EXTRACTING COSMOLOGICAL INFORMATION FROM THE SKY, FROM THE COSMIC MICROWAVE BACKGROUND TO INTENSITY MAPPING WITH NEUTRAL HYDROGEN

by

ANH PHAN

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

Doctor of Philosophy
(Physics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2022

Date of final oral examination: 06/27/2022

The dissertation is approved by the following members of the Final Oral Committee:

Peter T. Timbie, Professor, Physics
Moritz Münchmeyer, Assistant Professor, Physics
Keith Bechtol, Assistant Professor, Physics
Ke Zhang, Assistant Professor, Astronomy

Abstract

In this thesis, I will introduce several methods to extract cosmological information from observations of the sky, ranging from the cosmic microwave background (CMB) to hydrogen line intensity mapping. Starting with the first light, the CMB, a cosmological Boltzmann code is presented to calculate the CMB power spectra from different theoretical models, for a flat ($\Omega_k = 0$) cosmology. It can rapidly compute the CMB power spectrum accurately up to high multipoles. After reviewing the chronological evolution of the Universe, I will then introduce the basic ideas of intensity mapping of neutral hydrogen after recombination, which tells us about structure formation. I will present the first results from the Tianlai Dish interferometer array, which is a new instrument specifically designed and constructed for hydrogen intensity mapping between redshift $z = 0$ and $z = 2.55$. The array is still in its infancy, and a thorough understanding of the instrument through simulation, calibration, noise analysis is described. An eigen-decomposition method to remove the Sun signal from the timestream data of radio interferometers is discussed in detail. It is applied to Tianlai data, which helps the instrument collect usable data during the daytime. Finally, a machine learning method is presented that maximizes the cross-correlation between hydrogen intensity maps and galaxy redshift surveys as a tool for detecting the hydrogen signal in the presence of bright foregrounds.

Acknowledgments

I would like to acknowledge the warmest thanks to my advisor, Peter Timbie, who made this research possible. His guidance has always been extremely helpful throughout my graduate school journey. His thoughtful comments and suggestions on my papers and thesis made them better. I would also give special thanks to Santanu Das, who introduced me to programming and had always been patient with teaching me coding practice.

I would also like to thank my family, Drew, and Sherry for their continued support and love. Last but not least, I am most thankful to my wife, Elizabeth Scray, who had always cheered me on and given me the motivation to go through difficult times. I wish we had more time together, but you will be missed forever and always. Rest in peace, Lizzy!

To Elizabeth Scray, who made this all possible

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.0.1 History of Modern Cosmology

Cosmology is the scientific study of the origin and development of the Universe. Although it has been studied for thousands of years, modern cosmology only began with Albert Einstein's publication of his theory of general relativity. After that, several renowned cosmologists such as de Sitter, Schwarzschild, and Eddington contributed important ideas on the curvature of the Universe, the nature of spacetime, and the first confirmation of general relavity. Alexander Friedmann later introduced the idea of an expanding Universe that contained moving matter, in contrast to the static eternal Universe as was widely believed at the time. A few years later, in 1926, Hubble confirmed that the Universe was indeed expanding by showing that galaxies seemed to be receding at a rate proportional to their distance.

Our current understanding is that the Universe started with the Big Bang, where space-time and everything within it begin, followed by a rapid accelerating expansion called inflation during the first 10^{-32} second. During this period, quantum fluctuations formed, which later became the seeds for large scale structure in the Universe. A

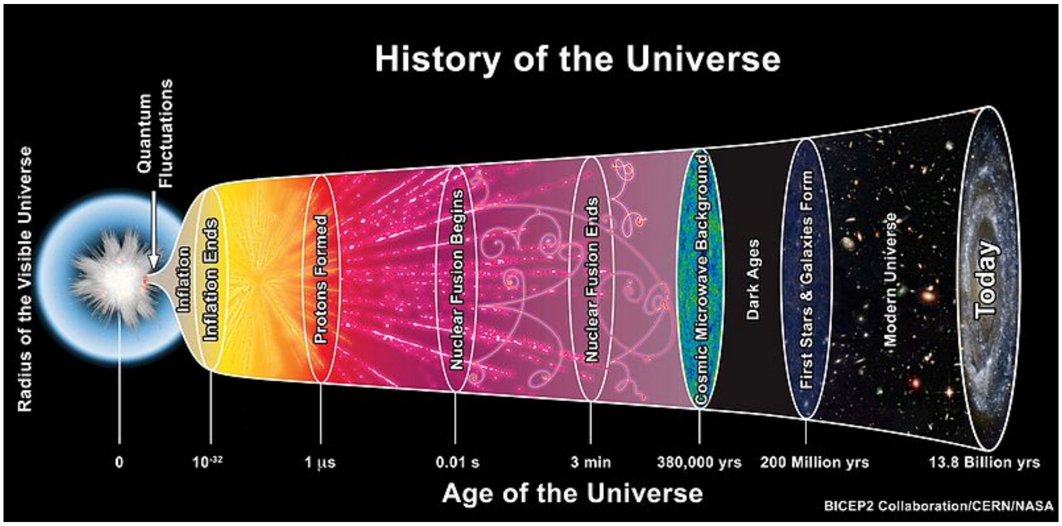

Figure 1.1: A history of the Universe

summary of the evolution of the Universe is shown in Figure 1.1. As the Universe expanded from about 0.01 seconds to 3 minutes after the Big Bang, the temperature cooled to below 10^9 K (0.1 MeV), during which the production of nuclei other than hydrogen began. This is known as Big Bang nucleosynthesis. Most of the Universe's helium (^{4}He), together with a small amount of helium-2 (^{2}He) and helium-3 (^{3}He), and even smaller amounts of lithium (^{7}Li), were produced during this phase. However, photons still remained tightly coupled to electrons via Compton scattering. Below 10^4 K (0.1 eV), free electrons could then couple with protons to form hydrogen atoms. This era is called recombination. At 380,000 years after the Big Bang, the coupling between photons and baryons becomes weak enough at the last scattering surface that photons could travel unimpeded through the Universe. The hot thermal photons from the last scattering surface are still traveling to us now, and these photons are called the Cosmic Microwave Background (CMB) because they are redshifted to the microwave range. The details of this process will be discussed further in the next chapter.

Because there are no luminous objects after the CMB, the Universe enters an epoch called the Dark Ages. At around 200 million years, the first generation of luminous objects starts to form inside dark matter halos. Collapsed dark matter makes dark matter halos in over-dense regions. This is the structure formation era, beginning at the Cosmic Dawn. Inside dark matter halos, baryonic matter gathers to form primordial stars from the metal-poor gas. Larger luminous objects such as X-ray binaries, galaxies, supernovas, quasars also start to form and illuminate the Universe. As more structures are formed, ionizing photons produced by stars and galaxies start to ionize the neutral hydrogen cloud. This epoch is called the Epoch of Reionization (EoR). Those illuminating objects affect the thermal and ionization history of the intergalacic medium (IGM). The temperature of the IGM first decreases adiabatically as the Universe expands, but X-ray photons heat the IGM significantly afterward. Details of the exact redshifts of the Dark Ages and EoR are still ongoing subjects of investigation. After the EoR, the Universe experiences accelerated expansion due to dark energy. In addition, galaxies are not distributed in a random manner but form the Large Scale Structure (LSS).

Within the last 50 years, cosmologists have made important contributions to the understanding of our Universe, thanks to the rapid development of state-of-the-art instruments. Models for nucleosynthesis are constrained by measurements of the primordial metallicity abundance using absorption lines in hot and ionized regions of galaxies and distant quasars. The Big Bang picture was further confirmed by the discovery and measurements of the CMB. The COsmic Background Explorer (COBE) discovered anisotropies in the CMB temperature field (fractional variation $\delta T/T \approx 10^{-5}$) and determined the blackbody temperature of the CMB to be 2.725 ± 0.001 K. Later, the WMAP and Planck satellites measured the anisotropies of the CMB at higher angular resolution. Progress was also made in characterizing the Large Scale Structure with 3D galaxy surveys, primarily using optical telescopes. Intensity mapping, measuring the 3D Large Scale Structure with emission lines, is one of the latest newcomers to observational cosmology and is actively being developed.

The original Big Bang model had several fundamental shortcomings, known as the horizon, flatness, and monopole problems. These were solved by the concept of inflation, in which the early Universe underwent a brief, rapid expansion from a singularity, a point of infinite density and gravity. The current 'standard model' of cosmology, called the ΛCDM model, posits an inflationary beginning to a Universe now dominated by a cosmological constant (Λ), or dark energy, and cold dark matter (CDM). It offers remarkably good explanations for a wide range of observed phenomena, including the cosmic microwave background (CMB) radiation, the abundance of light elements, and large-scale structure, which formed from the quantum fluctuations that were magnified during inflation. We need a large amount of observational data to fit all the parameters in the inflationary ΛCDM model. In addition, how the Hubble rate, which describes the rate of expansion of the Universe, varies with redshift is an open question. The expansion rate of the Universe is decelerated by regular matter but is accelerated by dark energy. Fortunately, these questions can be answered by studying the fluctuations in the CMB temperature and the distribution of matter in the Universe. In addition, there is another challenge: local measurements of the Hubble parameter from supernovae and lensing time delays disagree with the value inferred from a ΛCDM fit to the CMB, with local measurements indicating higher values. This disagreement does not come from known systematic effects in either measurement and is called the *Hubble tension*. This may imply new physics beyond the standard ΛCDM paradigm. In this thesis, I will introduce a tool for CMB analysis, named `CMBAns`, and show how it can be used to set constraints on the ΛCDM model parameters. In `CMBAns`, the Hubble parameter can be visually modified as a function of redshift through a MATLAB GUI, then the program will output the CMB power spectrum of the corresponding model. Later, I will introduce intensity mapping using radio interferometric observations of neutral hydrogen, a new technique for mapping the large scale structure of the Universe.

1.0.2 From the CMB to Intensity Mapping

The CMB is the Universe's first light, which was released about 380,000 years after the Big Bang. The CMB thus offers a remarkable view of the Universe right after recombination, when charged electrons and protons first became bound to form electrically neutral hydrogen atoms. The CMB angular power spectrum can tell us about the geometric properties of the Universe, inflation, and the composition of the Universe including baryonic matter, dark matter, and dark energy, because it is highly sensitive to the ΛCDM model's cosmological parameters. However, the CMB, being only a 2-dimensional map of the Universe, does not directly observe structures formed after recombination. There are also secondary anisotropies, such as gravitational lensing and the (S-Z) effect, that arise from processes along the line of sight. However, the secondary anisotropies are integrated along the line of sight, so the redshift dependence is mostly washed out. This inspires us to turn to intensity mapping as a way to learn about structure formation (galaxies, galaxy clusters and larger structures). By mapping the 21 cm emission line due to the hyperfine structure of neutral hydrogen in 3-dimensions, we can probe the Universe from the dark ages, to reionization, all the way to the present day. Although the ΛCDM model is very good at predicting the very large-scale distribution of cosmological objects, on the scale of galaxy clusters or individual galaxies, nonlinear processes including baryonic physics, gas heating and cooling, star formation, etc. can complicate the prediction. Intensity mapping can augment the ΛCDM model in such cases. It can be used to determine the power spectrum of matter fluctuations. Both the CMB and intensity mapping will complement each other to provide further insights into our understanding of the Universe.

1.0.3 Outline of this thesis

This thesis is organized as follows. In Chapter 2, I will briefly review our understanding of the history of the Universe. We start with the Hubble Law and use the Friedmann equation to summarize the evolution of the scale factor with time. We then present the basic ideas on the CMB and its angular power spectrum.

In Chapter 3, we will introduce a new cosmological Boltzmann code, `CMBAns`, for calculating the CMB angular power spectra and estimating cosmological parameters. It is based on a paper I wrote with Santanu Das: "Cosmic Microwave Background Anisotropy numerical solution (`CMBAns`) I: An introduction to C_ℓ calculation" published in JCAP in May, 2020 [39]. We will present the capabilities of `CMBAns` for calculating the unlensed CMB scalar/tensor angular power spectra from standard cosmological parameters such as baryon or dark matter density, reionization optical depth, Hubble parameter, and so on.

Chapter 4 presents the fundamental physics of neutral hydrogen 21 cm emission, a promising probe to study reionization and structure formation after the CMB was emitted. We introduce different astrophysical processes that affect the 21 cm line and its observation. Our main results are presented in Chapter 5 and 6. In Chapter 5, we introduce the concept of intensity mapping and the radio interferometric technique used by most 21 cm intensity mapping instruments. I describe one such instrument in detail: the Tianlai Pathfinder Array. The array is in an experimental stage, and we discuss the array performance and calibration. I will present my work in a paper with the Tianlai collaboration named "The Tianlai Dish Pathfinder Array: design, operation and performance of a prototype transit radio interferometer" that appeared in MNRAS in 2021 [135].

Chapter 6 will deal with removing the solar contamination from radio interferometer data. Sun contamination generally makes daytime data unusable for any analysis. We introduced an eigenvalue - based technique to remove up to 95% of the solar contamination without affecting weaker sources in the sky. Most of the work in this chapter is based on a paper on which I am first author: "AlgoSCR: An algorithm for Solar Contamination Removal from radio interferometric data," that appeared in MNRAS in 2022. [109]

Chapter 7 presents a 21 cm foreground removal technique using machine learning (ML). Foreground signals are much brighter than the HI signal and in general very difficult to remove. Historically, blind techniques such as Principle Component Analysis (PCA) are used. However, Makinen [87] proposed a way to combine traditional PCA with machine learning to improve the foreground removal efficacy. In this chapter, we further refine the results by introducing cross-correlation with galaxy surveys. We will give preliminary results with potential for further expansion and optimization on a larger data set. Finally, Chapter 8 provides the conclusion and discusses future CMB and HI experiments.

Chapter 2

The Cosmic Microwave Background

Chapter 2 is a general introduction to modern cosmology, with emphasis on the Cosmic Microwave Background (CMB) and its polarization. This chapter is intended to give a foundation for the following chapters. We briefly describe the homogeneous Universe and the evolution of its components. We then describe how small inhomogeneities seeded by inflation in the early Universe imprint small anisotropies in the CMB. We show how the observed CMB power spectrum can test the ΛCDM model and constrain cosmological parameters.

