature of inflation?". Traditionally, we have sought to answer such questions using CMB observations, galaxy surveys and supernova observations, among other approaches. Much of what we know comes from the combination of CMB and galaxy surveys. Combining observations from both of these methods breaks degeneracies, serves as a double check and further tightens parameter constraints.

HIM could potentially serve as a third important LSS probe. In the post-EOR era, HIM would serve as a complement to galaxy surveys, providing a double check and higher sensitivity for $z < 6$. Such low redshift observations would be helpful for constraining dark energy, neutrino masses, inflationary physics and more. During the EOR ($6 < z < 20$), HIM would provide us with information about the reionization process. Lastly, HIM would provide a unique probe of the dark ages (observable for $20 < z < 150$). Such dark ages observations would provide a unique window into linear cosmology at small scales, which in turn would provide greater information about inflationary physics.

Although promising, HIM is still a young field and there are systematic effects that are not yet fully understood. Two severe systematic effects are mutual coupling and the presence of foregrounds. In the HIM context, the presence of unmitigated foregrounds would prevent one from observing the HI signal at low $k_{||}$. For post-EOR observations, such foregrounds would prevent some or perhaps all BAO scales from being observed. In Sailer et al. (2021), under pessimistic foreground assumptions, it was estimated that PUMA would still provide pretty good constraining power on cosmological parameters when its data is analyzed in conjuction with other experiments. However, these pessimistic assumptions still allow most BAO modes to be observed. In realistic conditions, we may find that more $k_{||}$ modes are inaccesible than what was assumed by the authors of Sailer et al. In such a case, the constraining power of HIM may be more severely limited in the post-EOR era.

Various cleaning methods have been proposed, such as PCA, GMCA, GNILC, GPR, FastICA and so on. These methods have been mostly been tested on simulations at low redshifts and under ideal assumptions about the instrument used. Moreover, they seem to all provide more or less equal performance. One exception to this is the work of Shaw et al. (2015), which was found to work well in the presence of polarized foregrounds at a somewhat higher redshift ($z \sim 2.15$). However, this method seems to require accurate beam calibration to work. My own tests of methods at this same redshift in the presence of polarized foregrounds seems to show none of the other methods (NKL, PCA, FastICA, GMCA, GNILC) working particularly well. Essentially, at least in my simulations, it seems like we can't extract much information from delays for which the foregrounds are dominant. In my tests, I found that none of these methods actually seem to work any better than avoidance when considering higher redshifts with polarized foregrounds included. Of course, this was a single test assuming a particular instrument and using the most pessimistic simulation software available.

We also have to consider the issue of mutual coupling. The works of Kwak et al. (2024) and Kern et al. (2019) show that port-to-port coupling will lead to a pedestal in one's data. It was shown in Kwak et al. (2024) that the pedestal caused by noise alone will likely be several orders of magnitude higher than the HI signal, even if a low level of of coupling, $S_{21} \leq -100\, dB$, is assumed. In my antenna paper, I sought to minimize the port-to-port coupling by employing deep dishes, collars and optimized Vivaldi feeds. Despite these efforts, I was not able to get S_{21} low enough to push the pedestal below the HI signal.

In Kern et al. (2019), an analysis effort was taken to remove this port-to-port coupling contribution. This analysis assumed that the pedestal would be limited to only a certain subset of (LST, delay) space. Under these assumptions, it was shown that the coupling effect could be effectively removed for baselines closer to E-W orientation. However, baselines that slowly vary with LST will not be cleanable with this approach. Moreover, we know that the frequency dependence of the pedestal will depend both on the chromaticity of the sky, the amplifier noise and the antenna S-parameters. In the case of chromatic noise or chromatic S_{21}, we may find that the pedestal is not well constrained in delay space either.

Thus, we see that port-to-port coupling is a major issue in HIM. Even with very low levels of coupling we will still end up with a large pedestal in our data that dwarfs the HI signal. Moreover, the SVD method proposed in Kern et al. (2019) will only be effective for baselines with high fringe rates and when the S-parameters and amplifier noise are spectrally smooth. As of now, despite our best efforts such port-to-port coupling will make much of the data collected by an HIM instrument unusable.

Mutual coupling will also lead to distortions in the beams. Such distortions are troublesome since they will make the beam patterns for each antenna more chromatic and more complicated in terms of angular dependence. Moreover, it will increase variation in beam patterns from antenna to antenna, possibly making redundant calibration methods ineffective. It is also possible that the presence of mutual coupling could increase the effective size of our antennas, increasing the farfield distance and making beam measurement from drones much more difficult.

9.2 Ideas for the Future

As mentioned earlier, HIM is still a young field and there is much work left to be done. In this section, I will present some ideas for how to handle certain systematics and for how HIM instruments may be designed differently.

Cleaning Up Noise Pedestal With Calibration

In my previous sections, I summarized how the LNAs used in HIM interferometers ought to create a pedestal of noise in one's data. In my own work, I found that it will be very difficult to get port-to-port coupling below $-100\,dB$, at least for closely spaced antennas. Comparing this to the test case shown in Kwak et al. (2024), such coupling ought to get us to within 2 orders of magnitude of the HI signal.

Another way we could remove the noise pedestal is through calibration. Recall that the noise pedestal should be given by equation 5.52, which itself depends on LNA temperatures, reflection coefficients and S-parameters. It is not unreasonable to imagine attaching additional hardware to the antennas in order to measure the LNA noise temperatures along with the S-parameters that govern the coupling. Assuming an instrument with $-100\,dB$ coupling, we'd really only need to measure such parameters down to the percent level, which does not seem impossible. The frequency with which we would need to perform such measurements would depend on how stable these parameters are with respect to time.

Minimize Polarization Leakage Through Antenna Design

As mentioned in earlier in this thesis, polarized foregrounds are expected to be more chromatic and thus more concerning than unpolarized foregrounds. In Shaw et al. (2015), it was found that polarized foregrounds may be removed by projecting onto the null-space of the polarized beam matrix. However, this approach may not work out for two reasons. For one, it requires a model of the beam, and it is unclear exactly how accurate that needs to be. Second, this method cleans polarized foregrounds on an m-mode by m-mode basis. As such, if we switch back to azimuthal space or build a map, we may find that the unpolarized foregrounds occupy many more degrees of freedom than they did before. This could be overcome to some extent in map-space with accurate calibration, though it is unclear how good calibration needs to be.

In my work, I sought to find a foreground cleaning method that could handle polarized foregrounds without needing the precise calibration required by the null-space projection approach. Unfortunately, I was unable to extract much cosmological information from polarized foreground-dominated delays. It is possible that someone may come along with a method that does much better than mine or the null-space projection method, but it is not clear to me what such a method would look like. As such, if I were to stay in the field, I would not bet on these foreground removal methods working out. Instead, I would focus on cutting down on polarized foregrounds as much as possible through optimizing the hardware. We can do this by minimizing cross-polarization in the beams and ensuring that the beam patterns have a high level of azimuthal symmetry.

Setting More Modest Bandwidth Goals

Oftentimes, one will see proposals for future HIM experiments describing very large bandwidths. For example CHORD and PUMA are both proposing factors of 5 or more in bandwidth. This is done to capitalize on the fact that for HIM, a large bandwidth corresponds to larger observational volume. However, this approach will come with downsides.

One problem with using such a large bandwidth would be the cost of digitization. For PUMA, we would be observing from $200\,MHz$ up to $1100\,MHz$. If we mix this signal down, we will be digitizing from $0\,MHz$ up to $900\,MHz$. In order to Nyquist sample this signal, we would need to collect 1800 megasamples per second. As of the time I'm writing this, Texas instruments sells an ADC called “ADC12D1800RFIUT" for which they charge over 2000 dollars. Note that this ADC has two inputs, so we would only need one per dish. On the other hand, they also produce a model called “ADS4128" that samples at 200 megasamples per second (corresponding to a 100 MHz bandwidth). This ADC only costs about 25 dollars if you buy in bulk from Mouser. Moreover, the outputs of these ADCs presumably will go to some sort of Field Programmable Gate Array (FPGA). In this context, the FPGA would be used to convert the ADC time stream into spectra. The high rate of data output from the large bandwidth ADCs will of course also push the need for higher grade FPGAs, further increasing electronics costs.

To be more accurate, we really have two competing effects in terms of cost. Electronics per antenna and number of antennas. Imagine our goal is to observe over $0.3 \leq z \leq 6$ and that we have two options. Option 1 is to use one instrument that covers the entire redshift range. Option 2 is to build several telescopes, each of which covers a subset of the redshift range. As mentioned already, option 1 leads to more money being spent on electronics. However, it also leads to less money being spent on antennas, since we only have to build the one array. Depending on how cheaply antennas can be built, we may save money by pursuing the second option. Of course, this may change in the future as the electronics involved become cheaper.

Budgetary concerns aside, one also deals with worse antenna performance when using a very wide bandwidth. In my experience, it is tough to design an instrument that has high performance over a $5.5:1$ bandwidth. In particular, I ended up with higher coupling than I would've liked at the low frequencies and ugly looking beams at the high frequencies. Such things would've been easier to control on a smaller bandwidth instrument. This would also give us some more freedom in terms of the type of antenna chosen. I used a Vivaldi style feed to maximize the bandwidth, but we of course would have more options available if this constraint were to be removed.

In summary, I believe it would likely be beneficial to set more modest bandwidth goals for future HIM instruments. This is because a more modest instrument will be easier to optimize, minimizing systematics and providing better sensitivity. Moreover, they may be more cost effective, at least in the near future.

Replacing Dishes With Phased Arrays

In the future, we in the field may want to consider replacing the dish-based antenna elements with phased arrays of similar size. I think this option ought to be considered for several reasons, which I will now provide in the form of a list

1. Using phased arrays would eliminate complications that would come with repointing dishes.
2. It would be possible to synthesize many beams simultaneously, increasing the ability to simulatenously observe the fast radio sky, including Fast Radio Bursts (FRB)s and increasing mapping speed.
3. Since the dipoles that make up the phased arrays are small, short-circuit scattering terms might be at an acceptable level.