2.1 Introduction to Cosmology

Modern cosmology starts with the *cosmological principle*, which states that on large enough scales (larger than about 300 Mpc), the Universe possesses two important properties:

- The Universe is *homogeneous*. Homogeneity means translational invariance; the Universe looks the same at each observational point.

- The Universe is *isotropic*. Isotropy means rotational invariance; the Universe looks the same in all directions.

Homogeneity and isotropy are symmetries of *space*, not *spacetime*. Spacetime can still be curved even if its spatial part is flat.

2.1.1 Hubble's law

In 1929, Edwin Hubble published a paper titled "A relation between distance and radial velocity among extra-galactic nebulae" [65]. The paper showed that the distant galaxies in all directions are receding at a rate that is proportional to their distance from the Earth. This does not imply that the Earth is the center of the Universe. An observer at any location in the Universe would observe that distant galaxies recede away from them. It is a consequence of the fact that space itself is expanding. Hubble's law is regarded as the first observational evidence for the expansion of the Universe. Note that the Universe would not be isotropic if every point saw a recessional velocity larger in the x direction than in the y direction, for example, but our Universe does not have any measurable velocity anisotropy. Because of this simplicity, the law is often expressed as

$$v = H(t)d \tag{2.1}$$

where v is the recessional velocity, conventionally expressed in km/s, $H(t)$ is the *Hubble parameter*, i.e. the instantaneous relative rate of expansion of the Universe at time t, which has units km/s/Mpc. Its present value is called the *Hubble constant,* H_0. d is the *proper distance* in parsec (which can vary with time, unlike comoving distance, which is static). A plot of Hubble's original finding of the linear relationship between recessional velocity and proper distance is shown in Fig. 2.1.

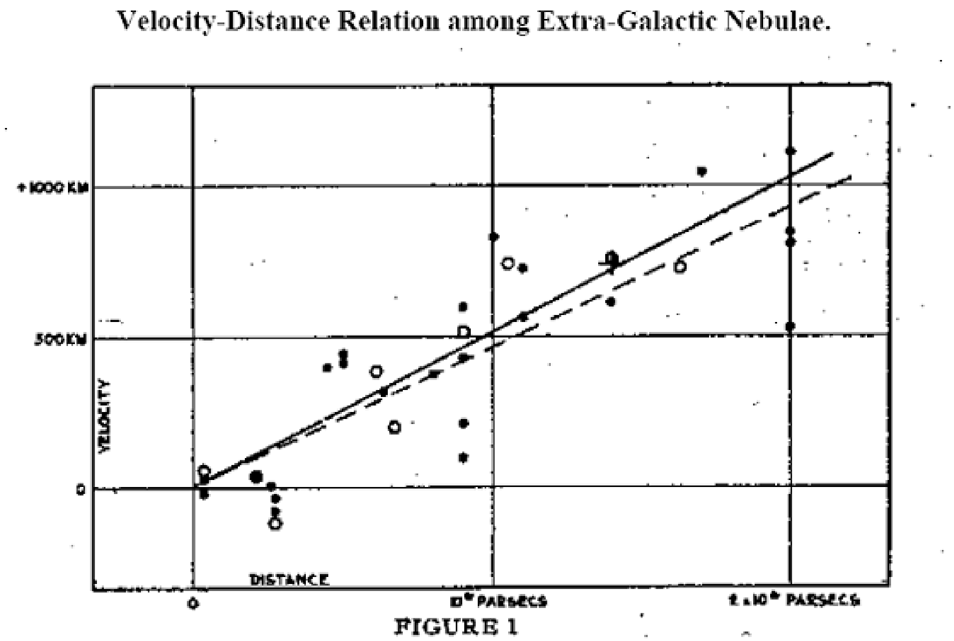

Figure 2.1: Velocity versus time graph for various nebulae outside the Milky Way galaxy. Figure from 1929 paper by Edwin Hubble [65].

2.1.2 General relativity

In 1915, fourteen years before Hubble's finding, Albert Einstein published the first geometric theory of gravitation, called general relativity, which formed another basis for modern cosmology. The theory generalizes special relativity and Newton's law of universal gravitation. It relates the curvature of spacetime to the energy and momentum of matter or radiation present. In general relativity, the fundamental quantity is the *metric*, which describes the geometry of spacetime by giving the separation distance between neighboring points. In four-dimensional spacetime, the infinitesimal separation is

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu, \tag{2.2}$$

where $g_{\mu\nu}$ is the metric, μ and ν are indices taking the values $0, 1, 2$ and 3, x^0 is the time coordinate and x^1, x^2 and x^3 are the three spatial coordinates. The Einstein summation notation is assumed. The metric evolves in accordance with the Einstein field equation:

$$\boxed{R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}}, \tag{2.3}$$

where $R_{\mu\nu}$ is the Ricci curvature tensor, R is the Ricci scalar curvature, $g_{\mu\nu}$ is the metric tensor, Λ is the cosmological constant, G is Newton's gravitational constant, c is the speed of light in vacuum, and $T_{\mu\nu}$ is the energy-momentum tensor.

The energy-momentum tensor can be written as

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu - pg_{\mu\nu}, \tag{2.4}$$

where ρ is the total energy density, p is the pressure, and u is the four-velocity. Conservation of energy requires that the covariant derivative of the energy-momentum tensor must vanish:

$$\nabla_\mu T^{\mu\nu} = 0. \tag{2.5}$$

2.1.3 Friedmann-Lemaître-Robertson-Walker (FLRW) metric

In 1922, Russian physicist Alexander Friedmann attempted to solve the most general case of general relativity, based on the *cosmological principle* that, at any given time, the Universe should not have any preferred location. This requires that the spatial part of the metric have a constant curvature. With this assumption, Friedmann was able to find the exact solution to the Einstein field equation (the solution is known as the Friedmann-Lemaître-Robertson-Walker (FLRW) metric; it is the most general metric which has constant spatial curvature). The solution also describes an expanding

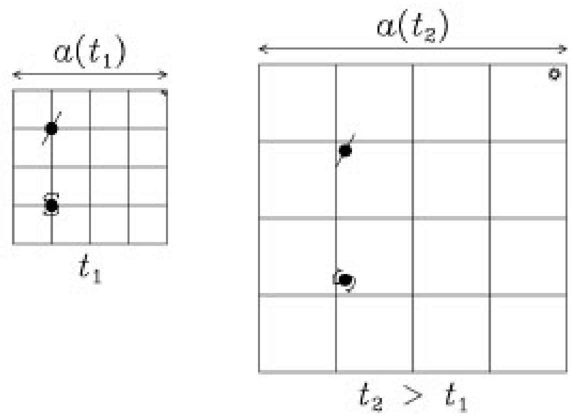

Figure 2.2: Comoving coordinate grid, expanding corresponding to increasing cosmic time. Image taken from `https://ned.ipac.caltech.edu/level5/March02/Bertschinger/Figures/figure1.jpg`

Universe, in agreement with Hubble's finding. We will take a brief look at the derivation of the FLRW metric, but first the idea of comoving coordinates is introduced.

The fact that the Universe is homogeneous and expands uniformly allows us to change to a different coordinate system, known as the **comoving coordinates**. They expand and contract with the evolution of the Universe, as shown in Figure 2.2. The comoving coordinates (X, Y, Z) are related to the physical coordinates (x, y, z) by

$$\begin{aligned} X &= a(t)x \\ Y &= a(t)y \\ Z &= a(t)z. \end{aligned} \tag{2.6}$$

The distance between two points in the Universe is thus given by $D = \sqrt{X^2 + Y^2 + Z^2}$, and if $d = \sqrt{x^2 + y^2 + z^2}$ is the proper distance between those two points at the present time, we can write $d = a(t)D$. Normally, at present time t_0, $a(t_0)$ is set to equal 1, for simplicity.

We can write the metric in (2.2) as

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}l^2, \tag{2.7}$$

where

$$\mathrm{d}l^2 = g_{ij}\mathrm{d}x^i\mathrm{d}x^j. \tag{2.8}$$

The FLRW metric has the form:

$$\begin{aligned}\mathrm{d}s^2 &= -c^2\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2\right] \\ &= -c^2\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right],\end{aligned} \tag{2.9}$$

where k is a constant representing the curvature of space.

2.1.4 Geometry of the Universe

Notice that the following substitution

$$k \to \frac{k}{|k|}, \qquad r \to \sqrt{|k|}r, \qquad a(t) \to \frac{a(t)}{\sqrt{|k|}} \tag{2.10}$$

leaves (2.9) invariant, in cases where k is nonzero. Therefore, the only relevant parameter is $k/|k|$, and we can rescale (2.9) so that k takes on only the following values:

1. $k = -1$, negative curvature. This describes an open Universe.

2. $k = 0$, zero curvature. The spatial part of the Universe is flat.

3. $k = +1$, positive curvature. This describes a closed Universe.

The line (spatial) element of the FLRW metric is found by setting $\mathrm{d}t = 0$ in (2.9):

$$\mathrm{d}l^2 = a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2\right)\right] \tag{2.11}$$

- First, consider $k = 0$, defining

$$r = \chi. \tag{2.12}$$

This may seem redundant, but it generalizes to the case of $k = \pm 1$. The line element of the hypersurface at any moment t_0 is

$$\begin{aligned} \mathrm{d}l^2 &= a^2(t_0)\left[\mathrm{d}\chi^2 + \chi^2\left(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2\right)\right] \\ &= a^2(t_0)\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right) \\ &= a^2(t_0)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j, \end{aligned} \tag{2.13}$$

which is simply flat Euclidean space with an added scale factor. Although the 3D spatial surface has Euclidean geometry, the 4D spacetime hypersurface is still curved. The co-moving distance can still change due to the changing scale factor, hence the Riemann curvature tensor is not necessarily zero. The metric is clearly invariant under spatial translation $x^i \rightarrow x^i + a^i$, with a^i being any arbitrary constant, and under rotation $x^i \rightarrow R^i{}_k x^k$, with $\delta_{ij}R^i{}_k R^j{}_l = \delta_{kl}$.

- For $k = 1$, we can define a new coordinate $\chi\left(r\right)$ such that

$$\mathrm{d}\chi^2 = \frac{\mathrm{d}r^2}{1-r^2} \tag{2.14}$$

and $\chi(0) = 0$. After integrating both sides, we get

$$r = \sin\chi. \tag{2.15}$$

The line element (2.11) for the space at $t = t_0$ is

$$\mathrm{d}l^2 = a^2(t_0)\left[\mathrm{d}\chi^2 + \sin^2\xi\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right]. \tag{2.16}$$

This is the metric for the three-sphere of radius $a(t_0)$, i.e. the set of points equidistant $a(t_0)$ from a fixed central point in 4-dimensional Euclidean space. Recall that a three-dimensional surface in four-dimensional Euclidean space is given by the Cartesian representation $x^2+y^2+z^2+w^2 = a^2(t_0)$, where $a(t_0)$ is the radius of the sphere. The line element in (2.16) can be found by parametrizing $x^1 = a(t_0)\cos\chi$, $y = a(t_0)\sin\chi\cos\theta$, $z = a(t_0)\sin\chi\sin\theta\cos\phi$, $w = a(t_0)\sin\chi\sin\theta\sin\phi$ and then calculating the Jacobian. Note that the curvature of the three-sphere is an intrinsic property and there is no need of a higher dimensional space for it to live in. This model describes a closed, spherical Universe, directly analogous to the surface of a sphere.

- Finally, for $k = -1$, define a new coordinate $\chi(r)$ such that

$$\mathrm{d}\chi^2 = \frac{\mathrm{d}r^2}{1 + r^2} \tag{2.17}$$

and $\chi(0) = 0$. Integrating both sides leads to the result

$$r = \sinh\chi. \tag{2.18}$$

The line element (2.11) for the space at $t = t_0$ is

$$\mathrm{d}l^2 = a^2(t_0)\left[\mathrm{d}\chi^2 + \sinh^2\chi\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right]. \tag{2.19}$$

We see that in each case, the line (spatial) element of the FLRW metric is the same at any one point in space (homogeneous) and spherically symmetric about that point (isotropic). There is neither a special point nor a special direction.

2.1.5 Conformal Time

We need to find an interchangeable relationship between time and distance. We will need to define a new quantity, called the conformal time η, so that the ratio between the proper time t and conformal time η is the same as the ratio between proper distance d and comoving distance χ - namely the scale factor a. We define the conformal time at a certain time t as

$$\eta(t) = \int_0^t \frac{\mathrm{d}t'}{a(t')}, \tag{2.20}$$

where we have taken the Big Bang to be at $t = 0$. By convention, $\eta_0 = \eta(t_0) = 1.48 \times 10^{18}$ s is the conformal time today, which is the radius of the current observable Universe with diameter of 93 billion light years. With

$$\eta - \eta_i = \int_{t_i}^t \frac{dt'}{a\left(t'\right)}, \tag{2.21}$$

$c\left(\eta - \eta_i\right)$ represents the comoving distance travelled by a photon between the time η_i and η, or t_i and t. The conformal time allows us to rewrite the metric in 2.9 as

$$\mathrm{d}s^2 = a^2(\eta)\left[-c^2\mathrm{d}\eta^2 + \frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2\right)\right] \tag{2.22}$$

so that the scale factor becomes a conformal factor. If a has dimensions then $c\eta$ is dimensionless, and if a is dimensionless, then η is time.