However, using phased arrays would also come with downsides compared to the deep parabolic dishes considered by the HIRAX collaboration and myself. For one, the phased array beams may become quite messy as we steer them away from zenith, leading to more difficult calibration and perhaps decreased sensitivity. Moreover, the wide beams used by individual elements could increase port-to-port coupling to the point where it is difficult to handle. Lastly, forming multiple beams at a time would likely require a lot of computing power after digitization unless a suitable beam forming chip could be found.

9.3 Concluding Thoughts

HIM is a technique with significant scientific promise, but it does suffer several systematic effects that still need to be understood and minimized. Although the list is longer, I have mostly focused on foregrounds and mutual coupling.

In chapter 5, I discussed how mutual coupling comes in the form of beam effects and a noise pedestal. On this topic, I think there is good reason to be hopeful. Supposing $|S_{21}| < 100\,dB$, then we could effectively remove the pedestal as long as we calibrate the appropriate parameters, such as amplifier temperature and antenna impedances, down to the percent level.

For the beam effects of mutual coupling, the discussion is a bit more complicated. Simulated tests of antenna delay response at zenith from both myself and Saliwanchik et al. (2021) seem to indicate that we won't take on too much additional chromaticity as long as the antennas are properly designed. However, these simulations used "periodic boundary conditions" which just an approximation of a large array. It is unclear to me whether this approximation will accurately account for the chromaticity caused by mutual coupling. Moreover, dish-based instruments such as PUMA will need to cycle through several pointings in order to capture a larger fraction of the sky. The simulations conducted by Saliwanchik et al. (2021) and I were done with antennas pointing straight up towards the sky. It's reasonable to imagine this chromaticity becoming worse once the antennas are pointed away from zenith. Moreover, it is unclear to what extent this coupling will affect the calibration process and how much this will matter when we estimate the power spectrum.

In my work on foregrounds, I found that none of the available cleaning methods seem to work any better than avoidance, at least in the case of using simulated CRIME maps in the 400 to 500 MHz band. The methods outlined in Shaw et al. (2015) could in principle work, but I am not optimistic that we will be able to calibrate the beams well enough given the strict requirements and the fact that we have to contend with mutual coupling. However, this may not be the end of the world. In the post-EOR context, some of the BAO will still be accessible as long as antenna beams have an adequately narrow delay response and low polarization-leakage. To really understand the implications of this, we'd have to conduct a rigorous analysis in the vein of Sailer et al. (2021).

Overall, HIM is still facing many challenges, some of which might not be solvable. The main impact of these systematics is that they limit access to low k modes. In the dark-ages context, this likely won't be a big deal since high k modes carry a lot of interesting cosmological information during that era. For post-EOR observations, these systematic effects will certainly limit our ability to extract cosmological information. It is unclear to me to exactly what extent this will be the case, but there is certainly hope that HIM will be fruitful for post-EOR cosmology.

REFERENCES

Adame, A. G., 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. Carnero 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. Gontcho 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. Le Goff, L. Le 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. 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, C. Ravoux, 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. 2024. Desi 2024 vi: Cosmological constraints from the measurements of baryon acoustic oscillations. `2404.03002`.

Ade, P. A. R., Z. Ahmed, M. Amiri, D. Barkats, R. Basu Thakur, C. A. Bischoff, D. Beck, J. J. Bock, H. Boenish, E. Bullock, V. Buza, J. R. Cheshire, J. Connors, J. Cornelison, M. Crumrine, A. Cukierman, E. V. Denison, M. Dierickx, L. Duband, M. Eiben, S. Fatigoni, J. P. Filippini, S. Fliescher, N. Goeckner-Wald, D. C. Goldfinger, J. Grayson, P. Grimes, G. Hall, G. Halal,

M. Halpern, E. Hand, S. Harrison, S. Henderson, S. R. Hildebrandt, G. C. Hilton, J. Hubmayr, H. Hui, K. D. Irwin, J. Kang, K. S. Karkare, E. Karpel, S. Kefeli, S. A. Kernasovskiy, J. M. Kovac, C. L. Kuo, K. Lau, E. M. Leitch, A. Lennox, K. G. Megerian, L. Minutolo, L. Moncelsi, Y. Nakato, T. Namikawa, H. T. Nguyen, R. O'Brient, R. W. Ogburn, S. Palladino, T. Prouve, C. Pryke, B. Racine, C. D. Reintsema, S. Richter, A. Schillaci, R. Schwarz, B. L. Schmitt, C. D. Sheehy, A. Soliman, T. St. Germaine, B. Steinbach, R. V. Sudiwala, G. P. Teply, K. L. Thompson, J. E. Tolan, C. Tucker, A. D. Turner, C. Umiltà, C. Vergès, A. G. Vieregg, A. Wandui, A. C. Weber, D. V. Wiebe, J. Willmert, C. L. Wong, W. L. K. Wu, H. Yang, K. W. Yoon, E. Young, C. Yu, L. Zeng, C. Zhang, and S. Zhang. 2021. Improved constraints on primordial gravitational waves using planck, wmap, and bicep/keck observations through the 2018 observing season. *Phys. Rev. Lett.* 127:151301.

Ade, P. A. R., R. W. Aikin, M. Amiri, D. Barkats, S. J. Benton, C. A. Bischoff, J. J. Bock, J. A. Brevik, I. Buder, E. Bullock, G. Davis, P. K. Day, C. D. Dowell, L. Duband, J. P. Filippini, S. Fliescher, S. R. Golwala, M. Halpern, M. Hasselfield, S. R. Hildebrandt, G. C. Hilton, K. D. Irwin, K. S. Karkare, J. P. Kaufman, B. G. Keating, S. A. Kernasovskiy, J. M. Kovac, C. L. Kuo, E. M. Leitch, N. Llombart, M. Lueker, C. B. Netterfield, H. T. Nguyen, R. O'Brient, R. W. Ogburn, A. Orlando, C. Pryke, C. D. Reintsema, S. Richter, R. Schwarz, C. D. Sheehy, Z. K. Staniszewski, K. T. Story, R. V. Sudiwala, G. P. Teply, J. E. Tolan, A. D. Turner, A. G. Vieregg, P. Wilson, C. L. Wong, and K. W. Yoon. 2014. Bicep2. ii. experiment and three-year data set. *The Astrophysical Journal* 792(1):62.

Ade, Peter, James Aguirre, Zeeshan Ahmed, Simone Aiola, Aamir Ali, David Alonso, Marcelo A. Alvarez, Kam Arnold, Peter Ashton, Jason Austermann, Humna Awan, Carlo Baccigalupi, Taylor Baildon, Darcy Barron, Nick Battaglia, Richard Battye, Eric Baxter, Andrew Bazarko, James A. Beall, Rachel Bean, Dominic Beck, Shawn Beckman, Benjamin Beringue, Federico Bianchini, Steven Boada, David Boettger, J. Richard Bond, Julian Borrill, Michael L. Brown, Sarah Marie Bruno, Sean Bryan, Erminia Calabrese, Victoria Calafut, Paolo Calisse, Julien Carron, Anthony Challinor, Grace Chesmore, Yuji Chinone, Jens Chluba, Hsiao-Mei Sherry Cho, Steve Choi, Gabriele Coppi, Nicholas F. Cothard, Kevin Coughlin, Devin Crichton, Kevin D. Crowley, Kevin T. Crowley, Ari Cukierman, John M. D'Ewart, Rolando Dünner, Tijmen de Haan, Mark Devlin, Simon Dicker, Joy Didier, Matt Dobbs, Bradley Dober, Cody J. Duell, Shannon Duff, Adri Duivenvoorden, Jo Dunkley, John Dusatko, Josquin Errard, Giulio Fabbian, Stephen Feeney, Simone Ferraro, Pedro Fluxà, Katherine Freese, Josef C. Frisch, Andrei Frolov, George Fuller, Brittany Fuzia, Nicholas Galitzki, Patricio A. Gallardo, Jose Tomas Galvez Ghersi, Jiansong Gao, Eric Gawiser, Martina Gerbino, Vera Gluscevic, Neil Goeckner-Wald, Joseph Golec, Sam Gordon, Megan Gralla, Daniel Green, Arpi Grigorian, John Groh, Chris Groppi, Yilun Guan, Jon E. Gudmundsson, Dongwon Han, Peter Hargrave, Masaya Hasegawa, Matthew Hasselfield, Makoto Hattori, Victor Haynes, Masashi Hazumi, Yizhou He, Erin Healy, Shawn W. Henderson, Carlos Hervias-Caimapo, Charles A. Hill, J. Colin Hill, Gene Hilton, Matt Hilton, Adam D. Hincks, Gary Hinshaw, Renée Hložek, Shirley Ho, Shuay-Pwu Patty Ho, Logan Howe, Zhiqi Huang, Johannes Hubmayr, Kevin Huffenberger, John P. Hughes, Anna Ijjas, Margaret Ikape, Kent Irwin, Andrew H. Jaffe, Bhuvnesh Jain, Oliver Jeong, Daisuke Kaneko, Ethan D. Karpel, Nobuhiko Katayama, Brian Keating, Sarah S. Kernasovskiy, Reijo Keskitalo, Theodore Kisner, Kenji Kiuchi, Jeff Klein, Kenda Knowles, Brian Koopman, Arthur Kosowsky, Nicoletta Krachmalnicoff, Stephen E. Kuenstner, Chao-Lin Kuo, Akito Kusaka, Jacob Lashner, Adrian Lee, Eunseong Lee, David Leon, Jason S.-Y. Leung, Antony Lewis, Yaqiong Li, Zack Li, Michele Limon, Eric Linder, Carlos Lopez-Caraballo, Thibaut Louis, Lindsay Lowry, Marius Lungu, Mathew Madhavacheril, Daisy Mak, Felipe Maldonado, Hamdi Mani, Ben Mates, Frederick Matsuda, Loïc Maurin, Phil Mauskopf, Andrew May, Nialh McCallum, Chris McKenney, Jeff McMahon, P. Daniel Meerburg, Joel Meyers, Amber Miller, Mark Mirmelstein, Kavilan Moodley, Moritz Munchmeyer, Charles Munson, Sigurd Naess, Federico Nati, Martin Navaroli, Laura Newburgh, Ho Nam Nguyen, Michael Niemack, Haruki Nishino, John Orlowski-Scherer, Lyman Page, Bruce Partridge, Julien Peloton, Francesca Perrotta, Lucio Piccirillo, Giampaolo Pisano, Davide Poletti, Roberto Puddu, Giuseppe Puglisi, Chris Raum, Christian L. Reichardt, Mathieu Remazeilles, Yoel Rephaeli, Dominik Riechers, Felipe Rojas, Anirban Roy, Sharon Sadeh, Yuki Sakurai, Maria Salatino, Mayuri Sathyanarayana Rao, Emmanuel Schaan, Marcel Schmittfull, Neelima Sehgal, Joseph Seibert, Uros Seljak, Blake Sherwin, Meir Shimon, Carlos Sierra, Jonathan Sievers, Precious Sikhosana, Maximiliano Silva-Feaver, Sara M. Simon, Adrian Sinclair, Praween Siritanasak, Kendrick Smith, Stephen R. Smith, David Spergel, Suzanne T. Staggs, George Stein, Jason R. Stevens, Radek Stompor, Aritoki Suzuki, Osamu Tajima, Satoru Takakura, Grant Teply, Daniel B. Thomas, Ben Thorne, Robert Thornton, Hy Trac, Calvin Tsai, Carole Tucker, Joel Ullom, Sunny Vagnozzi, Alexander van Engelen, Jeff Van Lanen, Daniel D. Van Winkle, Eve M. Vavagiakis, Clara Vergès, Michael Vissers, Kasey Wagoner, Samantha Walker, Jon Ward, Ben Westbrook, Nathan Whitehorn, Jason Williams, Joel Williams, Edward J. Wollack, Zhilei Xu, Byeonghee Yu, Cyndia Yu, Fernando Zago, Hezi Zhang, Ningfeng Zhu, and The Simons Observatory collaboration. 2019. The simons observatory: science goals and forecasts. *Journal of Cosmology and Astroparticle Physics* 2019(02):056.

Aghanim, N., Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J.-F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J.-M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini,

G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J.-M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y.-Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M.-A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J.-L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A.-S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. 2020. Planck2018 results: Vi. cosmological parameters. *Astronomy and Astrophysics* 641:A6.

Akaike, H. 1974. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* 19:716–723.

Alonso, David, Pedro G. Ferreira, and Mario G. Santos. 2014. Fast simulations for intensity mapping experiments. *Monthly Notices of the Royal Astronomical Society* 444(4):3183–3197.

Amiri, Mandana, Kevin Bandura, Tianyue Chen, Meiling Deng, Matt Dobbs, Mateus Fandino, Simon Foreman, Mark Halpern, Alex S. Hill, Gary Hinshaw, Carolin Höfer, Joseph Kania, T. L. Landecker, Joshua MacEachern, Kiyoshi Masui, Juan Mena-Parra, Nikola Milutinovic, Arash Mirhosseini, Laura Newburgh, Anna Ordog, Ue-Li Pen, Tristan Pinsonneault-Marotte, Ava Polzin, Alex Reda, Andre Renard, J. Richard Shaw, Seth R. Siegel, Saurabh Singh, Keith Vanderlinde, Haochen Wang, Donald V. Wiebe, and Dallas Wulf. 2023. Detection of cosmological 21 cm emission with the canadian hydrogen intensity mapping experiment. *The Astrophysical Journal* 947(1):16.

Anderson, C J, N J Luciw, Y C Li, C Y Kuo, J Yadav, K W Masui, T-C Chang, X Chen, N Oppermann, Y-W Liao, U-L Pen, D C Price, L Staveley-Smith, E R Switzer, P T Timbie, and L Wolz. 2018. Low-amplitude clustering in low-redshift 21-cm intensity maps cross-correlated with 2dF galaxy densities. *Monthly Notices of the Royal Astronomical Society* 476(3):3382–3392. `https://academic.oup.com/mnras/article-pdf/476/3/3382/25076525/sty346.pdf`.

Ansari, Réza, Evan J. Arena, Kevin Bandura, Philip Bull, Emanuele Castorina, Tzu-Ching Chang, Shi-Fan Chen, Liam Connor, Simon Foreman, Josef Frisch, Daniel Green, Matthew C. Johnson, Dionysios Karagiannis, Adrian Liu, Kiyoshi W. Masui, P. Daniel Meerburg, Moritz Münchmeyer, Laura B. Newburgh, Andrej Obuljen, Paul O'Connor, Hamsa Padmanabhan, J. Richard Shaw, Christopher Sheehy, Anže Slosar, Kendrick Smith, Paul Stankus, Albert Stebbins, Peter Timbie, Francisco Villaescusa-Navarro, Benjamin Wallisch, and Martin White. 2019. Inflation and early dark energy with a stage ii hydrogen intensity mapping experiment. `1810.09572`.

Bandura, Kevin, Graeme E. Addison, Mandana Amiri, J. Richard Bond, Duncan Campbell-Wilson, Liam Connor, Jean-François Cliche, Greg Davis, Meiling Deng, Nolan Denman, and et al. 2014. Canadian hydrogen intensity mapping experiment (chime) pathfinder. *Ground-based and Airborne Telescopes V*.

Battye, Richard A., Rod D. Davies, and Jochen Weller. 2004. Neutral hydrogen surveys for high-redshift galaxy clusters and protoclusters. *Monthly Notices of the Royal Astronomical Society* 355(4):1339–1347. `https://academic.oup.com/mnras/article-pdf/355/4/1339/6275970/355-4-1339.pdf`.

Beukman, Theunis S., Marianna V. Ivashina, Rob Maaskant, Petrie Meyer, and Carlo Bencivenni. 2014. A quadraxial feed for ultra-wide bandwidth quadruple-ridged flared horn antennas. In *The 8th european conference on antennas and propagation (eucap 2014)*, 3312–3316.

Bharadwaj, Somnath, and Shiv K. Sethi. 2001. HI fluctuations at large redshifts: I-visibility correlation. *Journal of Astrophysics and Astronomy* 22(4):293–307.

Bigot-Sazy, Marie-Anne, Yin-Zhe Ma, Richard A. Battye, Ian W. A. Browne, Tianyue Chen, Clive Dickinson, Stuart Harper, Bruno Maffei, Lucas C. Olivari, and Peter N. Wilkinson. 2016. HI intensity mapping with FAST. *ASP Conf. Ser.* 502:41. `1511.03006`.

Bobin, Jerome. 2017. LGMCA: Local-Generalized Morphological Component Analysis. Astrophysics Source Code Library, record ascl:1710.016. `1710.016`.

Bull, Philip, Pedro G. Ferreira, Prina Patel, and Mário G. Santos. 2015. Late-time cosmology with 21 cm intensity mapping experiments. *The Astrophysical Journal* 803(1):21.

Carucci, Isabella P, Melis O Irfan, and Jé rôme Bobin. 2020. Recovery of 21-cm intensity maps with sparse component separation. *Monthly Notices of the Royal Astronomical Society* 499(1): 304–319.

Castorina, Emanuele, and Francisco Villaescusa-Navarro. 2017. On the spatial distribution of neutral hydrogen in the universe: bias and shot-noise of the h i power spectrum. *Monthly Notices of the Royal Astronomical Society* 471(2):1788–1796.

Chang, Tzu-Ching, Ue-Li Pen, Jeffrey B. Peterson, and Patrick McDonald. 2008. Baryon acoustic oscillation intensity mapping of dark energy. *prl* 100(9):091303. `0709.3672`.

Chapman, Emma, Filipe B. Abdalla, J. Bobin, J.-L. Starck, Geraint Harker, Vibor Jelić , Panagiotis Labropoulos, Saleem Zaroubi, Michiel A. Brentjens, A. G. de Bruyn, and L. V. E. Koopmans. 2012a. The scale of the problem: recovering images of reionization with generalized morphological component analysis. *Monthly Notices of the Royal Astronomical Society* 429(1):165–176.

Chapman, Emma, Filipe B. Abdalla, Geraint Harker, Vibor Jelić , Panagiotis Labropoulos, Saleem Zaroubi, Michiel A. Brentjens, A. G. de Bruyn, and L. V. E. Koopmans. 2012b. Foreground removal usingscpfastica/scp: a showcase of LOFAR-EoR. *Monthly Notices of the Royal Astronomical Society* 423(3):2518–2532.

Chen, Xingang, P. Daniel Meerburg, and Moritz Münchmeyer. 2016. The future of primordial features with 21 cm tomography. *Journal of Cosmology and Astroparticle Physics* 2016(09): 023–023.

Chen, Xuelei. 2012. The Tianlai project: a 21cm cosmology experiment. *Int. J. Mod. Phys. Conf. Ser.* 12:256–263. `1212.6278`.

Choi, Steve, Matthew Hasselfield, Shuay-Pwu Ho, Brian Koopman, Marius Lungu, Maximilian Abitbol, Graeme Addison, Peter Ade, Simone Aiola, David Alonso, M. Amiri, Stefania Amodeo, Elio Angile, Jason Austermann, Taylor Baildon, Nick Battaglia, J. Beall, Rachel Bean, Daniel Becker, and Ningfeng Zhu. 2020. The atacama cosmology telescope: a measurement of the cosmic microwave background power spectra at 98 and 150 ghz. *Journal of Cosmology and Astroparticle Physics* 2020:045–045.

Coffey, D., Salvato, M., Merloni, A., Boller, Th., Nandra, K., Dwelly, T., Comparat, J., Schulze, A., Del Moro, A., and Schneider, D. P. 2019. Sdss-iv/spiders: A catalogue of x-ray selected agn properties - spectral properties and black hole mass estimates for spiders sdss dr14 type 1 agn. *AA* 625:A123.

Cole, Philippa S, and Joseph Silk. 2020. Small-scale primordial fluctuations in the 21 cm dark ages signal. *Monthly Notices of the Royal Astronomical Society* 501(2):2627–2634.

Collin, Robert, and Francis Zucker. 1969. *Antenna theory part i*. McGraw Hill.

Crighton, Neil H. M., Michael T. Murphy, J. Xavier Prochaska, Gá bor Worseck, Marc Rafelski, George D. Becker, Sara L. Ellison, Michele Fumagalli, Sebastian Lopez, Avery Meiksin, and John M. O'Meara. 2015. The neutral hydrogen cosmological mass density atiz/i= 5. *Monthly Notices of the Royal Astronomical Society* 452(1):217–234.