2.1.6 The Friedmann Equations

Having the FLRW metric, our next natural step is to plug it into Einstein's field equation to derive the Friedmann equations relating the scale factor $a(t_0)$ to the energy-momentum of the Universe. Since the Universe is not empty, we are not interested in the vacuum solutions to Einstein's field equation. We will model the matter and energy in the Universe as a perfect fluid, which is a fluid that is completely characterized by its rest frame energy density ρ and isotropic pressure p. It has no heat conduction ($T^{0i} = T^{i0} = 0$) and no viscosity ($T^{ij} = 0$ if $i \neq j$). The energy-momentum tensor for a perfect fluid, in units where $c = 1$, can be written

$$T_{\mu\nu} = (\rho + p)\, U_\mu U_\nu + p g_{\mu\nu}, \tag{2.23}$$

where $U^\mu = (1, 0, 0, 0)$ is the four-velocity of the fluid. The four-velocity of the fluid at rest in the comoving frame is

$$U^\mu = (1, 0, 0, 0). \tag{2.24}$$

The energy-momentum tensor in (2.23) becomes

$$T^{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & & & \\ 0 & & g_{ij}p & \\ 0 & & & \end{pmatrix}. \tag{2.25}$$

Conservation of energy requires that the covariant derivative of the energy-momentum tensor must vanish

$$\nabla_\mu T^{\mu\nu} = 0. \tag{2.26}$$

First, we consider the zero component of (2.26)

$$\begin{aligned} 0 &= \nabla_\mu T^\mu{}_0 \\ &= \partial_\mu T^\mu{}_0 + \Gamma^\mu_{\mu\lambda} T^\lambda{}_0 - \Gamma^\lambda_{\mu 0} T^\lambda{}_0 \\ &= -\partial_0 \rho - 3\frac{\dot{a}}{a}(\rho + p). \end{aligned} \tag{2.27}$$

Here, the time dependence of a, ρ, p is implicit. The relationship between ρ and p is described by the equation of state[1]

$$p = w\rho, \tag{2.28}$$

where w is a constant independent of time. The equation of state (2.27) becomes

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a}. \tag{2.29}$$

After integration, with w constant,

$$\rho \propto a^{-3(1+w)}. \tag{2.30}$$

Two important examples of cosmological fluids are **matter** and **radiation**. Matter, also known as dust, is any set of collisionless, non-relativistic particles, which has positive mass density but vanishing pressure and has $w = 0$. Matter includes dark matter and baryonic matter. The energy density in matter, denoted as ρ_m, falls off as

$$\rho_m \propto a^{-3}. \tag{2.31}$$

[1] For a more detailed derivation, see Appendix **??**.

For matter the energy density is dominated by the rest energy, which is proportional to the number density. Therefore, the number density decreases proportional to a^{-3} as the Universe expands. Furthermore, for radiation $w = 1/3$ and thus the energy density for radiation, denoted as ρ_r, falls off as

$$\rho_r \propto a^{-4}. \tag{2.32}$$

The energy density for radiation falls off faster than that for matter. The reason is that, in addition to the fact that the number density of photons decreases in the same manner as the number density of dust particles, and the energy of individual photons decreases $\propto a^{-1}$ as they redshift in an expanding Universe. In the early Universe, the matter energy density is dominated by the radiation energy density. In contrast, today the radiation energy density is dominated by the matter energy denstity, with $\rho_r/\rho_m \sim 10^{-3}$.

In addition, vacuum energy (dark energy) can also be modeled as a perfect fluid with $w = -1$ and equation of state $p_\Lambda = -\rho_\Lambda$. Note that dark energy has a *negative* pressure. The vacuum energy density is constant,

$$\rho_\Lambda \propto a^0. \tag{2.33}$$

If the vacuum energy is nonzero, it will dominate over matter and radiation energy over the long term because the vacuum energy density does not decrease as the Universe expands. If this happens, the Universe becomes **vacuum-dominated.**

To derive the Friedmann equation, let's go back to the Einstein equation (2.3) with $c = 1$. The $\mu\nu = 00$ equation is

$$
\begin{aligned}
R_{00} - \frac{1}{2} R g_{00} + \Lambda g_{00} &= 8\pi G T_{00} \\
\longrightarrow -3\frac{\ddot{a}}{a} + 3\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right] - \Lambda &= 8\pi G \rho.
\end{aligned}
$$

Simplifying this equation, we get the **first Friedmann equation:**

$$
\boxed{\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3}}. \tag{2.34}
$$

Similarly, the spatial part $\mu\nu = ij$ gives

$$
2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 = -8\pi G p - \frac{k}{a^2} + \Lambda \tag{2.35}
$$

Subtracting (2.34) from (2.35) and dividing by 2, we obtain the **second Friedmann equation (acceleration equation):**

$$
\boxed{\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}}. \tag{2.36}
$$

The cosmological constant Λ is shown here for historical reasons. In some physics literature, the vacuum energy density is absorbed into the total energy density term, where $\Lambda = 8\pi G \rho_\Lambda = -8\pi G p_\Lambda$, so the total energy density in this case is $\rho = \rho_m + \rho_r + \rho_\Lambda$. The Friedmann equations are

$$
\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \tag{2.37}
$$

$$
\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p). \tag{2.38}
$$

We can see that if the negative pressure associated with the cosmological constant is large enough, the right side of (2.38) is positive, and the Universe expansion is accelerating. This is indeed the case with the current observable Universe. The expansion rate is characterized by the Hubble parameter, introduced in (2.1):

$$\boxed{H(t) = \frac{\dot{a}(t)}{a(t)}}. \tag{2.39}$$

Its present value (the Hubble constant, H_0) is $H_0 = \dot{a}(t_0)/a(t_0) = \dot{a}(t_0)/1 = 67 \pm 10$ (km/s)/Mpc. The reduced Hubble constant, h, is a dimensionless quantity defined as $h \equiv H_0/100$.

As a simple example, the de Sitter model assumes a homogeneous, isotropic Universe with zero curvature ($k = 0$), zero cosmological constant ($\Lambda = 0$) and zero pressure ($p = 0$). The energy density for this spatially flat Universe is called the critical density ρ_c, and the first Friedmann equation (2.34) is

$$H(t)^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho_c. \tag{2.40}$$

We can solve for the critical density, ρ_c, as a function of time[2]

$$\rho_c = \frac{3H_0^2}{8\pi G} \tag{2.41}$$

[2]Even though ρ_c is a function of time, it is common to define ρ_c as a constant, the critical density as of today: $\rho_c = 3H^2/8\pi G$.

The density parameter is defined as the ratio of the actual total density, ρ, in our observable Universe to the critical density, ρ_c

$$\Omega = \frac{8\pi G}{3H(t)^2}\rho = \frac{\rho}{\rho_c}. \tag{2.42}$$

The density parameter for matter (dark plus baryonic) Ω_m, radiation Ω_r, and vacuum Ω_Λ are defined similarly

$$\Omega_m = \frac{\rho_m}{\rho_c}, \qquad \Omega_r = \frac{\rho_r}{\rho_c}, \qquad \Omega_\Lambda = \frac{\rho_\Lambda}{\rho_c}, \tag{2.43}$$

so that $\Omega = \Omega_m + \Omega_r + \Omega_\Lambda$. In addition, the first Friedmann equation 2.37 can be written as

$$\Omega - 1 = \frac{k}{H(t)^2 a^2} \tag{2.44}$$

The density parameter associated with curvature Ω_k is defined as

$$\Omega_k = -\frac{k}{H(t)^2 a^2}, \tag{2.45}$$

so that

$$\Omega + \Omega_k = 1 \qquad \Leftrightarrow \qquad \Omega_m + \Omega_r + \Omega_\Lambda + \Omega_k = 1. \tag{2.46}$$

The sign of Ω_k determines whether Ω is greater than, equal to, or less than unity, which corresponds to a closed, flat, and open Universe respectively, as shown in **??**. The density parameter at the current time is denoted as Ω_0.

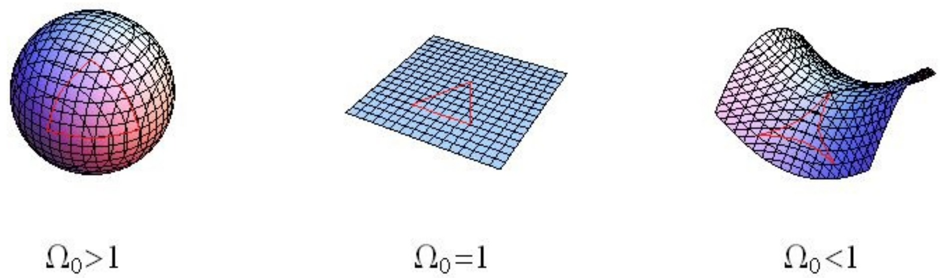

Figure 2.3: $\Omega_0 > 1$ corresponds to a closed Universe, $\Omega_0 = 0$ corresponds to a flat Universe, and $\Omega_0 < 1$ corresponds to an open Universe.

We can rewrite the Friedmann equation (2.37) in terms of the current values of the density parameters, noting that $\rho_i/\rho_{0,i} = (a(t)/a(t_0))^{-3(1+w_i)} = a(t)^{-3(1+w_i)}$:

$$\begin{aligned}
H(t)^2 = \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}(\rho_m + \rho_r + \rho_\Lambda) - \frac{k}{a^2} \\
&= \frac{8\pi G}{3}\left(\rho_{0,m}a^{-3} + \rho_{0,r}a^{-4} + \rho_{0,\Lambda}a^0\right) - \frac{k}{a^2} \\
&= H_0^2(\Omega_{0,m}a^{-3} + \Omega_{0,r}a^{-4} + \Omega_{0,\Lambda}a^0 + \Omega_{0,k}a^{-2}) \\
\Longrightarrow \frac{H(t)^2}{H_0^2} &= \Omega_{0,m}a^{-3} + \Omega_{0,r}a^{-4} + \Omega_{0,\Lambda}a^0 + \Omega_{0,k}a^{-2}. \qquad (2.47)
\end{aligned}$$

Current data from the Cosmology Supernova Project [126] suggest that we live in a flat Universe, with Ω_k very close to zero. The matter density parameter Ω_m is approximately 0.3, and the cosmological constant density parameter Ω_Λ is about 0.7. The matter density is actually higher than what can be accounted for from baryonic matter alone (matter made of protons, neutrons and electrons). In fact, most of the matter in the Universe is non-baryonic, and neither emits nor interacts with electromagnetic radiation. This type of matter is called dark matter and apparently it can only be detected indirectly from its gravitational effects. Hence, we can split Ω_m

Fit	Ω_m	Ω_Λ	Ω_k	Ω_b	Ω_c
BAO+CMB+H_0	0.267	0.733	0 (fixed)		
SNe	0.277	0.723	0 (fixed)		
SNE+BAO+H_0	0.288	0.712	0 (fixed)		
SNE+CMB	0.272	0.728	0 (fixed)		
SNE+CMB+H_0	0.262	0.738	0 (fixed)		
SNE+BAO+CMB	0.278	0.722	0 (fixed)		
SNE+BAO+CMB+H_0	0.271	0.729	0 (fixed)		
Planck	0.3175	0.6825	0 (fixed)	0.0490	0.2685
Planck + lensing	0.3036	0.6964	0 (fixed)	0.0479	0.2557
Planck + WMAP	0.315	0.685	0 (fixed)	0.049	0.266

Table 2.1: Fit results on density parameters from the Cosmology Supernova Project [126] and the Planck mission [3]. The radiation density parameter, Ω_r, is not included here.

into the baryonic matter component Ω_b and dark matter component Ω_d:

$$\Omega_m = \Omega_b + \Omega_d. \tag{2.48}$$

Data from the Planck mission [3] indicate that $\Omega_b \approx 0.049$ and $\Omega_d \approx 0.26$. In other words, the Universe contains about 4.9% baryonic matter, 26% dark matter and the rest is dark energy. Current data on the various density parameters are shown in Table 2.1.6.

2.1.7 Evolution of the scale factor

- Matter-dominated era:

The energy density for matter obeys the relationship $\rho_m = \rho_{0,m}/a^3$. Substituting for ρ_m in the Friedmann equation (2.37) and assuming $k = 0$ gives

$$\dot{a}^2 = \frac{8\pi G \rho_{0,m}}{3}\frac{1}{a}. \tag{2.49}$$

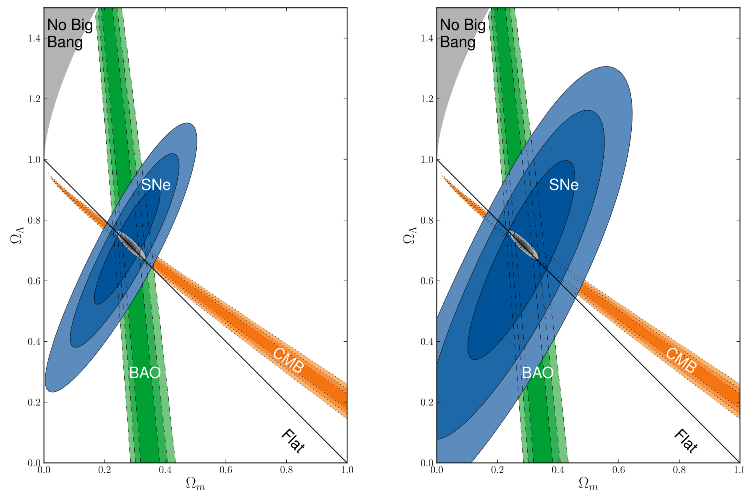

Figure 2.4: LCDM model:: 68.3%, 95.4%, and 99.7% confidence regions of the $(\Omega_m, \Omega_\Lambda)$ plane from supernovae Ia data (SNe Ia) combined with the constraints from Baryon Acoustic Oscillation (BAO) data and CMB. The left panel shows the SNe Ia confidence region only including statistical errors while the right panel shows the SN Ia confidence region with both statistical and systematic errors. Figure taken from [126].