Cunnington, Steven, Melis O Irfan, Isabella P Carucci, Alkistis Pourtsidou, and Jérôme Bobin. 2021. 21-cm foregrounds and polarization leakage: cleaning and mitigation strategies. *Monthly Notices of the Royal Astronomical Society* 504(1):208–227. `https://academic.oup.com/mnras/article-pdf/504/1/208/37091350/stab856.pdf`.

Cunnington, Steven, Yichao Li, Mario G Santos, Jingying Wang, Isabella P Carucci, Melis O Irfan, Alkistis Pourtsidou, Marta Spinelli, Laura Wolz, Paula S Soares, Chris Blake, Philip Bull, Brandon Engelbrecht, José Fonseca, Keith Grainge, and Yin-Zhe Ma. 2022. HI intensity mapping with MeerKAT: power spectrum detection in cross-correlation with WiggleZ galaxies. *Monthly Notices of the Royal Astronomical Society* 518(4):6262–6272. `https://academic.oup.com/mnras/article-pdf/518/4/6262/48302259/stac3060.pdf`.

Cunnington, Steven, Laura Wolz, Philip Bull, Isabella P Carucci, Keith Grainge, Melis O Irfan, Yichao Li, Alkistis Pourtsidou, Mario G Santos, Marta Spinelli, and Jingying Wang. 2023. The foreground transfer function for Hi intensity mapping signal reconstruction: MeerKLASS and precision cosmology applications. *Monthly Notices of the Royal Astronomical Society* 523(2): 2453–2477. `https://academic.oup.com/mnras/article-pdf/523/2/2453/50519870/stad1567.pdf`.

Datta, A., J. D. Bowman, and C. L. Carilli. 2010. BRIGHT SOURCE SUBTRACTION REQUIREMENTS FOR REDSHIFTED 21 cm MEASUREMENTS. *The Astrophysical Journal* 724(1): 526–538.

Datta, Kanan K., T. Roy Choudhury, and Somnath Bharadwaj. 2007. The multifrequency angular power spectrum of the epoch of reionization 21-cm signal. *Monthly Notices of the Royal Astronomical Society* 378(1):119–128. `https://academic.oup.com/mnras/article-pdf/378/1/119/3962809/mnras0378-0119.pdf`.

DeBoer, David R., and HERA. 2015. Hydrogen Epoch of Reionization Array (HERA). In *American astronomical society meeting abstracts #225*, vol. 225 of *American Astronomical Society Meeting Abstracts*, 328.03.

DeBoer, David R., Aaron R. Parsons, James E. Aguirre, Paul Alexander, Zaki S. Ali, Adam P. Beardsley, Gianni Bernardi, Judd D. Bowman, Richard F. Bradley, Chris L. Carilli, Carina Cheng, Eloy de Lera Acedo, Joshua S. Dillon, Aaron Ewall-Wice, Gcobisa Fadana, Nicolas Fagnoni, Randall Fritz, Steve R. Furlanetto, Brian Glendenning, Bradley Greig, Jasper Grobbelaar, Bryna J. Hazelton, Jacqueline N. Hewitt, Jack Hickish, Daniel C. Jacobs, Austin Julius, MacCalvin Kariseb, Saul A. Kohn, Telalo Lekalake, Adrian Liu, Anita Loots, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary Martinot, Nathan Mathison, Eunice Matsetela, Andrei Mesinger, Miguel F. Morales, Abraham R. Neben, Nipanjana Patra, Samantha Pieterse, Jonathan C. Pober, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Raddwine Sell, Craig Smith, Angelo Syce, Max Tegmark, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. 2017. Hydrogen epoch of reionization array (hera). *Publications of the Astronomical Society of the Pacific* 129(974):045001.

Delabrouille, J., Betoule, M., Melin, J.-B., Miville-Deschênes, M.-A., Gonzalez-Nuevo, J., Le Jeune, M., Castex, G., de Zotti, G., Basak, S., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Bernard, J.-P., Bouchet, F. R., Clements, D. L., da Silva, A., Dickinson, C., Dodu, F., Dolag, K., Elsner, F., Fauvet, L., Faÿ, G., Giardino, G., Leach, S., Lesgourgues, J., Liguori, M., Macías-Pérez, J. F., Massardi, M., Matarrese, S., Mazzotta, P., Montier, L., Mottet, S., Paladini, R., Partridge, B., Piffaretti, R., Prezeau, G., Prunet, S., Ricciardi, S., Roman, M., Schaefer, B., and Toffolatti, L. 2013. The pre-launch planck sky model: a model of sky emission at submillimetre to centimetre wavelengths. *A&A* 553:A96.

Delabrouille, J., Cardoso, J.-F., Le Jeune, M., Betoule, M., Fay, G., and Guilloux, F. 2009. A full sky, low foreground, high resolution cmb map from wmap. *A&A* 493(3):835–857.

DESI Collaboration, Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E. Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, Otger Ballester, Charles Baltay, Lucas Beaufore, Chris Bebek, Timothy C. Beers, Eric F. Bell, José Luis Bernal, Robert Besuner, Florian Beutler, Chris Blake, Hannes Bleuler, Michael Blomqvist, Robert Blum, Adam S. Bolton, Cesar Briceno, David Brooks, Joel R. Brownstein, Elizabeth Buckley-Geer, Angela Burden, Etienne Burtin, Nicolas G. Busca, Robert N. Cahn, Yan-Chuan Cai, Laia Cardiel-Sas, Raymond G. Carlberg, Pierre-Henri Carton, Ricard Casas, Francisco J. Castander, Jorge L. Cervantes-Cota, Todd M. Claybaugh, Madeline Close, Carl T. Coker, Shaun Cole, Johan Comparat, Andrew P. Cooper, M. C. Cousinou, Martin Crocce, Jean-Gabriel Cuby, Daniel P. Cunningham, Tamara M. Davis, Kyle S. Dawson, Axel de la Macorra, Juan De Vicente, Timothée Delubac, Mark Derwent, Arjun Dey, Govinda Dhungana, Zhejie Ding, Peter Doel, Yutong T. Duan, Anne Ealet, Jerry Edelstein, Sarah Eftekharzadeh, Daniel J. Eisenstein, Ann Elliott, Stéphanie Escoffier, Matthew Evatt, Parker Fagrelius, Xiaohui Fan, Kevin Fanning, Arya Farahi, Jay Farihi, Ginevra Favole,

Yu Feng, Enrique Fernandez, Joseph R. Findlay, Douglas P. Finkbeiner, Michael J. Fitzpatrick, Brenna Flaugher, Samuel Flender, Andreu Font-Ribera, Jaime E. Forero-Romero, Pablo Fosalba, Carlos S. Frenk, Michele Fumagalli, Boris T. Gaensicke, Giuseppe Gallo, Juan Garcia-Bellido, Enrique Gaztanaga, Nicola Pietro Gentile Fusillo, Terry Gerard, Irena Gershkovich, Tommaso Giannantonio, Denis Gillet, Guillermo Gonzalez de Rivera, Violeta Gonzalez-Perez, Shelby Gott, Or Graur, Gaston Gutierrez, Julien Guy, Salman Habib, Henry Heetderks, Ian Heetderks, Katrin Heitmann, Wojciech A. Hellwing, David A. Herrera, Shirley Ho, Stephen Holland, Klaus Honscheid, Eric Huff, Timothy A. Hutchinson, Dragan Huterer, Ho Seong Hwang, Joseph Maria Illa Laguna, Yuzo Ishikawa, Dianna Jacobs, Niall Jeffrey, Patrick Jelinsky, Elise Jennings, Linhua Jiang, Jorge Jimenez, Jennifer Johnson, Richard Joyce, Eric Jullo, Stéphanie Juneau, Sami Kama, Armin Karcher, Sonia Karkar, Robert Kehoe, Noble Kennamer, Stephen Kent, Martin Kilbinger, Alex G. Kim, David Kirkby, Theodore Kisner, Ellie Kitanidis, Jean-Paul Kneib, Sergey Koposov, Eve Kovacs, Kazuya Koyama, Anthony Kremin, Richard Kron, Luzius Kronig, Andrea Kueter-Young, Cedric G. Lacey, Robin Lafever, Ofer Lahav, Andrew Lambert, Michael Lampton, Martin Landriau, Dustin Lang, Tod R. Lauer, Jean-Marc Le Goff, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Su-Jeong Lee, Daniela Leitner, Michael Lesser, Michael E. Levi, Benjamin L'Huillier, Baojiu Li, Ming Liang, Huan Lin, Eric Linder, Sarah R. Loebman, Zarija Lukić, Jun Ma, Niall MacCrann, Christophe Magneville, Laleh Makarem, Marc Manera, Christopher J. Manser, Robert Marshall, Paul Martini, Richard Massey, Thomas Matheson, Jeremy McCauley, Patrick McDonald, Ian D. McGreer, Aaron Meisner, Nigel Metcalfe, Timothy N. Miller, Ramon Miquel, John Moustakas, Adam Myers, Milind Naik, Jeffrey A. Newman, Robert C. Nichol, Andrina Nicola, Luiz Nicolati da Costa, Jundan Nie, Gustavo Niz, Peder Norberg, Brian Nord, Dara Norman, Peter Nugent, Thomas O'Brien, Minji Oh, Knut A. G. Olsen, Cristobal Padilla, Hamsa Padmanabhan, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Antonella Palmese, Daniel Pappalardo, Isabelle Pâris, Changbom Park, Anna Patej, John A. Peacock, Hiranya V. Peiris, Xiyan Peng, Will J. Percival, Sandrine Perruchot, Matthew M. Pieri, Richard Pogge, Jennifer E. Pollack, Claire Poppett, Francisco Prada, Abhishek Prakash, Ronald G. Probst, David Rabinowitz, Anand Raichoor, Chang Hee Ree, Alexandre Refregier, Xavier Regal, Beth Reid, Kevin Reil, Mehdi Rezaie, Constance M. Rockosi, Natalie Roe, Samuel Ronayette, Aaron Roodman, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, Eduardo Rozo, Vanina Ruhlmann-Kleider, Eli S. Rykoff, Cristiano Sabiu, Lado Samushia, Eusebio Sanchez, Javier Sanchez, David J. Schlegel, Michael Schneider, Michael Schubnell, Aurélia Secroun, Uros Seljak, Hee-Jong Seo, Santiago Serrano, Arman Shafieloo, Huanyuan Shan, Ray Sharples, Michael J. Sholl, William V. Shourt, Joseph H. Silber, David R. Silva, Martin M. Sirk, Anze Slosar, Alex Smith, George F. Smoot, Debopam Som, Yong-Seon Song, David Sprayberry, Ryan Staten, Andy Stefanik, Gregory Tarle, Suk Sien Tie, Jeremy L. Tinker, Rita Tojeiro, Francisco Valdes, Octavio Valenzuela, Monica Valluri, Mariana

Vargas-Magana, Licia Verde, Alistair R. Walker, Jiali Wang, Yuting Wang, Benjamin A. Weaver, Curtis Weaverdyck, Risa H. Wechsler, David H. Weinberg, Martin White, Qian Yang, Christophe Yeche, Tianmeng Zhang, Gong-Bo Zhao, Yi Zheng, Xu Zhou, Zhimin Zhou, Yaling Zhu, Hu Zou, and Ying Zu. 2016. The desi experiment part i: Science,targeting, and survey design. `1611.00036`.