This is a separable differential equation. Since $a(t_0) = 1$ at the present time $t = t_0$, the solution is

$$a(t) = \left(\frac{t}{t_0}\right)^{2/3}, \qquad \rho_m(t) = \frac{\rho_{0,m}}{a^3} = \frac{\rho_{0,m} t_0^2}{t^2}. \tag{2.50}$$

In a matter-dominated Universe, the Universe expands forever. Nonetheless, the rate of expansion, $H(t)$, decreases with time

$$H \equiv \frac{\dot{a}(t)}{a(t)} = \frac{2}{3t}. \tag{2.51}$$

- Radiation-dominated era:

The energy density for radiation obeys $\rho_r = \rho_{0,r}/a^4$. After carrying out the same steps as in the matter-dominated case, we arrive at

$$a(t) = \left(\frac{t}{t_0}\right)^{1/2}, \qquad \rho_r(t) = \frac{\rho_{0,r}}{a^4} = \frac{\rho_{0,r} t_0^2}{t^2}. \tag{2.52}$$

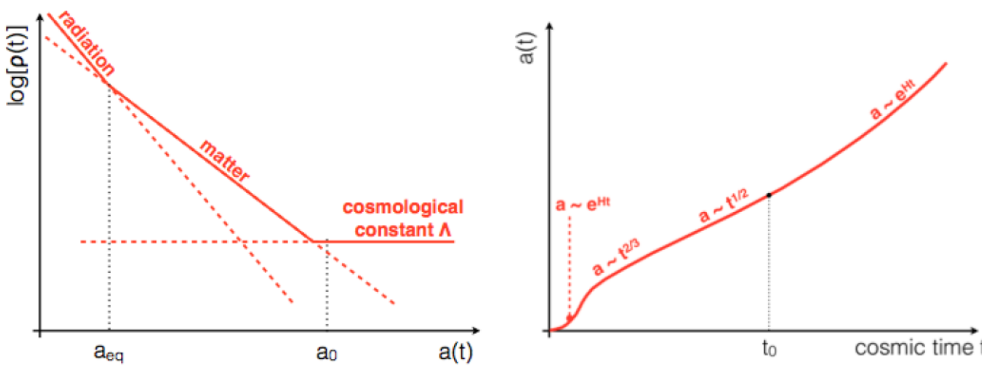

Figure 2.5: The log of the energy density ρ as a function of the scale factor $a(t)$ (left) and scale factor $a(t)$ versus time (right). In the beginning, the Universe underwent inflation. As time increases, the dominant component is, first, radiation ($a \propto t^{2/3}$), then matter ($a \propto t^{1/2}$), and now cosmological constant ($a \propto e^{Ht}$).

In a radiation-dominated Universe, the Universe also expands forever. The expansion rate, $H(t)$, also decreases with time, but it is slower than the matter-dominated case due to the pressure that radiation provides:

$$H \equiv \frac{\dot{a}(t)}{a(t)} = \frac{1}{2t}. \tag{2.53}$$

- Cosmological constant-dominated era:

The energy density for dark energy stays constant: $\rho_\Lambda = \rho_{0,\Lambda} a^0$, and the scale factor evolves as

$$a(t) = \frac{e^{Ht}}{e^{Ht_0}}, \qquad H = \sqrt{\frac{8\pi G}{3}\rho_\Lambda}. \tag{2.54}$$

Redshifts and distances

It is common in cosmology to use the term "redshift" to describe how far away a distant object is and the epochs of the Universe. This cosmological redshift is caused by the expansion of space. As we saw in previous sections, the Universe is expanding and the galaxies and clusters are moving away from each other. The light from distant stars and galaxies are found to be shifted toward the red end of the spectrum (the wavelengths are stretched), and the farther away they are, the faster they are moving.

In cosmology, it is customary to characterize redshift based on a dimensionless quantity z :

$$z = \frac{\lambda_{\text{observed}} - \lambda_{\text{emitted}}}{\lambda_{\text{emitted}}} \quad \longrightarrow \quad \boxed{1 + z = \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}}}. \tag{2.55}$$

If an object is at redshift z, it means that it is at a distance which emitted light that has been stretched by a factor of $1 + z$. At the time of writing, the highest observed redshift (besides the cosmic microwave background) has redshift of $z = 11.1$.

We can relate the redshift z to the scale factor $a(t)$ using general relativity. Light propagation obeys the FLRW metric (2.9) with $\mathrm{d}s = 0$, and for simplicity we assume the light ray propagates radially from $r = 0$ to $r = R$ so that $\mathrm{d}\theta = \mathrm{d}\phi = 0$:

$$0 = -c^2 \mathrm{d}t^2 + a^2(t) \left[\frac{\mathrm{d}r^2}{1 - kr^2} \right]. \tag{2.56}$$

This simplifies to

$$\frac{c\mathrm{d}t}{a(t)} = \frac{\mathrm{d}r}{\sqrt{1 - kr^2}}. \tag{2.57}$$

The time it takes for the ray to get from $r = 0$ to $r = R$ can be found by integrating the above equation

$$\int_{t_e}^{t_o} \frac{c\mathrm{d}t}{a(t)} = \int_0^R \frac{\mathrm{d}r}{\sqrt{1 - kr^2}}, \tag{2.58}$$

where 'e' stands for 'emitted' and 'o' stands for 'observed'. Now the subsequent crest is emitted at time $t_e + \mathrm{d}t_e = t_e + \lambda_e/c$ and is observed at time $t_o + \mathrm{d}t_o = t_o + \lambda_o/c$. The spatial coordinates in the FLRW metric are comoving, so galaxies are still at the same coordinate, and the radial integral remains the same

$$\int_{t_e+\mathrm{d}t_e}^{t_o+\mathrm{d}t_o} \frac{c\mathrm{d}t}{a(t)} = \int_0^R \frac{\mathrm{d}r}{\sqrt{1-kr^2}}. \tag{2.59}$$

From the two previous equations,

$$\begin{aligned}
0 &= \int_{t_e}^{t_o} \frac{\mathrm{d}t}{a(t)} - \int_{t_e+\mathrm{d}t_e}^{t_o+\mathrm{d}t_o} \frac{\mathrm{d}t}{a(t)} \\
&= \int_{t_e}^{t_e+\mathrm{d}t_e} \frac{\mathrm{d}t}{a(t)} + \int_{t_e+\mathrm{d}t_e}^{t_o} \frac{\mathrm{d}t}{a(t)} - \int_{t_e+\mathrm{d}t_e}^{t_o+\mathrm{d}t_o} \frac{\mathrm{d}t}{a(t)} \\
&= \int_{t_e}^{t_e+\mathrm{d}t_e} \frac{\mathrm{d}t}{a(t)} - \left(\int_{t_o}^{t_e+\mathrm{d}t_e} \frac{\mathrm{d}t}{a(t)} + \int_{t_e+\mathrm{d}t_e}^{t_o+\mathrm{d}t_o} \frac{\mathrm{d}t}{a(t)} \right) \\
&= \int_{t_e}^{t_e+\mathrm{d}t_e} \frac{\mathrm{d}t}{a(t)} - \int_{t_o}^{t_o+\mathrm{d}t_o} \frac{\mathrm{d}t}{a(t)}
\end{aligned}$$

For small variations of $a(t)$ with time, we assume it is constant inside the integral. This gives

$$\frac{\mathrm{d}t_e}{a(t_e)} = \frac{\mathrm{d}t_o}{a(t_o)}. \tag{2.60}$$

Since $\mathrm{d}t_e = \lambda_e/c$ and $\mathrm{d}t_o = \lambda_o/c$, we get the ratio

$$\frac{\lambda_o}{\lambda_e} = \frac{a(t_o)}{a(t_e)}. \tag{2.61}$$

This shows that light is stretched in an expanding Universe ($a(t_o) > a(t_e)$). If there were an intermediate observer comoving with the expansion, he would see the light with a wavelength between the emitted and final observed wavelength. We want to apply the above equation to the light received by us today, so $t_o = t_0$. Comparing with

(2.55), we see that

$$1 + z = \frac{a(t_0)}{a(t_e)} \tag{2.62}$$

Note that light emitted closer to the Big Bang ($a(t_e) \to 0$) will have $z \to \infty$,

2.1.8 Problems with the original Big Bang model

Unfortunately, the original Big Bang model still has some shortcomings. It does not completely describe some of the observed properties of the Universe, namely the Horizon Problem, the Flatness Problem, and the magnetic monopole problem.

The Horizon Problem

One of the biggest problems with the original Big Bang model is that the cosmic microwave background is uniform over the whole sky to only a few parts in 10^5. This is quite remarkable, because the only way for two regions on the sky to have the same temperature is that they are close enough to be in causal contact with each other so that equilibrium conditions can be established. Nonetheless, the fastest speed that information can travel is the speed of light, and if two said regions are far enough so that light has not had enough time to travel between them, the regions are isolated from each other. Yet this appears to be the case with our observed Universe. To an observer, the CMB radiation coming from the opposite sides of the sky is from isolated regions and has been travelling towards us since decoupling. Since the light has only reached us now, there is not enough time for it to travel all the way to the other side of the sky, yet the temperatures of these regions are very similar. This is known as the horizon problem (or the homogeneity problem).

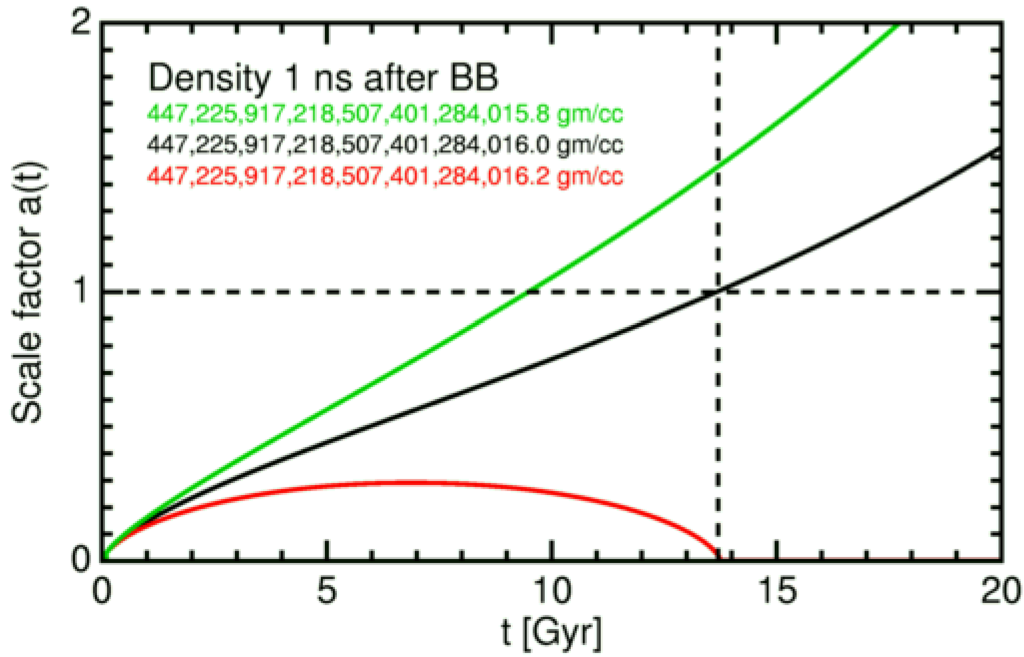

Figure 2.6: A figure illustrating $a(t)$ for three models with three different densities at 1 ns after the Big Bang. If the density of the Universe were to deviate from one by one part in 10^{27} at 1 ns after the Big Bang, the curvature of the present Universe would steer away from the flat geometry and be inconsistent with experimental data. Figure from http://www.astro.ucla.edu/~wright/cosmo_03.htm

The Flatness Problem

The flatness problem is an example of a fine-tuning problem, in which the current Universe could only result from a very finely tuned set of initial conditions. The observed density of the Universe is very close to the critical density. Hence, the geometry of the Universe is close to flat, perfectly balanced between an open Universe and a closed one (see Figure 2.6). Let's visit Equation (2.44), which reads

$$|\Omega(t) - 1| = \frac{|k|}{H(t)^2 a(t)^2}. \tag{2.63}$$

From (2.50), (2.51), (2.52), (2.53), we can deduce the product $H(t)^2a(t)^2$ in the matter- and radiation-dominated Universe:

$$H(t)^2a(t)^2 \propto t^{-2/3}, \qquad \text{matter-dominated} \tag{2.64}$$

$$H(t)^2a(t)^2 \propto t^{-1}, \qquad \text{radiation-dominated.} \tag{2.65}$$

Therefore,

$$|\Omega(t) - 1| \propto t^{2/3}, \qquad \text{matter-dominated} \tag{2.66}$$

$$|\Omega(t) - 1| \propto t, \qquad \text{radiation-dominated} \tag{2.67}$$

In either case, $|\Omega(t) - 1|$ is an increasing function of time, and the flat geometry is an unstable solution for the Universe. If $\Omega(t)$ deviates ever so slightly from unity at early time, the Universe will be very quickly curved. We know that the Universe is very close to flat today at $t_0 \approx 4 \times 10^{17}$ seconds. Assuming a radiation-dominated Universe (for simplicity), at a time 1 ns after the Big Bang, $|\Omega(t) - 1|$ needs to be less than 10^{-27}! Indeed, the flatness problem would be resolved if the Universe were exactly at the critical density, but there is no reason for it to prefer this value over the others.

Magnetic Monopole Problem

A magnetic monopole is a magnet with only one pole; it has a net magnetic charge. As a consequence of models of unification of fundamental forces, the Grand Unified Theories, magnetic monopoles were produced in large quantities in the very early Universe. In the standard Big Bang model, given the density in which magnetic monopoles were created, we would expect to observe magnetic monopoles in our present day Universe. However, we know that this is not the case, which means the magnetic monopole density is far lower than expected.

2.1.9 Inflation

The original Big Bang model can be extended to answer those aforementioned problems by adding cosmological inflation. Inflation is a theory of perfectly exponential expansion of space in the early Universe. The inflationary epoch describes a rapid exponential expansion of the Universe between 10^{-36} and 10^{-34} seconds after the Big Bang. At the end of inflation, a is increased by a factor of at least 10^{27}. After the inflationary epoch, the Universe continued to expand, but at a slower rate (see Figure 2.5). Inflation model was developed in the 1980's by Alexei Starobinsky, Alan Guth and Andrei Linde.

During this period of rapid exponential expansion, quantum fluctuations are magnified to cosmic size, which later form the seeds for the large scale structure in the Universe. This rapid expansion offers the answer as to why opposite ends of the Universe have the same temperature (the horizon problem). It also provides explanation to the flatness and magnetic monopole problems.

Inflation is currently an active field of research, both in cosmology and particle physics. The details on the mechanism of inflation are still unknown. The general hypothesis is that inflation is driven by a phase transition of a scalar field. A phase transition corresponds to a sudden change in the properties of a physical system, and the phase transition of a hypothetical scalar called the "inflaton" is conjectured to have driven inflation in the early Universe. There are currently a very large number of different models of inflation, and observational cosmology will provide evidence to narrow this down to a few favored models.

Inflation was needed to solved the three problems (horizon, flatness, monopole). The ΛCDM model, now the Standard Model of Cosmology, was developed after the idea of inflation was introduced, and it lives within the inflationary paradigm. When inflation was proposed, the Universe was not known to be accelerating and it was not known that the dark matter couldn't be all baryonic. It only became clear later that the inflationary Universe had to be dominated by Λ and cold dark matter.

The ΛCDM is the simplest model that currently offers the best explanation for the observed properties of our current Universe. It is a parametrization of the Big Bang cosmological model in which the Universe contains a cosmological constant Λ, corresponding to dark energy, and cold dark matter (CDM). The ΛCDM model correctly predicts the existence and structure of the cosmic microwave background [45], the large scale structure of galaxies and galaxy clusters [43], the abundances of light elements [11], and the accelerating expansion of the Universe [65].

The nature of both the dark energy and the dark matter are still unclear. Measurements of the CMB and of the LSS with intensity mapping are meant to shed light on these and other cosmological mysteries.

2.2 The Cosmic Microwave Background

The Cosmic Microwave Background (CMB) is the blackbody radiation left over from the Recombination Era in Big Bang cosmology. Before the creation of the CMB, the Universe was a hot, dense and opaque plasma in which photons could not travel freely. During the Recombination Era, the Universe cooled down to a temperature of 2700 K, cool enough to allow electrons and protons to form hydrogen atoms. Due to the expansion of the Universe, the CMB is redshifted and hence its blackbody spectrum now peaks at 2.72 K [134].

We can measure three characteristics of the CMB:

1. The frequency spectrum $f(\nu)$.

2. The temperature $T(\hat{\mathbf{n}})$, where $\hat{\mathbf{n}}$ is the direction in the sky, $\hat{\mathbf{n}} = (\theta, \phi)$.

3. The polarization (Stokes) parameters.

2.2.1 Redshifting of the CMB

At the end of the Recombination Era, the temperature of the CMB is estimated to be 3700 K. Because of the expansion of the Universe, its wavelength has been stretched into the microwave regime.

The energy density of radiation per unit frequency interval $u(\nu)$ for black-body radiation is described by the Planck distribution function:

$$u(\nu, T)\mathrm{d}\nu = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/k_B T} - 1} \mathrm{d}\nu \tag{2.68}$$

The radiation is uniform in all directions and propagates at the speed of light, hence the spectral radiance of a blackbody in thermal equilibrium is

$$B(\nu, T) = \frac{u(\nu, T)c}{4\pi} = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/k_B T} - 1}. \tag{2.69}$$

The CMB spectrum was accurately measured by the FIRAS instrument on COBE, and it follows the Planck spectrum to very high precision, as shown in Figure 2.7. The peak temperature of the CMB, determined from its spectral radiance, is $T_{\mathrm{CMB}} =$

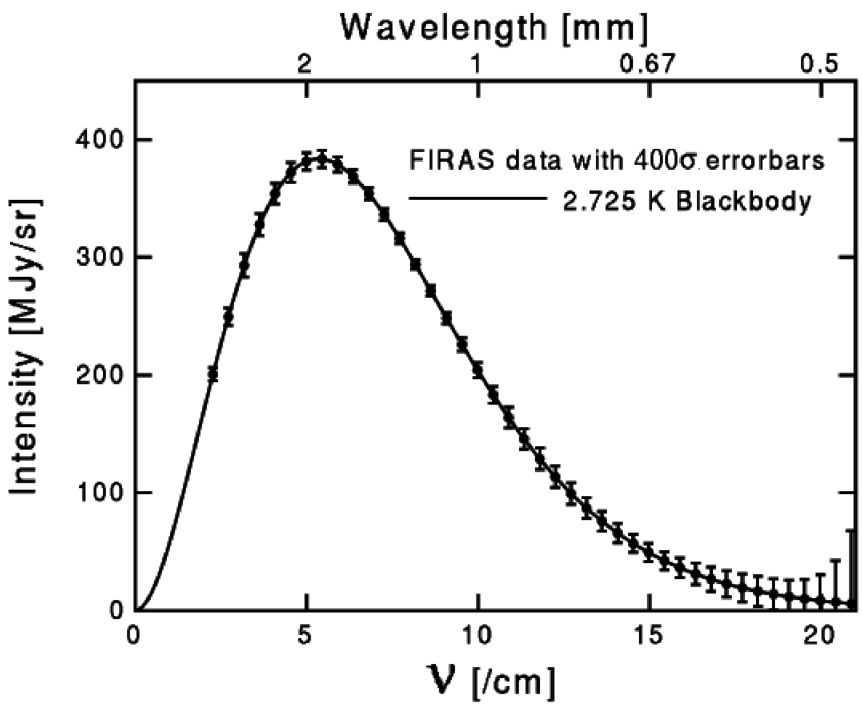

Figure 2.7: CMB spectrum measured by the FIRAS instrument on COBE is plotted along with the theoretical curve for a blackbody in thermal equilibrium at 2.725K [55]. The error bars are exaggerated by 400 times.

2.725 ± 0.001 K. The CMB temperature across the sky is extremely uniform, within a few parts in 10^5. This implies that the observable Universe was in thermal equilibrium when the CMB was created.

The Planck distribution can be written in terms of the photon number density in a frequency interval dν:

$$n(\nu, T)\mathrm{d}\nu = \frac{u(v, T)\mathrm{d}\nu}{h\nu} = \frac{8\pi\nu^2\mathrm{d}\nu}{c^3}\frac{1}{e^{h\nu/k_BT} - 1}. \tag{2.70}$$

Now we can relate the temperature at time t_1 to a temperature at redshift t_2. Suppose at time t_1 the Universe is populated with a blackbody radiation of temperature T_1. Then a volume V_1 contains

$$\mathrm{d}N_1 = V_1\frac{8\pi\nu_1^2\mathrm{d}\nu_1}{c^3}\frac{1}{e^{h\nu_1/k_BT_1} - 1} \tag{2.71}$$

photons in the frequency interval between ν_1 and $\nu_1 + \mathrm{d}\nu_1$. At time $t_2 > t_1$, these photons' wavelengths are stretched by the expansion of the Universe; if we let $r \equiv a(t_1)/a(t_2)$, then $\nu_2 = r\nu_1$ and these photons occupy a new frequency interval

$$(\nu_2, \nu_2 + \mathrm{d}\nu_2) \iff r(\nu_1, \nu_1 + \mathrm{d}\nu_1), \qquad r = \frac{a(t_1)}{a(t_2)} < 1 \text{ if } t_2 > t_1. \tag{2.72}$$

The photons will occupy a new volume $V_2 = V_1/r^3$.

$$\begin{aligned} \mathrm{d}N_1 &= V_1 \frac{8\pi\nu_1^2 \mathrm{d}\nu_1}{c^3} \frac{1}{e^{h\nu_1/k_B T_1} - 1} \\ &= \frac{V_1}{r^3} \frac{8\pi r^3 \nu_1^2 \mathrm{d}\nu_1}{c^3} \frac{1}{e^{hr\nu_1/k_B T_1} - 1} \\ &= V_2 \frac{8\pi\nu_2^2 \mathrm{d}\nu_2}{c^3} \frac{1}{e^{h\nu_2/k_B T_2} - 1} \\ &= \mathrm{d}N_2 \end{aligned} \tag{2.73}$$

We made the assumption $T_2 = rT_1$ so that the number of photons is conserved: $\mathrm{d}N_1 = \mathrm{d}N_2$. This is true for any frequency interval. Therefore, the radiation spectrum at time t_2 is also a black-body spectrum with temperature $T_2 = T_1 a(t_1)/a(t_2)$. By definition of redshift, we can relate the present day temperature to a temperature at a redshift z with the equation

$$T_0 = \frac{T(z)}{1+z}. \tag{2.74}$$

We can use this equation to estimate the temperature of the CMB at the time it was created. The redshift factor of the surface of last scattering is $z_{LS} = 1100$, which gives a temperature of about 3000 K at the time of last scattering.

2.2.2 Temperature anisotropies

We now know that redshift gives us information about the temperature at the time of last scattering, but in reality most of the information lies in the CMB's temperature field. There are small variations in the temperature of the CMB from point to point in the sky. These variations are called anisotropies. They were first detected by the COBE satellite in 1992. The temperature anisotropies captured by the Planck satellite is shown in Figure 2.8 below. The predicted temperature anisotropies are very sensitive

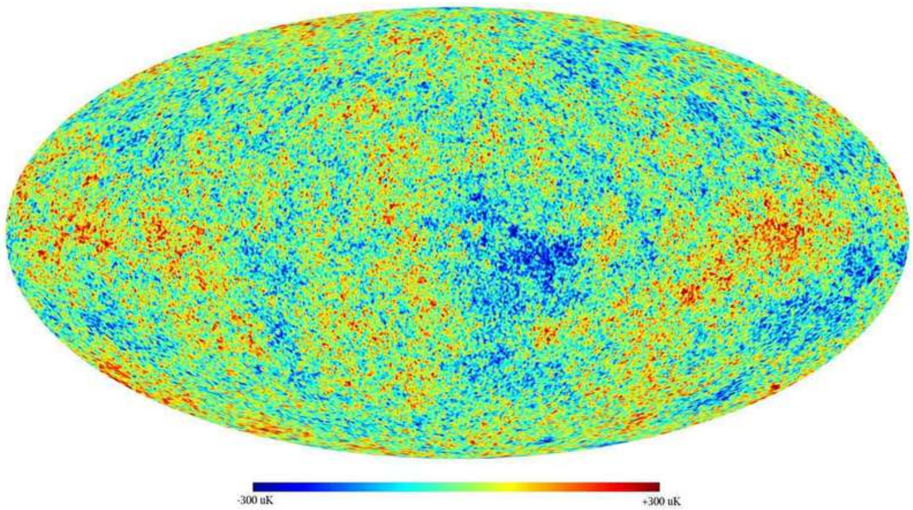

Figure 2.8: The Planck mission's full sky map of the temperature anisotropies in the CMB in the Galactic coordinate system (ESA/Planck Project). The map shows deviations of $\pm 300\,\mu$K from the average temperature of 2.725 K. The contribution from the Galaxy and the dipole anisotropy has been removed.

to a large range of cosmological parameters, so that an accurate measurement of them can provide constraints on cosmological models.

Let's now dive into the mathematical framework. Since we are interested in deviations from the average temperature, we can define a dimensionless quantity:

$$\Theta(\hat{\mathbf{n}}, \mathbf{x}, \eta) = \frac{T(\hat{\mathbf{n}}, \mathbf{x}, \eta) - T(\eta)}{T(\eta)}, \tag{2.75}$$

where $\hat{\mathbf{n}}$ is the direction on the sky, $\mathbf{x}$ is the position of the observer, and η is the conformal time. However, we can only measure $\Theta(\hat{\mathbf{n}}, \mathbf{x}, \eta)$ in some locations $\mathbf{x}_0$ and η_0 (i.e. locations around the Earth *now*). This is the source of cosmic variance, which will be defined later.

We can project the temperature deviation $\Theta(\hat{\mathbf{n}}, \mathbf{x}_0, \eta_0)$ onto a 2-dimensional sky. Mathematically, this means expanding the temperature fluctuations using spherical harmonics $Y_{\ell m}(\theta, \phi)$, which form a complete orthonormal set on the unit sphere. Recall that the spherical harmonics are defined as

$$Y_{\ell m}(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}} P_\ell^m(\cos\theta)e^{im\phi}, \tag{2.76}$$

where ℓ is the multipole, which has range $\ell = 0, \ldots, \infty$. The index ℓ can be thought of as giving the angular scale, with small ℓ corresponding to large angular scale and large ℓ corresponding to small angular scale. For each value of ℓ, there are $2\ell+1$ values of m corresponding to $m = -\ell, \ldots, \ell$. $P_\ell^m(\cos\theta)$ are the associated Legendre polynomials. The expansion of $\Theta(\hat{\mathbf{n}}, \mathbf{x}, \eta)$ is therefore

$$\Theta(\hat{\mathbf{n}}, \mathbf{x}, \eta) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} a_{\ell m}(\mathbf{x}, \eta) Y_{\ell m}(\hat{\mathbf{n}}). \tag{2.77}$$

Equation (2.77) can be inverted to find $a_{\ell m}(\mathbf{x}, \eta)$ by Fourier transform:

$$a_{\ell m}(\mathbf{x}, \eta) = \int\int \Theta(\hat{\mathbf{n}}, \mathbf{k}, \eta)\, \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{(2\pi)^3}\, Y_{\ell m}^*(\hat{\mathbf{n}})\ \mathrm{d}^3k\, \mathrm{d}\Omega. \tag{2.78}$$

In particular, at a location $\mathbf{x}_0$ and conformal time η_0,

$$\Theta(\hat{\mathbf{n}}, \mathbf{x}_0, \eta_0) = \Theta(\theta, \phi) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\int_{\theta'=0}^{\pi}\int_{\phi'=0}^{2\pi}\Theta(\theta', \phi')Y_{\ell m}^*(\theta', \phi')Y_{\ell m}(\theta, \phi)\,\mathrm{d}\Omega' \tag{2.79}$$

and

$$a_{\ell m}(\mathbf{x}_0, \eta_0) = \int_{\theta'=0}^{\pi}\int_{\phi'=0}^{2\pi}\Theta(\theta', \phi')Y_{\ell m}^*(\theta', \phi')\,\mathrm{d}\Omega'. \tag{2.80}$$

We have $\Theta^*(\hat{\mathbf{n}}, \mathbf{x}_0, \eta_0) = \Theta(\hat{\mathbf{n}}, \mathbf{x}_0, \eta_0)$ and $Y_{\ell m}^*(\hat{\mathbf{n}}) = Y_{\ell(-m)}(\hat{\mathbf{n}})$, so $a_{\ell m}^*(\mathbf{x}_0, \eta_0) = a_{\ell(-m)}(\mathbf{x}_0, \eta_0)$.

The coefficient $a_{\ell m}$ tells us about the magnitude of these fluctuations on different angular scales. Since the temperature fluctuation coefficient $a_{\ell m}$ is assumed to be a random Gaussian variable in inflation theory, the mean value of all the $a_{\ell m}$ is zero: $\langle a_{\ell m} \rangle = 0$. Thus, the variance $\langle |a_{\ell m}|^2 \rangle$ measures the deviation of $a_{\ell m}$ from zero and gives the magnitude of the temperature anisotropies.