Dodelson, Scott. 2003. *Modern Cosmology*. 1st ed. Academic Press.

Dodelson, Scott, and Fabian Schmidt. 2020. *Modern Cosmology*. 2nd ed. Elsevier Science.

Ewall-Wice, Aaron, Richard Bradley, David Deboer, Jacqueline Hewitt, Aaron Parsons, James Aguirre, Zaki S. Ali, Judd Bowman, Carina Cheng, Abraham R. Neben, Nipanjana Patra, Nithyanandan Thyagarajan, Mariet Venter, Eloy de Lera Acedo, Joshua S. Dillon, Roger Dickenson, Phillip Doolittle, Dennis Egan, Mike Hedrick, Patricia Klima, Saul Kohn, Patrick Schaffner, John Shelton, Benjamin Saliwanchik, H. A. Taylor, Rusty Taylor, Max Tegmark, and Butch Wirt. 2016. The hydrogen epoch of reionization array dish. ii. characterization of spectral structure with electromagnetic simulations and its science implications. *The Astrophysical Journal* 831(2):196.

Ewall-Wice, Aaron, Nicholas Kern, Joshua S Dillon, Adrian Liu, Aaron Parsons, Saurabh Singh, Adam Lanman, Paul La Plante, Nicolas Fagnoni, Eloy de Lera Acedo, David R DeBoer, Chuneeta Nunhokee, Philip Bull, Tzu-Ching Chang, T Joseph W Lazio, James Aguirre, and Sean Weinberg. 2020. ttDAYENU:/tt a simple filter of smooth foregrounds for intensity mapping power spectra. *Monthly Notices of the Royal Astronomical Society* 500(4):5195–5213.

Fagnoni, Nicolas, Eloy de Lera Acedo, David R DeBoer, Zara Abdurashidova, James E Aguirre, Paul Alexander, Zaki S Ali, Yanga Balfour, Adam P Beardsley, Gianni Bernardi, Tashalee S Billings, Judd D Bowman, Richard F Bradley, Phil Bull, Jacob Burba, Chris L Carilli, Carina Cheng, Matt Dexter, Joshua S Dillon, Aaron Ewall-Wice, Randall Fritz, Steve R Furlanetto, Kingsley Gale-Sides, Brian Glendenning, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J Hazelton, Jacqueline N Hewitt, Jack Hickish, Daniel C Jacobs, Alec Josaitis, Austin Julius, Nicholas S Kern, Joshua Kerrigan, Honggeun Kim, Piyanat Kittiwisit, Saul A Kohn, Matthew Kolopanis, Adam Lanman, Paul La Plante, Telalo Lekalake, Adrian Liu, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary E Martinot, Eunice Matsetela, Juan Mena Parra, Andrei Mesinger, Mathakane Molewa, Miguel F Morales, Tshegofalang Mosiane, Abraham R Neben, Bojan Nikolic, Aaron R Parsons, Nipanjana Patra, Samantha Pieterse, Jonathan C Pober, Nima Razavi-Ghods, James Robnett, Kathryn Rosie, Peter Sims, Craig Smith, Angelo Syce, Nithyanandan Thyagarajan, Peter K G Williams, and Haoxuan Zheng. 2020. Understanding the hera phase i receiver system with simulations and its impact on the detectability of the eor delay power spectrum. *Monthly Notices of the Royal Astronomical Society* 500(1):1232–1242.

Fagnoni, Nicolas, Eloy de Lera Acedo, Nick Drought, David R. DeBoer, Daniel Riley, Nima Razavi-Ghods, Steven Carey, and Aaron R. Parsons. 2021. Design of the new wideband vivaldi feed for the hera radio-telescope phase ii. *IEEE Transactions on Antennas and Propagation* 69(12): 8143–8157.

and Flaugher, Brenna. 2005. The dark energy survey. *International Journal of Modern Physics A* 20(14):3121–3123. `https://doi.org/10.1142/S0217751X05025917`.

Ghojogh, Benyamin, Fakhri Karray, and Mark Crowley. 2023. Eigenvalue and generalized eigenvalue problems: Tutorial. `1903.11240`.

Gupta, I.J., K.P. Ericksen, and W.D. Burnside. 1990. A method to design blended rolled edges for compact range reflectors. *IEEE Transactions on Antennas and Propagation* 38(6):853–861.

van Haarlem, M. P., M. W. Wise, A. W. Gunst, G. Heald, J. P. McKean, J. W. T. Hessels, A. G. de Bruyn, R. Nijboer, J. Swinbank, R. Fallows, M. Brentjens, A. Nelles, R. Beck, H. Falcke, R. Fender, J. Hörandel, L. V. E. Koopmans, G. Mann, G. Miley, H. Röttgering, B. W. Stappers, R. A. M. J. Wijers, S. Zaroubi, M. van den Akker, A. Alexov, J. Anderson, K. Anderson, A. van Ardenne, M. Arts, A. Asgekar, I. M. Avruch, F. Batejat, L. Bähren, M. E. Bell, M. R. Bell, I. van Bemmel, P. Bennema, M. J. Bentum, G. Bernardi, P. Best, L. Bîrzan, A. Bonafede, A.-J. Boonstra, R. Braun, J. Bregman, F. Breitling, R. H. van de Brink, J. Broderick, P. C. Broekema, W. N. Brouw, M. Brüggen, H. R. Butcher, W. van Cappellen, B. Ciardi, T. Coenen, J. Conway, A. Coolen, A. Corstanje, S. Damstra, O. Davies, A. T. Deller, R.-J. Dettmar, G. van Diepen, K. Dijkstra, P. Donker, A. Doorduin, J. Dromer, M. Drost, A. van Duin, J. Eislöffel, J. van Enst, C. Ferrari, W. Frieswijk, H. Gankema, M. A. Garrett, F. de Gasperin, M. Gerbers, E. de Geus, J.-M. Grießmeier, T. Grit, P. Gruppen, J. P. Hamaker, T. Hassall, M. Hoeft, H. A. Holties, A. Horneffer, A. van der Horst, A. van Houwelingen, A. Huijgen, M. Iacobelli, H. Intema, N. Jackson, V. Jelic, A. de Jong, E. Juette, D. Kant, A. Karastergiou, A. Koers, H. Kollen, V. I. Kondratiev, E. Kooistra, Y. Koopman, A. Koster, M. Kuniyoshi, M. Kramer, G. Kuper, P. Lambropoulos, C. Law, J. van Leeuwen, J. Lemaitre, M. Loose, P. Maat, G. Macario, S. Markoff, J. Masters, R. A. McFadden, D. McKay-Bukowski, H. Meijering, H. Meulman, M. Mevius, E. Middelberg, R. Millenaar, J. C. A. Miller-Jones, R. N. Mohan, J. D. Mol, J. Morawietz, R. Morganti, D. D. Mulcahy, E. Mulder, H. Munk, L. Nieuwenhuis, R. van Nieuwpoort, J. E. Noordam, M. Norden, A. Noutsos, A. R. Offringa, H. Olofsson, A. Omar, E. Orrú, R. Overeem, H. Paas, M. Pandey-Pommier, V. N. Pandey, R. Pizzo, A. Polatidis, D. Rafferty, S. Rawlings, W. Reich, J.-P. de Reijer, J. Reitsma, G. A. Renting, P. Riemers, E. Rol, J. W. Romein, J. Roosjen, M. Ruiter, A. Scaife, K. van der Schaaf, B. Scheers, P. Schellart, A. Schoenmakers, G. Schoonderbeek, M. Serylak, A. Shulevski, J. Sluman, O. Smirnov, C. Sobey, H. Spreeuw, M. Steinmetz, C. G. M. Sterks, H.-J. Stiepel, K. Stuurwold,

M. Tagger, Y. Tang, C. Tasse, I. Thomas, S. Thoudam, M. C. Toribio, B. van der Tol, O. Usov, M. van Veelen, A.-J. van der Veen, S. ter Veen, J. P. W. Verbiest, R. Vermeulen, N. Vermaas, C. Vocks, C. Vogt, M. de Vos, E. van der Wal, R. van Weeren, H. Weggemans, P. Weltevrede, S. White, S. J. Wijnholds, T. Wilhelmsson, O. Wucknitz, S. Yatawatta, P. Zarka, A. Zensus, and J. van Zwieten. 2013. Lofar: The low-frequency array. *Astronomy and Astrophysics* 556:A2.

Hansen, N., S.D. Muller, and P. Koumoutsakos. 2003. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation* 11(1):1–18.

Hothi, Ian, Emma Chapman, Jonathan R Pritchard, F G Mertens, L V E Koopmans, B Ciardi, B K Gehlot, R Ghara, A Ghosh, S K Giri, I T Iliev, V Jelić , and S Zaroubi. 2020. Comparing foreground removal techniques for recovery of the lofar-eor 21cm power spectrum. *Monthly Notices of the Royal Astronomical Society* 500(2):2264–2277.