The monopole term ($\ell = 0$) gives the average temperature of the whole sky (which is at the moment 2.725 K). We can see this immediately because

$$\begin{aligned} \Theta(\theta, \phi)_{\ell=0} &= \int_{\theta'=0}^{\pi} \int_{\phi'=0}^{2\pi} \Theta(\theta', \phi') Y_{00}^*(\theta', \phi') Y_{00}(\theta, \phi) \, \mathrm{d}\Omega' \\ &= \frac{1}{4\pi} \int_{\theta'=0}^{\pi} \int_{\phi'=0}^{2\pi} \Theta(\theta', \phi') \, \mathrm{d}\Omega' \\ &= \langle \Theta(\theta, \phi) \rangle \\ &= 0, \end{aligned} \tag{2.81}$$

where the last line follows from the definition of $\Theta(\hat{\mathbf{n}})$.

The dipole term ($\ell = 1$) tells us about the relative motion of the observer and the CMB photons. CMB photons coming toward the observer will appear blueshifted and those going away from the observer will appear redshifted. By convention, the sum in (2.80) starts at $\ell = 2$ and goes to a given ℓ_{max}, depending on the resolution of the data.

2.2.3 Angular power spectrum

We can define the (theoretical) *angular power spectrum* of these fluctuations as the variance of the harmonic coefficients $a_{\ell m}(\mathbf{x}, \eta)$ as follows

$$\langle a^*_{\ell m} a_{\ell' m'} \rangle = \delta_{\ell \ell'} \delta_{m m'} C_\ell. \tag{2.82}$$

The spatial average is taken over different positions $\mathbf{x}$. Now, the expectation value for the squared deviation from the average temperature is

$$\begin{aligned}
\langle \Theta(\hat{\mathbf{n}}, \mathbf{x}, \eta)\, \Theta^*(\hat{\mathbf{n}}, \mathbf{x}, \eta) \rangle &= \left\langle \sum_{\ell m} a_{\ell m} Y_{\ell m}(\theta, \phi) \sum_{\ell' m'} a_{\ell' m'} Y_{\ell' m'}(\theta, \phi) \right\rangle \\
&= \sum_{\ell \ell'} \sum_{m m'} Y_{\ell m}(\theta, \phi) Y^*_{\ell' m'}(\theta, \phi) \langle a_{\ell m} a^*_{\ell' m'} \rangle \\
&= \sum_{\ell} C_\ell \sum_{m} |Y_{\ell m}(\theta, \phi)|^2 \\
&= \sum_{\ell} \frac{2\ell + 1}{4\pi} C_\ell,
\end{aligned} \tag{2.83}$$

where in the last line we have used the closure relation for spherical harmonics. Therefore, if $(2\ell + 1)C_\ell/4\pi$ is plotted on a linear ℓ scale, or $\ell(2\ell + 1)C_\ell/4\pi$ on a logarithmic ℓ scale, the area under the curve is the temperature variance. However, the convention is to plot the angular power spectrum as $\ell(\ell + 1)C_\ell/2\pi$.

Observed angular spectrum The theoretical angular spectrum in 2.82 can't be measured, because we only have data at a particular location $\mathbf{x}_0$. What is actually observed is the *unbiased estimato*r $\hat{C}_\ell$ of C_ℓ:

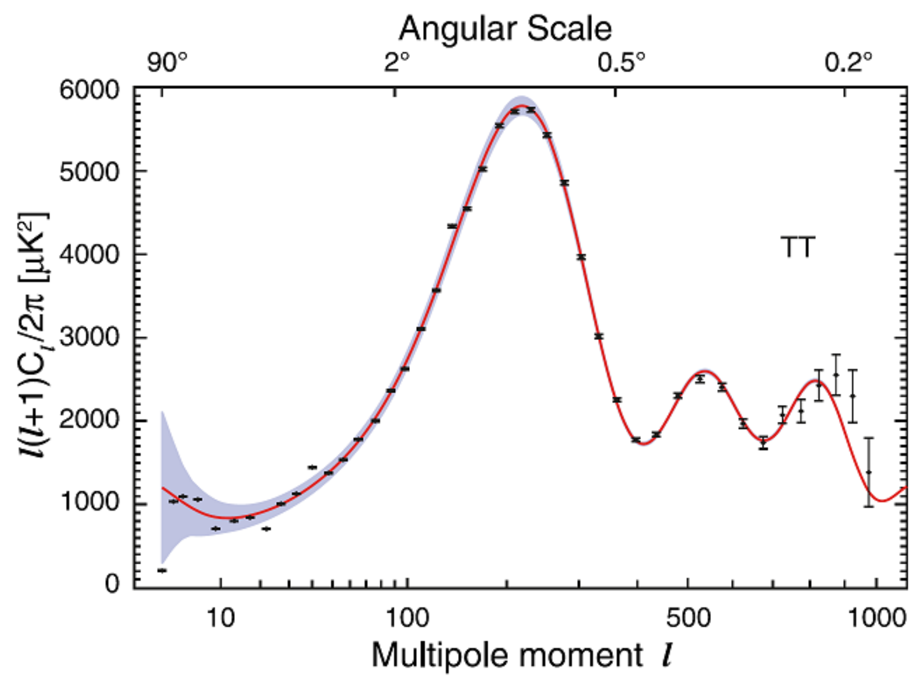

Figure 2.9: The observed angular power spectrum $\hat{C}_\ell$ based on the 5 year WMAP data [132]. The observational results are the data points with error bars. The red curve is the theoretical $C_\ell$from a best-fit model, and the blue band around it represents the cosmic variance corresponding to this C_ℓ.

$$\hat{C}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} a_{\ell m} a^*_{\ell m} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} a_{\ell m} a_{\ell(-m)}. \tag{2.84}$$

The spatial average of $\hat{C}_\ell$ equals C_ℓ:

$$\langle \hat{C}_\ell \rangle = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \langle a_{\ell m} a_{\ell(-m)} \rangle = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} C_\ell = C_\ell. \tag{2.85}$$

We want to know how $\hat{C}_\ell$ compares to the actual C_ℓ, so we want to find the *cosmic variance*, defined as

$$\frac{\Delta C_\ell}{C_\ell} \equiv \frac{\sqrt{\langle (C_\ell - \hat{C}_\ell)^2 \rangle}}{C_\ell} = \sqrt{\frac{2}{2\ell+1}}. \tag{2.86}$$

This is the cosmic variance for idealized full sky coverage. If the observed patch covers a solid angle $\Omega < 4\pi$, the cosmic variance is increased by a factor of $4\pi/\Omega$. For small values of ℓ, such as $\ell = 2$, we only have 5 independent measurements to average. However, cosmic variance is much smaller for higher values of ℓ. In summary, cosmic variance is a fundamental limitation, not an experimental limitation, because we can only observe one Universe at a particular time, at a particular place, with a limited number of m-modes.

2.2.4 Acoustic Oscillations

Before the formation of neutral hydrogen, the early Universe was hot, dense and ionized. Photons and baryonic matter were tightly coupled by Thomson and Coulomb scattering. The acoustic oscillations of the photon-baryon fluid produce the peaks and troughs in the CMB angular power spectrum. The evolution of anisotropies is determined by fluid dynamics. Neglecting the dynamical effects of gravity and baryons, the temperature perturbations in Fourier space satisfy the continuity equation [61] (below) and the Euler equation.

$$\dot{\Theta}(\mathbf{k}) = -\frac{1}{3}kv_\gamma, \tag{2.87}$$

where $\dot{\Theta}(\mathbf{k})$ is the derivative of the monopole or temperature fluctuation $\Theta(\mathbf{k})$ with respect to conformal time $\eta = \int \mathrm{d}t/a(t)$ (the maximum comoving distance a particle could have traveled since $t = 0$), and v_γ is the dipole or the bulk velocity of the photon fluid.

The coupled Euler equation is

$$\dot{v}_\gamma = k\Theta. \tag{2.88}$$

Differentiating $\dot{\Theta}$ and combining it with the equation for $\dot{v}_\gamma$ forms the simple harmonic oscillator equation

$$\ddot{\Theta} + \frac{k^2}{3}\Theta = 0. \tag{2.89}$$

For photon domination (dynamically baryon free), the speed of sound in the fluid squared is $c_s^2 = \dot{p}/\dot{\rho} = 1/3$, so the oscillation equation becomes

$$\ddot{\Theta} + c_s^2 k^2 \Theta = 0.$$

The solution is

$$\Theta(\eta_*) = \Theta(0)\cos(ks), \tag{2.90}$$

where the asterisks denote evaluation at recombination and $s = \int_0^\eta c_s \mathrm{d}\eta'$ is the distance sound can travel in η, or the sound horizon. The effective temperature accounts for the redshift a photon experiences when climbing out of the gravitational well, also known as the Sachs-Wolfe effect. In the limit of large scales compared with the sound horizon $ks \ll 1$, the perturbation has very little difference from the initial conditions. Smaller scale modes oscillate more rapidly, and those that are trapped in a maximum or minimum at recombination result in peaks in the power spectrum. The peaks in Fourier space follow a harmonic relationship $k_n = n\pi/s_*$ where n is an integer. In real space, this corresponds to approximately 1° spacing betweens the peaks.

The initial conditions $\Theta(0)$ of the temperature fluctuations arise within the framework of inflation. Inflationary theory relates scale-invariant curvature fluctuations to the initial temperature fluctuations. It postulates the existence of a scalar field that drove the exponential expansion of the Universe. Quantum fluctuations in the scalar field grew during the inflationary epoch and introduce spatial curvature variations. General relativity also says that the Newtonian potential (a time-time perturbation in the metric) gives a temporal shift of $\delta t/t = \Psi$. The CMB temperature varies as the inverse of the scale factor that depends on time as

$$a \propto t^{\frac{2}{3(1+p/\rho)}}. \tag{2.91}$$

The fractional change in the CMB temperature is

$$\Theta = -\frac{\delta a}{a} = -\frac{2}{3}\left(1+\frac{p}{\rho}\right)^{-1}\frac{\delta t}{t}. \tag{2.92}$$

We see that the initial temperature fluctuations depend directly on the initial gravitational potential perturbations. The temperature perturbation is $-\Psi/2$ in the radiation-dominated ($p = \rho/3$) era and $-2\Psi/3$ in the matter-dominated ($p = 0$) era.

The above treatment neglects the effect of gravity. The Newtonian potential Ψ and the curvature fluctuation Φ change the acoustic oscillation by providing gravitational force on the oscillator. The primary effect of gravity is to make the oscillations a competing force between pressure gradient $k\Theta$ and potential gradient $k\Psi$ with an equilibrium when $\Phi + \Psi = 0$. The oscillator equation and the Euler equation in the previous section become

$$\begin{aligned} \dot{\Theta}(\mathbf{k}) &= -\frac{1}{3}kv_\gamma - \dot{\Phi} \\ \dot{v}_\gamma &= k(\Theta + \Psi). \end{aligned} \tag{2.93}$$

The solution is now

$$[\Theta + \Psi](\eta) = [\Theta + \Psi](\eta_{\rm md}) \cos(ks), \tag{2.94}$$

where $\eta_{\rm md}$ represents the start of the matter dominated epoch. The acoustic oscillation now includes the effect of infall and compression of the fluids into gravitational wells, and it is the effective (observed) temperature $\Theta+\Psi$ that oscillates. After recombination, photons must climb out of the gravitational well and get gravitationally redshifted by $\Delta T/T = \Psi$. Overdense regions in the sky are colder because photons lose energy climbing out of gravitational wells and the opposite is true for underdense region; this is known as the Sachs-Wolfe effect.

If one includes baryons in the model, the addition produces an extra inertia for the photon-baryon fluid, lowering the sound speed and decreasing the sound horizon. The addition of baryons also deepens the gravitational potential wells, leading to an enhanced compressional phase over the rarefactional phase. Since the baryons enhance only the compressional phase, this results in an elevated power in the odd numbered peaks in the power spectrum. The photon-baryon fluid also has slight imperfections caused by shear viscosity and heat conduction in the fluid. The inhomogeneities are damped by an exponential factor of order $e^{-k^2\eta/\dot{\tau}}$ where $\dot{\tau}$ is the differential Thomson optical depth. The damping scale k_d is of order $\sqrt{\dot{\tau}/\eta}$, which is the geometric mean of the horizon and the mean free path. Numerical calculations show that the damping scale is of order $k_d s_* \approx 10$, which suppresses the oscillations beyond the third peak.

The CMB power spectrum is very sensitive to the variations in fundamental cosmological parameters. Increasing the baryon density leads to an enhancement of the odd numbered peaks. An increase in dark matter also increases the total matter content at a fixed baryon density and decreases the overall amplitude of the peaks due to the weakened effect of radiation driving. Spatial curvature and dark energy change the angular distance to recombination and this shifts the peaks left and right and the power of the Sachs-Wolfe plateau. Gravitational waves contribute to large angle anisotropy more than small angle anisotropy and this lowers the relative heights between the peaks. Figure 2.10 shows some variations of the acoustic temperature power spectrum to some fundamental cosmological parameters.

2.2.5 Polarization of the CMB

Even though the acoustic peaks in the temperature power spectrum can reveal a lot of information about the early Universe, cosmological parameters in the temperature power spectrum exhibit degeneracies. Inflation is also not uniquely supported from the temperature power spectrum alone. On the other hand, the polarization of the CMB allows us to discover more information about the early Universe and whether inflation did actually occur.

Any electromagnetic field can be described by the Stokes parameters. For an electromagnetic wave propagating in the z direction with a single frequency ω_0, the components of the wave's electric field vector are

$$E_x = E_{0x} \cos\left[\omega_0 t - \theta_x(t)\right], \qquad E_y = E_{0y} \cos\left[\omega_0 t - \theta_y(t)\right]. \tag{2.95}$$

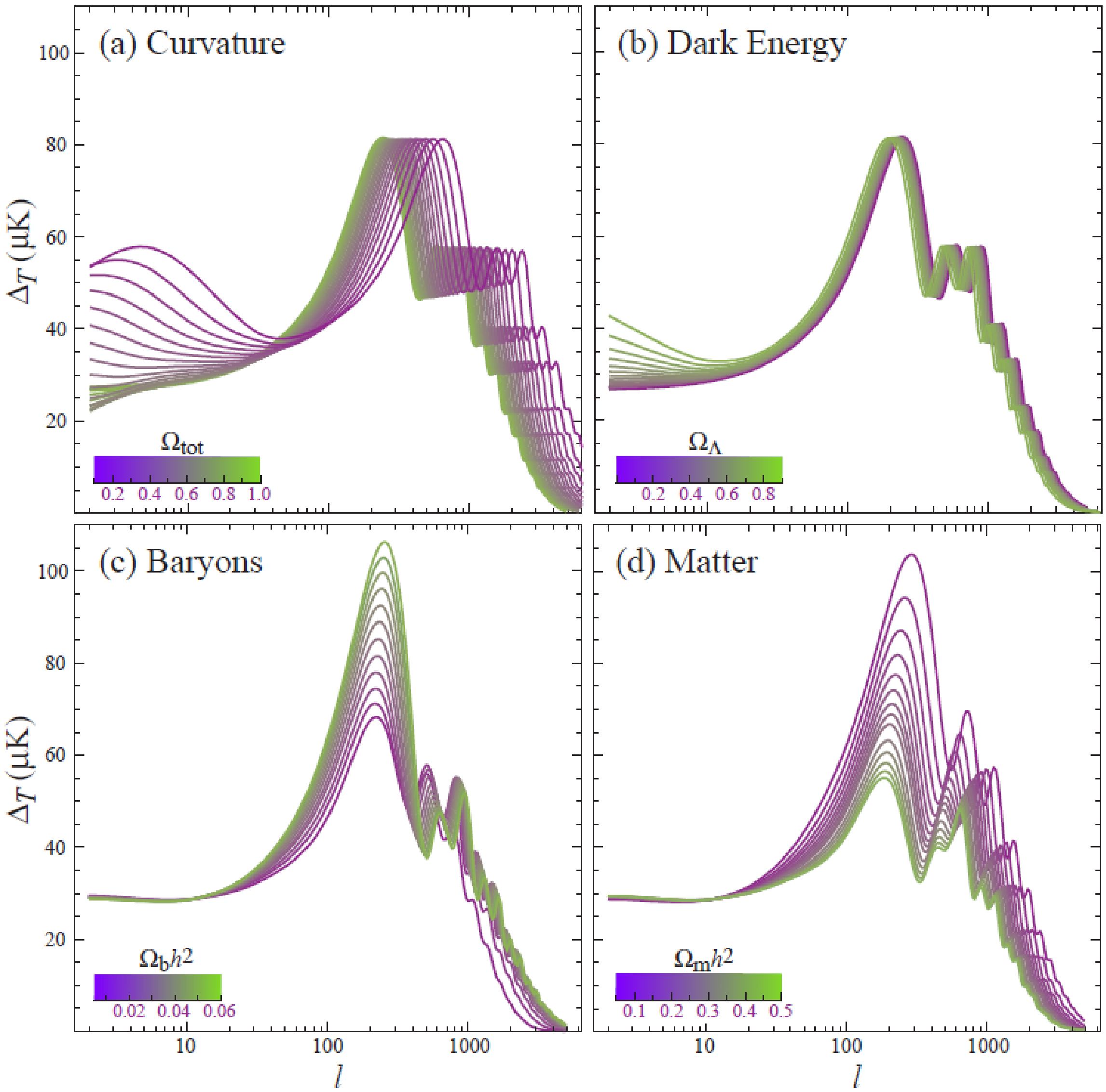

Figure 2.10: Variations of the acoustic temperature power spectrum to four fundamental cosmological parameters (a) the curvature defined by $\Omega_{\rm tot}$, (b) dark energy defined by the cosmological constant $\Omega_\Lambda(w_\Lambda = -1)$, (c) the physical baryon density $\Omega_b h^2$, (d) the physical matter density $\Omega_m h^2$, all varied around a model with $\Omega_{\rm tot} = 1, \Omega_b h^2 = 0.02, \Omega_m h^2 = 0.147$. Figure from [61].

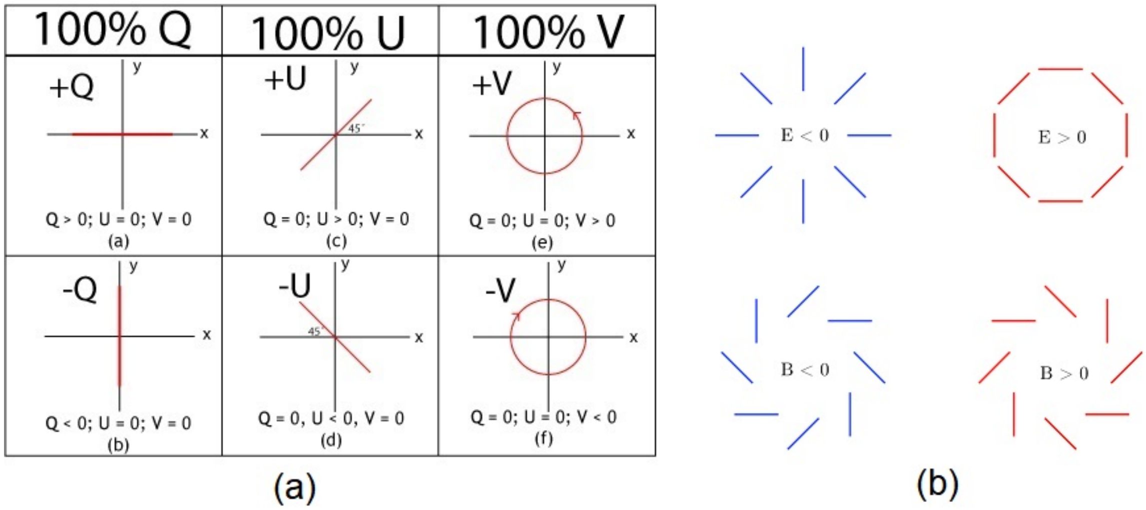

Figure 2.11: (a) Stokes parameters describing the polarization for a wave propagating along the positive z-axis. (b) The Q and U are mathematically decomposed into curl-free E-mode and divergence-free B-mode.

There are four Stokes parameters I, Q, U and V. In the context of the CMB, I is the radiation intensity or temperature, Q and U describe linear polarization, and V describes circular polarization. For unpolarized radiation, $Q = U = V = 0$.

$$I \equiv \langle E_{0x}^2 \rangle + \langle E_{0y}^2 \rangle \tag{2.96}$$

$$Q \equiv \langle E_{0x}^2 \rangle - \langle E_{0y}^2 \rangle$$

$$U \equiv \langle 2E_{0x}E_{0y}\cos(\theta_x - \theta_y) \rangle \tag{2.97}$$

$$V \equiv \langle 2E_{0x}E_{0y}\sin(\theta_x - \theta_y) . \rangle$$

The Stokes parameters I, Q and U are convenient tools to extract polarization data from the CMB. As shown in Figure 2.11 (a), the Q polarizations are orthogonal to the wave vector $\vec{\mathbf{k}}$ and parallel to the $x - y$ axis, while U polarizations are orthogonal to the wave vector $\vec{\mathbf{k}}$ and aligned at an angle of 45° from the $x - y$ axis. Since CMB polarization is generated through Thomson scattering, which results in only linear polarization, the circular polarization V component of the CMB is expected to be zero. However, cosmologists often decompose the CMB polarization into E-mode and B-mode. Note that E-mode has parity symmetry while the B-mode does not. Reflecting the wave vector $\vec{\mathbf{k}}$ over the xy plane does not change the E-modes while the B-modes flip. Accordingly, E-modes are curl-free (no handedness, like the electric field) while B-modes are divergence-free (having handedness like the magnetic field).

From (2.97), we can see that when the coordinate system is rotated by an angle α, the new Stokes parameters can be written as

$$\begin{aligned} Q' &= Q\cos(2\alpha) + U\sin(2\alpha) \\ U' &= -Q\sin(2\alpha) + U\cos(2\alpha). \end{aligned} \tag{2.98}$$

These parameters transform not like a vector but like a two-dimensional, second-rank symmetric trace-free tensor. In spherical polar coordinates (θ, ϕ), the metric tensor is

$$g_{ab} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2(\theta) \end{pmatrix}. \tag{2.99}$$

Following the treatment by Kamionkowski [68], the polarization tensor describes the CMB polarization and is given by

$$\mathcal{P}_{ab}(\hat{\mathbf{n}}) = \frac{1}{2}\begin{pmatrix} Q(\hat{\mathbf{n}}) & -U(\hat{\mathbf{n}})\sin(\theta) \\ -U(\hat{\mathbf{n}})\sin(\theta) & -Q(\hat{\mathbf{n}})\sin^2(\theta) \end{pmatrix}. \tag{2.100}$$

Note that $\mathcal{P}_{ab}(\hat{\mathbf{n}})$ is symmetric and trace ($g^{ab}P_{ab}$) free. The polarization tensor $\mathcal{P}_{ab}(\hat{\mathbf{n}})$ can be decomposed in terms of the spherical harmonics, which form a complete set of orthonormal basis functions on the sphere

$$\frac{\mathcal{P}_{ab}(\hat{\mathbf{n}})}{T_0} = \sum_{\ell=2}^{\infty} \sum_{m=\ell}^{\ell} \left[a^E_{(\ell m)} Y^E_{(\ell m)ab}(\hat{\mathbf{n}}) + a^B_{(\ell m)} Y^B_{(\ell m)ab}(\hat{\mathbf{n}}) \right].$$

The polarization tensor $\mathcal{P}_{ab}(\hat{\mathbf{n}})$ is broken into two orthonormal components $Y^E_{(\ell m)ab}(\hat{\mathbf{n}})$ and $Y^B_{(\ell m)ab}(\hat{\mathbf{n}})$ (which are the E-mode and B-mode). Assuming Gaussianity, the statistics of the CMB temperature/polarization can be fully described by power spectra. Along with temperature, there are three power spectra TT, EE and BB and three cross-power spectra:

$$\begin{aligned}
\langle a^{T*}_{(\ell m)} a^{T}_{(\ell' m')} \rangle &= C^{TT}_\ell \delta_{\ell\ell'} \delta_{mm'} & \langle a^{T*}_{(\ell m)} a^{E}_{(\ell' m')} \rangle &= C^{TE}_\ell \delta_{\ell\ell'} \delta_{mm'} \\
\langle a^{E*}_{(\ell m)} a^{E}_{(\ell' m')} \rangle &= C^{EE}_\ell \delta_{\ell\ell'} \delta_{mm'} & \langle a^{T*}_{(\ell m)} a^{B}_{(\ell' m')} \rangle &= C^{TB}_\ell \delta_{\ell\ell'} \delta_{mm'} \\
\langle a^{B*}_{(\ell m)} a^{B}_{(\ell' m')} \rangle &= C^{BB}_\ell \delta_{\ell\ell'} \delta_{mm'} & \langle a^{E*}_{(\ell m)} a^{B}_{(\ell' m')} \rangle &= C^{EB}_\ell \delta_{\ell\ell'} \delta_{mm'}.
\end{aligned} \tag{2.101}$$

The scalar spherical harmonics $Y_{(\ell m)}$ and the E tensor harmonics $Y^E_{(\ell m)}$ have parity $(-1)^\ell$, but the B tensor harmonics $Y^B_{(\ell m)}$ have parity $(-1)^{\ell+1}$. Since $(-1)^{\ell+\ell'+1} = (-1)$ when $\ell = \ell'$, symmetry under parity tranformations requires that $C^{TB}_\ell = C^{EB}_\ell = 0$.

2.2.6 Sources of CMB polarization

CMB polarization is created from Thomson scattering of electrons with surrounding photons possessing quadrupolar variation in intensity or temperature at the time of last scattering, as shown in Figure 2.13. On the other hand, a free electron surrounded by photons having an isotropic or dipole variation in temperatre will scatter photons with no net polarization. If Thomson scattering is rapid, the random scatterings would destroy quadrupole anisotropy and polarization. An understanding of the quadupolar