Hu, Jian-Ping, and Fa-Yin Wang. 2023. Hubble tension: The evidence of new physics. *Universe* 9: 94.

Hu, Wenkai, Xin Wang, Fengquan Wu, Yougang Wang, Pengjie Zhang, and Xuelei Chen. 2020. Forecast for FAST: from galaxies survey to intensity mapping. *Monthly Notices of the Royal Astronomical Society* 493(4):5854–5870. `https://academic.oup.com/mnras/article-pdf/493/4/5854/32978406/staa650.pdf`.

Hu, Yue, and A Lazarian. 2023. Mapping the galactic magnetic field orientation and strength in three dimensions. *Monthly Notices of the Royal Astronomical Society* 524(2):2379–2394.

Hui, Howard, P. A. R. Ade, Zeeshan Ahmed, Randol Aikin, Kate D. Alexander, Denis Barkats, Steve J. Benton, Colin A. Bischoff, James J. Bock, Rachel Bowens-Rubin, Justus A. Brevik, Immanuel Buder, Eric Bullock, Victor Buza, Jake Connors, James Cornelison, Brendan Crill, Michael Crumrine, Marion Dierickx, Lionel Duband, Cora Dvorkin, Jeff P. Filippini, Stefan Fliescher, James A. Grayson, Grantland Hall, Mark Halpern, Sam A. Harrison, Sergi Hildebrandt, Gene C. Hilton, Kent D. Irwin, Jae Hwan Kang, Kirit S. Karkare, Ethan Karpel, Jonathan Kaufman, Brian G. Keating, Sinan Kefeli, Sarah A. Kernasovskiy, J. M. Kovac, Chao-Lin Kuo, Nicole Larson, King Lau, Erik M. Leitch, Martin Lueker, K. G. Megerian, Lorenzo Moncelsi, Toshiya Namikawa, C. B. Netterfield, H. T. Nguyen, Roger O'brient, R. Walt Ogburn IV, Stephen Palladino, Clement Pryke, Benjamin Racine, Steffen Richter, Alessandro Schillaci, Robert Schwarz, Chris D. Sheehy, Ahmed Soliman, Tyler St Germaine, Zachary K. Staniszewski, Bryan Steinbach, Rashmi Sudiwala, Grant P. Teply, Keith L. Thompson, James E. Tolan, Carole Tucker, Anthony D. Turner, Caterina Umiltà, Abigail G. Vieregg, Albert Wandui, Alexis C. Weber, Don Wiebe, Justin Willmert, Chin Lin Wong, Wai Ling K. Wu, Eric Yang, Ki Won Yoon, and Cheng Zheng. 2018. Bicep array: a multi-frequency degree-scale cmb polarimeter. In *Millimeter, submillimeter, and far-infrared detectors and instrumentation for astronomy ix*, ed. Jonas Zmuidzinas and Jian-Rong Gao. SPIE.

Jaffe, Tess, Andre Waelkens, M. Reinecke, F. S. Kitaura, and T. A. Ensslin. 2012. Hammurabi: Simulating polarized Galactic synchrotron emission. Astrophysics Source Code Library, record ascl:1201.014. `1201.014`.

Josaitis, Alec T, Aaron Ewall-Wice, Nicolas Fagnoni, and Eloy de Lera Acedo. 2022. Array element coupling in radio interferometry I: a semi-analytic approach. *Monthly Notices of the Royal Astronomical Society* 514(2):1804–1827. `https://academic.oup.com/mnras/article-pdf/514/2/1804/44055382/stac916.pdf`.

Kern, Nicholas S., Aaron R. Parsons, Joshua S. Dillon, Adam E. Lanman, Nicolas Fagnoni, and Eloy de Lera Acedo. 2019. Mitigating internal instrument coupling for 21 cm cosmology. i. temporal and spectral modeling in simulations. *The Astrophysical Journal* 884(2):105.

Kollmeier, Juna, S. F. Anderson, G. A. Blanc, M. R. Blanton, K. R. Covey, J. Crane, N. Drory, P. M. Frinchaboy, C. S. Froning, J. A. Johnson, J. P. Kneib, K. Kreckel, A. Merloni, E. W. Pellegrini, R. W. Pogge, S. V. Ramirez, H. W. Rix, C. Sayres, José Sánchez-Gallego, Yue Shen, A. Tkachenko, J. R. Trump, S. E. Tuttle, A. Weijmans, G. Zasowski, B. Barbuy, R. L. Beaton, M. Bergemann, J. J. Bochanski, W. N. Brandt, A. R. Casey, B. A. Cherinka, M. Eracleous, X. Fan, R. A. García, P. J. Green, S. Hekker, R. R. Lane, P. Longa-Peña, S. Mathur, A. Meza, I. Minchev, A. D. Myers, D. L. Nidever, C. Nitschelm, J. E. O'Connell, A. M. Price-Whelan, M. J. Raddick, G. Rossi, R. Sankrit, J. D. Simon, A. M. Stutz, Y. S. Ting, B. Trakhtenbrot, B. A. Weaver, C. N. A. Willmer, and D. H. Weinberg. 2019. SDSS-V Pioneering Panoptic Spectroscopy. In *Bulletin of the american astronomical society*, vol. 51, 274.

Kwak, Juhun, John Podczerwinski, Peter Timbie, Réza Ansari, John Marriner, Albert Stebbins, Fengquan Wu, Haotian Cao, Xuelei Chen, Kai He, Jixia Li, Shijie Sun, and Jiacong Zhu. 2024. The effects of the local environment on a compact radio interferometer i: Cross-coupling in the tianlai dish pathfinder array. *Journal of Astronomical Instrumentation* 13(01):2450002. `https://doi.org/10.1142/S2251171724500028`.

Lamarre, Jean-Michel, Jean-Loup Puget, Peter Ade, François Bouchet, Guy Guyot, A. Lange, F. Pajot, A. Arondel, K. Benabed, Jean-Luc Beney, Alain Benoît, J.-Ph Bernard, Rishabh Bhatia, Y. Blanc, J. Bock, E. Bréelle, T. Bradshaw, Philippe Camus, A. Catalano, and D. Yvon. 2010. Planck pre-launch status: The hfi instrument, from specification to actual performance. *Astronomy and Astrophysics* 520.

Leistedt, B., J. D. McEwen, P. Vandergheynst, and Y. Wiaux. 2013. S2let: A code to perform fast wavelet analysis on the sphere. *Astronomy and Astrophysics* 558:A128.

Levi, Michael, Chris Bebek, Timothy Beers, Robert Blum, Robert Cahn, Daniel Eisenstein, Brenna Flaugher, Klaus Honscheid, Richard Kron, Ofer Lahav, Patrick McDonald, Natalie Roe, David Schlegel, and representing the DESI collaboration. 2013. The desi experiment, a whitepaper for snowmass 2013. `1308.0847`.

Liu, Adrian, and J. Richard Shaw. 2020. Data analysis for precision 21 cm cosmology. *Publications of the Astronomical Society of the Pacific* 132(1012):062001.

Liu, Adrian, Max Tegmark, Scott Morrison, Andrew Lutomirski, and Matias Zaldarriaga. 2010. Precision calibration of radio interferometers using redundant baselines. *Monthly Notices of the Royal Astronomical Society* 408(2):1029–1050. `https://academic.oup.com/mnras/article-pdf/408/2/1029/18440559/mnras0408-1029.pdf`.

Liu, Adrian, Yunfan Zhang, and Aaron R. Parsons. 2016. Spherical harmonic analyses of intensity mapping power spectra. *The Astrophysical Journal* 833(2):242.

Ludwig, A. 1973. The definition of cross polarization. *IEEE Transactions on Antennas and Propagation* 21(1):116–119.

Lui, Hoi-Shun, Hon Tat Hui, and Mook Seng Leong. 2009. A note on the mutual-coupling problems in transmitting and receiving antenna arrays. *IEEE Antennas and Propagation Magazine* 51(5):171–176.

Madau, Piero, Avery Meiksin, and Martin J. Rees. 1997. 21 centimeter tomography of the intergalactic medium at high redshift. *The Astrophysical Journal* 475(2):429–444.

Marins, Alessandro, Filipe B. Abdalla, Karin S. F. Fornazier, Elcio Abdalla, Luiz H. F. Assis, Mathieu Remazeilles, Carlos Alexandre Wuensche, Luciano Barosi, Amilcar R. Queiroz, Thyrso Villela, Bin Wang, Chang Feng, Ricardo Landim, Vincenzo Liccardo, Camila P. Novaes, Larissa Santos, Marcelo V. dos Santos, and Jiajun Zhang. 2022. Foreground removal and 21 cm signal estimates: comparing different blind methods for the bingo telescope. `2209.11701`.

Marinucci, D., D. Pietrobon, A. Balbi, P. Baldi, P. Cabella, G. Kerkyacharian, P. Natoli, D. Picard, and N. Vittorio. 2007. Spherical needlets for cosmic microwave background data analysis. *Monthly Notices of the Royal Astronomical Society* 383(2):539–545.

Martin, Jérôme, Christophe Ringeval, Roberto Trotta, and Vincent Vennin. 2014. The best inflationary models after planck. *Journal of Cosmology and Astroparticle Physics* 2014(03): 039–039.

Mertens, F G, A Ghosh, and L V E Koopmans. 2018. Statistical 21-cm signal separation via Gaussian Process Regression analysis. *Monthly Notices of the Royal Astronomical Society* 478(3): 3640–3652. `https://academic.oup.com/mnras/article-pdf/478/3/3640/25072259/sty1207.pdf`.

Meys, R.P. 1978. A wave approach to the noise properties of linear microwave devices. *IEEE Transactions on Microwave Theory and Techniques* 26(1):34–37.

Mirocha, Jordan, Julian B Muñoz, Steven R Furlanetto, Adrian Liu, and Andrei Mesinger. 2022. A galaxy-free phenomenological model for the 21-cm power spectrum during reionization. *Monthly Notices of the Royal Astronomical Society* 514(2):2010–2030.

Moore, David F., James E. Aguirre, Saul A. Kohn, Aaron R. Parsons, Zaki S. Ali, Richard F. Bradley, Chris L. Carilli, David R. DeBoer, Matthew R. Dexter, Nicole E. Gugliucci, Daniel C. Jacobs, Pat Klima, Adrian Liu, David H. E. MacMahon, Jason R. Manley, Jonathan C. Pober, Irina I. Stefan, and William P. Walbrugh. 2017. Limits on polarized leakage for the PAPER epoch of reionization measurements at 126 and 164 MHz. *The Astrophysical Journal* 836(2):154.

Moore, David F., James E. Aguirre, Aaron R. Parsons, Daniel C. Jacobs, and Jonathan C. Pober. 2013. THE EFFECTS OF POLARIZED FOREGROUNDS ON 21 cm EPOCH OF REIONIZATION POWER SPECTRUM MEASUREMENTS. *The Astrophysical Journal* 769(2):154.

Morales, M. F. 2005. Design princples of the Mileura Wide-field Array Low Frequency Demonstrator (MWA-LFD). In *From clark lake to the long wavelength array: Bill erickson's radio science*, ed. N. Kassim, M. Perez, W. Junor, and P. Henning, vol. 345 of *Astronomical Society of the Pacific Conference Series*, 452.

Murtagh, Fionn, Jean-Luc Starck, and Mario Bertero. 2011. *Starlet transform in astronomical data processing: Application to source detection and image deconvolution*, 1489–1531. Springer.

Muñoz, Julian B., Yacine Ali-Haïmoud, and Marc Kamionkowski. 2015. Primordial non-gaussianity from the bispectrum of 21-cm fluctuations in the dark ages. *Physical Review D* 92(8).

Narcowich, Francis J., Pencho Petrushev, and Joseph D. Ward. 2006. Localized tight frames on spheres. *SIAM J. Math. Anal.* 38:574–594.

Norris, Ray P., Kaustuv Basu, Michael Brown, Ettore Carretti, Anna D. Kapinska, Isabella Prandoni, Lawrence Rudnick, and Nick Seymour. 2014. The ska mid-frequency all-sky continuum survey: Discovering the unexpected and transforming radio-astronomy. `1412.6076`.

Olivari, L. C., M. Remazeilles, and C. Dickinson. 2015. Extracting h i cosmological signal with generalized needlet internal linear combination. *Monthly Notices of the Royal Astronomical Society* 456(3):2749–2765.

Oppermann, N., H. Junklewitz, G. Robbers, M. R. Bell, T. A. Enßlin, A. Bonafede, R. Braun, J. C. Brown, T. E. Clarke, I. J. Feain, B. M. Gaensler, A. Hammond, L. Harvey-Smith, G. Heald, M. Johnston-Hollitt, U. Klein, P. P. Kronberg, S. A. Mao, N. M. McClure-Griffiths, S. P. O'Sullivan, L. Pratley, T. Robishaw, S. Roy, D. H. F. M. Schnitzeler, C. Sotomayor-Beltran, J. Stevens, J. M. Stil, C. Sunstrum, A. Tanna, A. R. Taylor, and C. L. Van Eck. 2012. An improved map of the galactic faraday sky. *Astronomy and Astrophysics* 542:A93.

Parsons, Aaron, Jonathan Pober, James Aguirre, Christopher Carilli, Daniel Jacobs, and David Moore. 2012a. A per-baseline, delay-spectrum technique for accessing the 21 cm cosmic reionization signature. *The Astrophysical Journal* 756(2):165.

Parsons, Aaron R., Donald C. Backer, Griffin S. Foster, Melvyn C. H. Wright, Richard F. Bradley, Nicole E. Gugliucci, Chaitali R. Parashare, Erin E. Benoit, James E. Aguirre, Daniel C. Jacobs, and et al. 2010. The precision array for probing the epoch of re-ionization: Eight station results. *The Astronomical Journal* 139(4):1468–1480.

Parsons, Aaron R., Adrian Liu, James E. Aguirre, Zaki S. Ali, Richard F. Bradley, Chris L. Carilli, David R. DeBoer, Matthew R. Dexter, Nicole E. Gugliucci, Daniel C. Jacobs, and et al. 2014. New limits on 21 cm epoch of reionization from paper-32 consistent with an x-ray heated intergalactic medium atz= 7.7. *The Astrophysical Journal* 788(2):106.

Parsons, Aaron R., Jonathan C. Pober, James E. Aguirre, Christopher L. Carilli, Daniel C. Jacobs, and David F. Moore. 2012b. A per-baseline, delay-spectrum technique for accessing the 21 cm cosmic reionization signature. *The Astrophysical Journal* 756(2):165.

Paul, Sourabh, Mario G. Santos, Zhaoting Chen, and Laura Wolz. 2023. A first detection of neutral hydrogen intensity mapping on mpc scales at $z \approx 0.32$ and $z \approx 0.44$. `2301.11943`.

Penzias, A. A., and R. W. Wilson. 1965. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *Apj* 142:419–421.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de

Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. 2020a. Planck 2018 results. VI. Cosmological parameters. *AAP* 641:A6. `1807.06209`.

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

N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, M. A. Miville-Deschênes, D. Molinari, A. Moneti, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, A. Zacchei, and A. Zonca. 2020b. Planck 2018 results. V. CMB power spectra and likelihoods. *AAP* 641:A5. `1907.12875`.

Price, Danny C. *Spectrometers and polyphase filterbanks in radio astronomy*, chap. Chapter 7, 159–179. `https://www.worldscientific.com/doi/pdf/10.1142/9789811203770_0007`.

Riess, Adam G., Wenlong Yuan, Lucas M. Macri, Dan Scolnic, Dillon Brout, Stefano Casertano, David O. Jones, Yukei Murakami, Gagandeep S. Anand, Louise Breuval, Thomas G. Brink, Alexei V. Filippenko, Samantha Hoffmann, Saurabh W. Jha, W. D'arcy Kenworthy, John Mackenty, Benjamin E. Stahl, and WeiKang Zheng. 2022. A comprehensive measurement of the local value of the hubble constant with 1 km s1 mpc1 uncertainty from the hubble space telescope and the sh0es team. *The Astrophysical Journal Letters* 934(1):L7.

Robertson, B. E., S. Tacchella, B. D. Johnson, K. Hainline, L. Whitler, D. J. Eisenstein, R. Endsley, M. Rieke, D. P. Stark, S. Alberts, A. Dressler, E. Egami, R. Hausen, G. Rieke, I. Shivaei, C. C. Williams, C. N. A. Willmer, S. Arribas, N. Bonaventura, A. Bunker, A. J. Cameron, S. Carniani, S. Charlot, J. Chevallard, M. Curti, E. Curtis-Lake, F. D'Eugenio, P. Jakobsen, T. J. Looser, N. Lützgendorf, R. Maiolino, M. V. Maseda, T. Rawle, H.-W. Rix, R. Smit, H. Übler, C. Willott, J. Witstok, S. Baum, R. Bhatawdekar, K. Boyett, Z. Chen, A. de Graaff, M. Florian, J. M. Helton, R. E. Hviding, Z. Ji, N. Kumari, J. Lyu, E. Nelson, L. Sandles, A. Saxena, K. A. Suess, F. Sun, M. Topping, and I. E. B. Wallace. 2023. Identification and properties of intense star-forming galaxies at redshifts zgt;10. *Nature Astronomy* 7(5):611–621.

Sailer, Noah, Emanuele Castorina, Simone Ferraro, and Martin White. 2021. Cosmology at high redshift — a probe of fundamental physics. *Journal of Cosmology and Astroparticle Physics* 2021(12):049.

Saliwanchik, Benjamin R. B., Aaron Ewall-Wice, Devin Crichton, Emily R. Kuhn, Deniz Ölçek, Kevin Bandura, Martin Bucher, Tzu-Ching Chang, H. Cynthia Chiang, Kit Gerodias, Kabelo Kesebonye, Vincent MacKay, Kavilan Moodley, Laura B. Newburgh, Viraj Nistane, Jeffrey B.

Peterson, Elizabeth Pieters, Carla Pieterse, Keith Vanderlinde, Jonathan L. Sievers, Amanda Weltman, and Dallas Wulf. 2021. Mechanical and optical design of the HIRAX radio telescope. In *Ground-based and airborne telescopes viii*, ed. Heather K. Marshall, Jason Spyromilio, and Tomonori Usuda, vol. 11445, 114455O. International Society for Optics and Photonics, SPIE.

Santos, M., P. Bull, S. Camera, S. Chen, J. Fonseca, I. Heywood, M. Hilton, M. Jarvis, G. I. G. Jozsa, K. Knowles, L. Leeuw, R. Maartens, E. Malefahlo, K. McAlpine, K. Moodley, P. Patel, A. Pourtsidou, M. Prescott, K. Spekkens, R. Taylor, A. Witzemann, and I. H. Whittam. 2016. A Large Sky Survey with MeerKAT. In *Meerkat science: On the pathway to the ska*, 32. `1709.06099`.

Schlegel, David J., Juna A. Kollmeier, Greg Aldering, Stephen Bailey, Charles Baltay, Christopher Bebek, Segev BenZvi, Robert Besuner, Guillermo Blanc, Adam S. Bolton, Ana Bonaca, Mohamed Bouri, David Brooks, Elizabeth Buckley-Geer, Zheng Cai, Jeffrey Crane, Regina Demina, Joseph DeRose, Arjun Dey, Peter Doel, Xiaohui Fan, Simone Ferraro, Douglas Finkbeiner, Andreu Font-Ribera, Satya Gontcho A Gontcho, Daniel Green, Gaston Gutierrez, Julien Guy, Henry Heetderks, Dragan Huterer, Leopoldo Infante, Patrick Jelinsky, Dionysios Karagiannis, Stephen M. Kent, Alex G. Kim, Jean-Paul Kneib, Anthony Kremin, Luzius Kronig, Nick Konidaris, Ofer Lahav, Michael L. Lampton, Martin Landriau, Dustin Lang, Alexie Leauthaud, Michael E. Levi, Michele Liguori, Eric V. Linder, Christophe Magneville, Paul Martini, Mario Mateo, Patrick McDonald, Christopher J. Miller, John Moustakas, Adam D. Myers, John Mulchaey, Jeffrey A. Newman, Peter E. Nugent, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Antonella Palmese Anthony L. Piro, Claire Poppett, Jason X. Prochaska, Anthony R. Pullen, David Rabinowitz, Anand Raichoor, Solange Ramirez, Hans-Walter Rix, Ashley J. Ross, Lado Samushia, Emmanuel Schaan, Michael Schubnell, Uros Seljak, Hee-Jong Seo, Stephen A. Shectman, Edward F. Schlafly, Joseph Silber, Joshua D. Simon, Zachary Slepian, Anže Slosar, Marcelle Soares-Santos, Greg Tarlé, Ian Thompson, Monica Valluri, Risa H. Wechsler, Martin White, Michael J. Wilson, Christophe Yèche, Dennis Zaritsky, and Rongpu Zhou. 2022. The megamapper: A stage-5 spectroscopic instrument concept for the study of inflation and dark energy. `2209.04322`.

Scolnic, D. M., D. O. Jones, A. Rest, Y. C. Pan, R. Chornock, R. J. Foley, M. E. Huber, R. Kessler, G. Narayan, A. G. Riess, S. Rodney, E. Berger, D. J. Brout, P. J. Challis, M. Drout, D. Finkbeiner, R. Lunnan, R. P. Kirshner, N. E. Sanders, E. Schlafly, S. Smartt, C. W. Stubbs, J. Tonry, W. M. Wood-Vasey, M. Foley, J. Hand, E. Johnson, W. S. Burgett, K. C. Chambers, P. W. Draper, K. W. Hodapp, N. Kaiser, R. P. Kudritzki, E. A. Magnier, N. Metcalfe, F. Bresolin, E. Gall, R. Kotak, M. McCrum, and K. W. Smith. 2018. The complete light-curve sample of spectroscopically confirmed sne ia from pan-starrs1 and cosmological constraints from the combined pantheon sample. *The Astrophysical Journal* 859(2):101.

Shaw, J. Richard, Kris Sigurdson, Michael Sitwell, Albert Stebbins, and Ue-Li Pen. 2015. Coaxing cosmic 21cm fluctuations from the polarized sky using m-mode analysis. *Physical Review D* 91(8).

Slepian, D. 1978. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V - The discrete case. *AT T Technical Journal* 57:1371–1430.

Soares, Paula S, Catherine A Watkinson, Steven Cunnington, and Alkistis Pourtsidou. 2021. Gaussian process regression for foreground removal in h scpi/scp intensity mapping experiments. *Monthly Notices of the Royal Astronomical Society* 510(4):5872–5890.

Switzer, E. R., K. W. Masui, K. Bandura, L.-M. Calin, T.-C. Chang, X.-L. Chen, Y.-C. Li, Y.-W. Liao, A. Natarajan, U.-L. Pen, and et al. 2013. Determination of z 0.8 neutral hydrogen fluctuations using the 21 cm intensity mapping autocorrelation. *Monthly Notices of the Royal Astronomical Society: Letters* 434(1):L46–L50.

Tegmark, Max. 1997. Measuring cosmological parameters with galaxy surveys. *Physical Review Letters* 79(20):3806–3809.

Thomas, Rajat M., Saleem Zaroubi, Benedetta Ciardi, Andreas H. Pawlik, Panagiotis Labropoulos, Vibor Jelić, Gianni Bernardi, Michiel A. Brentjens, A. G. de Bruyn, Geraint J. A. Harker, Leon V. E. Koopmans, Garrelt Mellema, V. N. Pandey, Joop Schaye, and Sarod Yatawatta. 2009. Fast large-scale reionization simulations. *Monthly Notices of the Royal Astronomical Society* 393(1): 32–48.

Thyagarajan, Nithyanandan, N. Udaya Shankar, Ravi Subrahmanyan, Wayne Arcus, Gianni Bernardi, Judd D. Bowman, Frank Briggs, John D. Bunton, Roger J. Cappallo, Brian E. Corey, Ludi deSouza, David Emrich, Bryan M. Gaensler, Robert F. Goeke, Lincoln J. Greenhill, Bryna J. Hazelton, David Herne, Jacqueline N. Hewitt, Melanie Johnston-Hollitt, David L. Kaplan, Justin C. Kasper, Barton B. Kincaid, Ronald Koenig, Eric Kratzenberg, Colin J. Lonsdale, Mervyn J. Lynch, S. Russell McWhirter, Daniel A. Mitchell, Miguel F. Morales, Edward H. Morgan, Divya Oberoi, Stephen M. Ord, Joseph Pathikulangara, Ronald A. Remillard, Alan E. E. Rogers, D. Anish Roshi, Joseph E. Salah, Robert J. Sault, K. S. Srivani, Jamie B. Stevens, Prabu Thiagaraj, Steven J. Tingay, Randall B. Wayth, Mark Waterson, Rachel L. Webster, Alan R. Whitney, Andrew J. Williams, Christopher L. Williams, and J. Stuart B. Wyithe. 2013. A study of fundamental limitations to statistical detection of redshifted hi from the epoch of reionization. *The Astrophysical Journal* 776(1):6.

Vanderlinde, Keith, Adrian Liu, Bryan Gaensler, Dick Bond, Gary Hinshaw, Cherry Ng, Cynthia Chiang, Ingrid Stairs, Jo-Anne Brown, Jonathan Sievers, Juan Mena, Kendrick Smith, Kevin Bandura, Kiyoshi Masui, Kristine Spekkens, Leo Belostotski, Matt Dobbs, Neil Turok, Patrick

Boyle, Michael Rupen, Tom Landecker, Ue-Li Pen, and Victoria Kaspi. 2019. The Canadian Hydrogen Observatory and Radio-transient Detector (CHORD). In *Canadian long range plan for astronomy and astrophysics white papers*, vol. 2020, 28. `1911.01777`.

Virone, Giuseppe, Pietro Bolli, Fabio Paonessa, Giuseppe Pupillo, Stefan J. Wijnholds, Stefania Matteoli, Andrea Maria Lingua, Jader Monari, Giuseppe Addamo, and Oscar A. Peverini. 2018. Strong mutual coupling effects on lofar: Modeling and in situ validation. *IEEE Transactions on Antennas and Propagation* 66(5):2581–2588.

Wolleben, M., T. L. Landecker, W. Reich, and R. Wielebinski. 2006. An absolutely calibrated survey of polarized emission from the northern sky at 1.4 GHz. *Astronomy and Astrophysics* 448(1):411–424.

Wolz, L., F. B. Abdalla, C. Blake, J. R. Shaw, E. Chapman, and S. Rawlings. 2014. The effect of foreground subtraction on cosmological measurements from intensity mapping. *Monthly Notices of the Royal Astronomical Society* 441(4):3271–3283. `https://academic.oup.com/mnras/article-pdf/441/4/3271/4053328/stu792.pdf`.

Wolz, Laura, Alkistis Pourtsidou, Kiyoshi W Masui, Tzu-Ching Chang, Julian E Bautista, Eva-Maria Müller, Santiago Avila, David Bacon, Will J Percival, Steven Cunnington, Chris Anderson, Xuelei Chen, Jean-Paul Kneib, Yi-Chao Li, Yu-Wei Liao, Ue-Li Pen, Jeffrey B Peterson, Graziano Rossi, Donald P Schneider, Jaswant Yadav, and Gong-Bo Zhao. 2021. H scpi/scp constraints from the cross-correlation of eBOSS galaxies and green bank telescope intensity maps. *Monthly Notices of the Royal Astronomical Society* 510(3):3495–3511.

Wu, Fengquan, Jixia Li, Shifan Zuo, Xuelei Chen, Santanu Das, John P Marriner, Trevor M Oxholm, Anh Phan, Albert Stebbins, Peter T Timbie, Reza Ansari, Jean-Eric Campagne, Zhiping Chen, Yanping Cong, Qizhi Huang, Juhun Kwak, Yichao Li, Tao Liu, Yingfeng Liu, Chenhui Niu, Calvin Osinga, Olivier Perdereau, Jeffrey B Peterson, John Podczerwinski, Huli Shi, Gage Siebert, Shijie Sun, Haijun Tian, Gregory S Tucker, Qunxiong Wang, Rongli Wang, Yougang Wang, Yanlin Wu, Yidong Xu, Kaifeng Yu, Zijie Yu, Jiao Zhang, Juyong Zhang, and Jialu Zhu. 2021. The tianlai dish pathfinder array: design, operation, and performance of a prototype transit radio interferometer. *Monthly Notices of the Royal Astronomical Society* 506(3):3455–3482.

Wuensche, C.A. 2019. The bingo telescope: a new instrument exploring the new 21 cm cosmology window. *Journal of Physics: Conference Series* 1269(1):012002.

Yoon, K. W., P. A. R. Ade, D. Barkats, J. O. Battle, E. M. Bierman, J. J. Bock, J. A. Brevik, H. C. Chiang, A. Crites, C. D. Dowell, L. Duband, G. S. Griffin, E. F. Hivon, W. L. Holzapfel, V. V. Hristov, B. G. Keating, J. M. Kovac, C. L. Kuo, A. E. Lange, E. M. Leitch, P. V. Mason, H. T.

Nguyen, N. Ponthieu, Y. D. Takahashi, T. Renbarger, L. C. Weintraub, and D. Woolsey. 2006. The robinson gravitational wave background telescope (bicep): a bolometric large angular scale cmb polarimeter. In *Millimeter and submillimeter detectors and instrumentation for astronomy iii*, ed. Jonas Zmuidzinas, Wayne S. Holland, Stafford Withington, and William D. Duncan. SPIE.

Zhang, Le, Emory F. Bunn, Ata Karakci, Andrei Korotkov, P. M. Sutter, Peter T. Timbie, Gregory S. Tucker, and Benjamin D. Wandelt. 2016. BAYESIAN SEMI-BLIND COMPONENT SEPARATION FOR FOREGROUND REMOVAL IN INTERFEROMETRIC 21 cm OBSERVATIONS. *The Astrophysical Journal Supplement Series* 222(1):3.

Zuo, Shifan, Xuelei Chen, and Yi Mao. 2023. A semiblind PCA-based foreground subtraction method for 21 cm intensity mapping. *The Astrophysical Journal* 945(1):38.