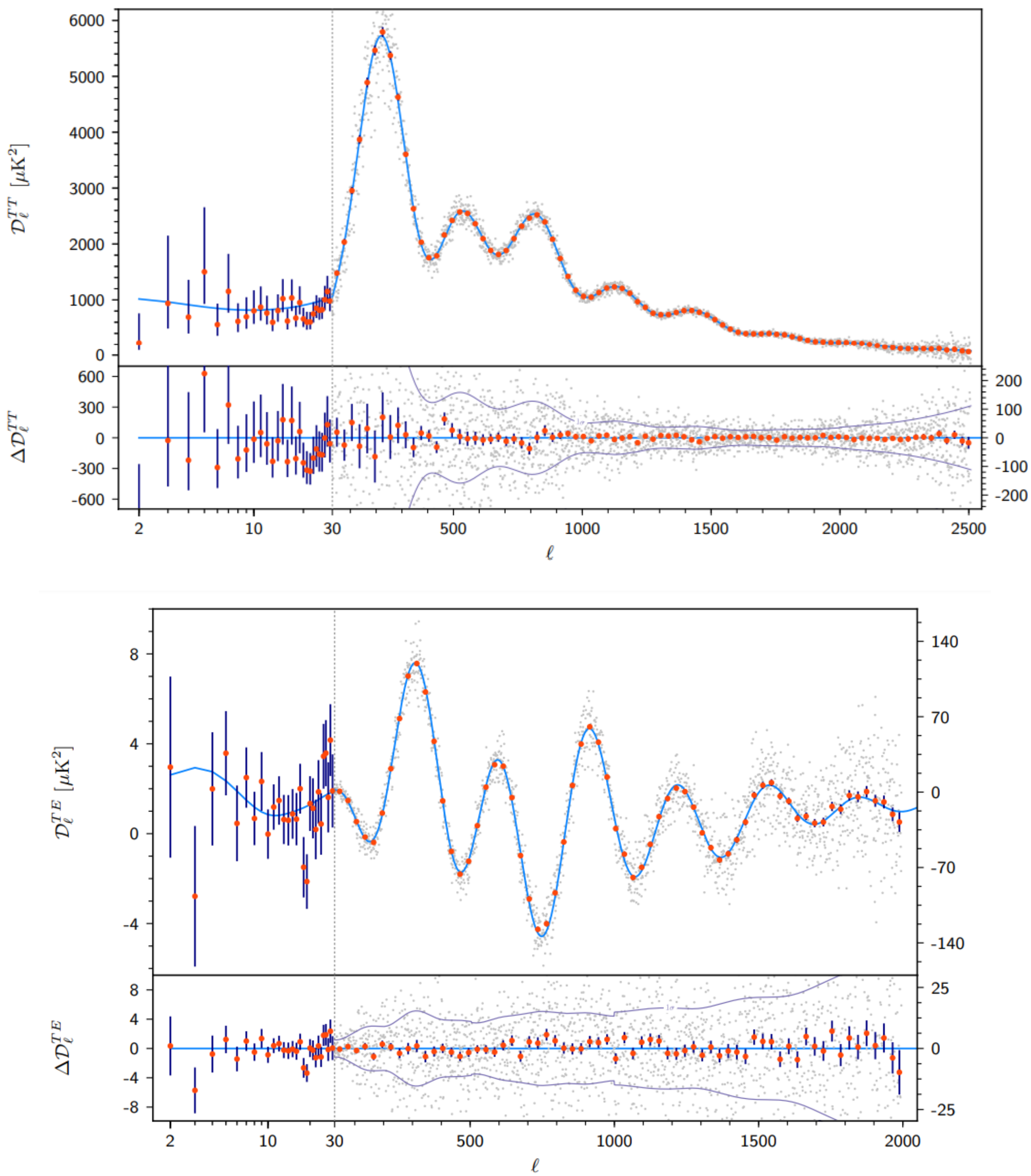

Figure 2.12: Planck 2018 CMB's TE, and TE power spectra [6], where $\mathcal{D}_\ell = \ell(\ell+1)C_\ell/2\pi$. The light blue line in each panel represents the ΛCDM model theoretical best fit. The residuals with respect to the fitted lines are shown in the smaller rectangular box in each panel. The horizontal axis switches from logarithmic to linear at $\ell = 30$ (grey vertical line in each panel). The error bars show $\pm 1\sigma$ diagonal uncertainties, including cosmic variance.

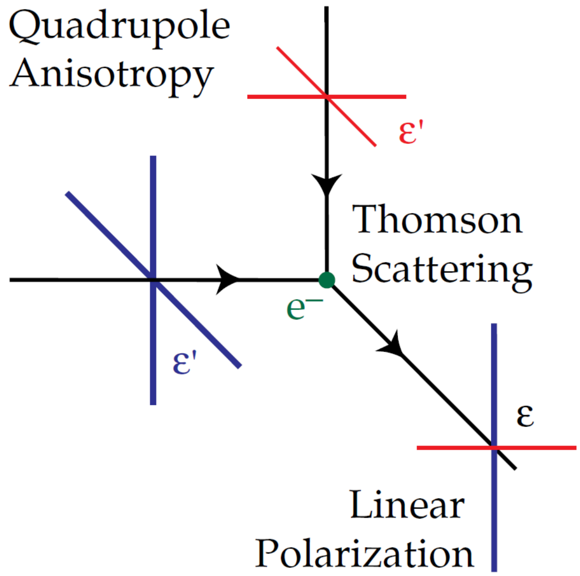

Figure 2.13: A depiction of Thomson scattering of radiation with a quadrupole anisotropy generating linear polarization. Hot radiation (blue) and cold radiation (red) combine to produce net polarization. Figure from [63].

temperature fluctuations at the last scattering surface ultimately leads to an understanding of the polarization of the CMB.

Sources of quadrupoles

There are three main types of perturbations that gives rise to quadrupole anisotropies and thus the CMB: scalar (compressional), vector (vortical), and tensor (gravitational wave) perturbation.

- The scalar perturbation is the most common type of perturbation. They represent perturbations in the energy density of the cosmological fluids at the time of last scattering and are the only fluctuations that can form structure through gravitational instability. The quadrupolar moment in the photon temperature distribution is produced when there are flows from locally underdense regions

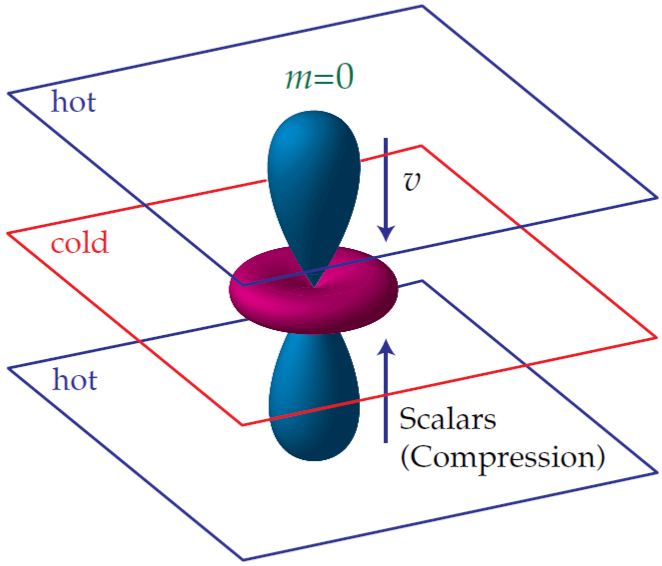

Figure 2.14: Flow from hot region into the cold region resultng in a scalar quadrupole moment that is azimuthally symmetric $Y_{\ell=2}^{m=0}$. Figure from [63].

(hot or crest) to locally overdense regions (cold or trough) (see Figure 2.14). An observer in the trough would see hotter photons from the crests flow into the trough from the $\pm\hat{\mathbf{k}}$ directions and colder photons surround them, resulting in a quadrupole pattern $Y_{\ell=2}^{m=0}$ seen in the trough. An observer in the crest would see the velocity and the sign of the quadrupole reversed. In Figure 2.15, Thomson scattering together with quadrupole anisotropies caused by scalar perturbations result in an E-mode polarization pattern. The strength and shape of the E-mode power spectrum can be predicted from the Euler equation (2.88) and the monopole or temperature oscillation equation (2.90). Since the dipole (or bulk velocity) term v_γ is proportional to the first derivative of the monopole (temperature) fluctuation Θ, the dipole and the monopole terms are $\pi/2$ out of phase but of the same order of magnitude at recombination. However, the quadrupole moment that generates scalar mode polarization is of order $kv_\gamma/\dot{\tau}$,

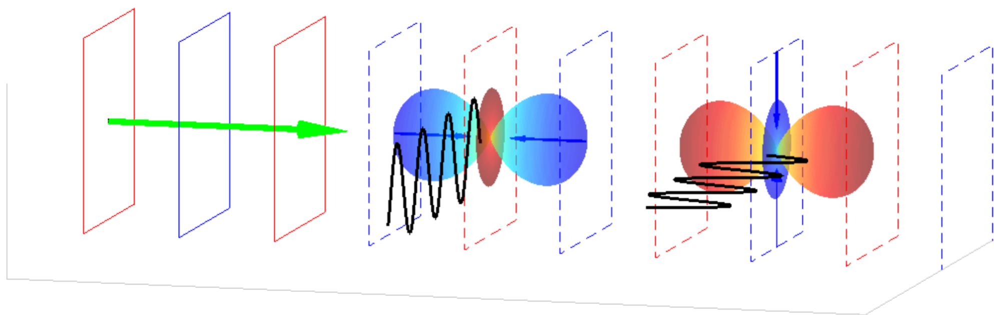

Figure 2.15: E-mode polarization resulting from a single Fourier mode of scalar perturbation. The wave vector is depicted in green. The scattered light is polarized at an angle 90° to the incoming wave vector. Figure from [101].

and the polarization spectrum is smaller than the temperature spectrum by a factor of order $k/\dot{\tau} \approx 10$ at recombination [61]. Thus, the EE power spectrum resulting from quadrupole moment is out of phase with the temperature anisotropy (Θ) and its amplitude is down from the temperature power spectrum by a factor of ten. The TE cross-power spectrum shows oscillations at twice the frequency of either the TT or EE power spectrum.

- Vector perturbations represent vortical motions of matter. The vorticity is damped as the Universe expands and thus vector perturbations are ignored in many standard cosmological models.

- Tensor perturbations are transverse-traceless perturbations to the metric, which can be considered as gravitational waves. A plane gravitational wave perturbations represents a quadrupolar stretching of space in plane of perturbation. The passage of gravitational waves through density fluctutations results in polarization with both E-mode and B-mode components, as shown in Figure 2.16. Primordial gravitational waves are the only expected source of B-mode polarization from the

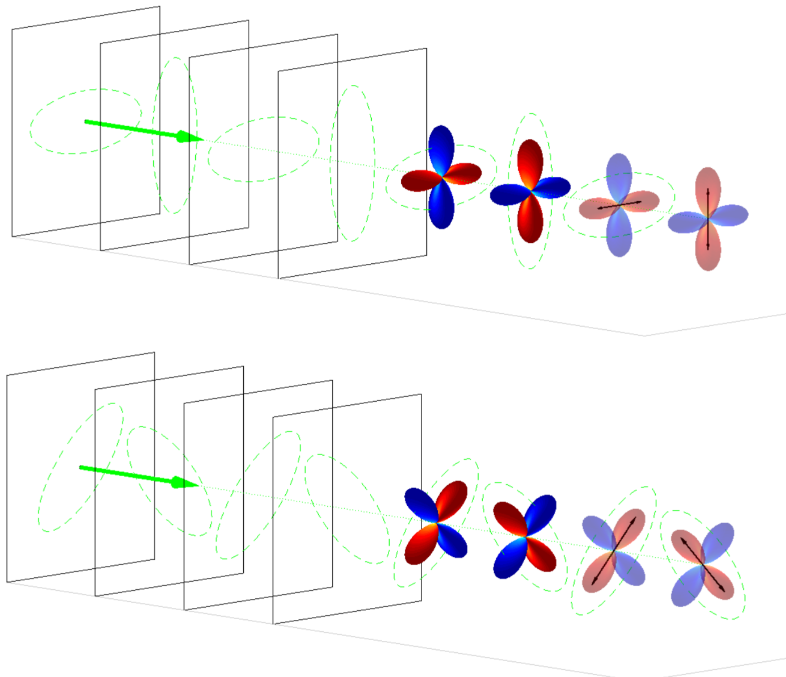

Figure 2.16: E-mode and B-mode polarization resulting from a single Fourier mode of a gravitational wave. The wave vector is depicted in green. A plus polarized gravitational wave h_+ (top) produces E-mode polarization. A cross polarized gravitational wave $h_\times$ (bottom) produces B-mode polarization. Figure from [101].

time of recombination. Many models predict that primordial gravitational waves were produced at the time of inflation, and hence the detection of primordial B-mode polarization would be direct evidence of the theory of inflation. It would also help to distinguish different models of inflation and provide understanding of the energy scale at which inflation occurred. The strength of tensor perturbations, or B-mode, is characterized by the tensor-to-scalar ratio r.

Gravitational lensing

In the presence of matter, the path of a light ray can be deflected by gravity. This is known as gravitational lensing. As CMB photons travel to Earth from the

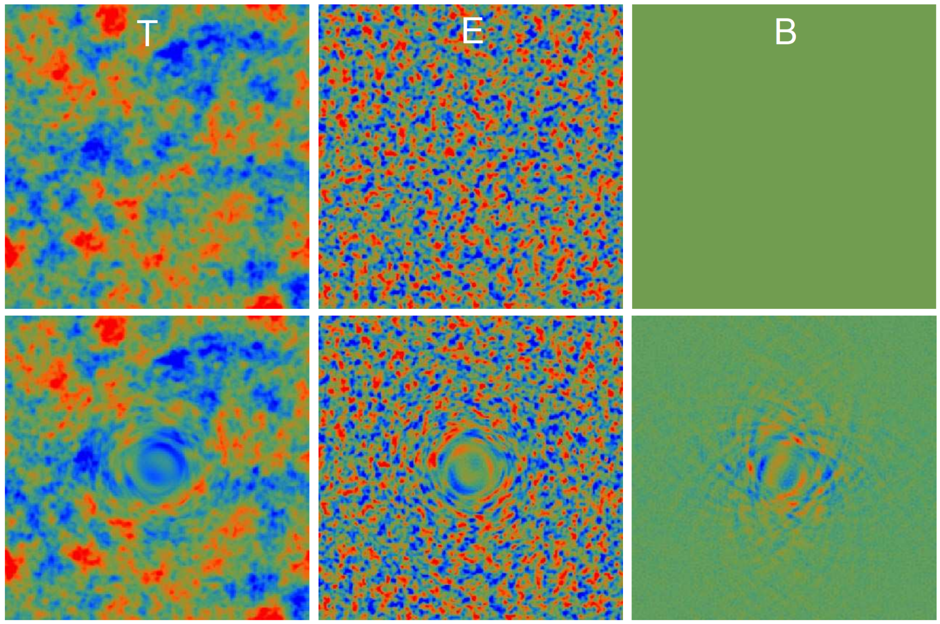

Figure 2.17: The gravitational lensing effect on temperature, E-polarization field and B-polarization field. Figure adapted from [62].

last scattering surface, their paths are deflected by foreground matter which distorts the observed pattern of CMB anisotropies. In addition, gravitational lensing affects the CMB's E and B polarization field much more than temperature. Gravitational potentials along the line of sight can mix the stronger E-mode signal into B-modes (see Figure 2.17). Lensing distortions obscure the information contained in the primordial B-mode polarization signal, and B-mode polarization measurements are expected to be dominated by lensing noise. For small values of the tensor-to-scalar ratio r, the gravitational lensing dominates the primordial B-mode, but fortunately the two signals peak at different angular scales (see Figure 2.18), making it possible to separate the two signals. Removing the lensing-induced noise (delensing) is essential in maximizing the information we can get about inflation.

Even though gravitational lensing contaminates the primordial CMB signals that contain clues about inflation, the information contained in lensing maps is important

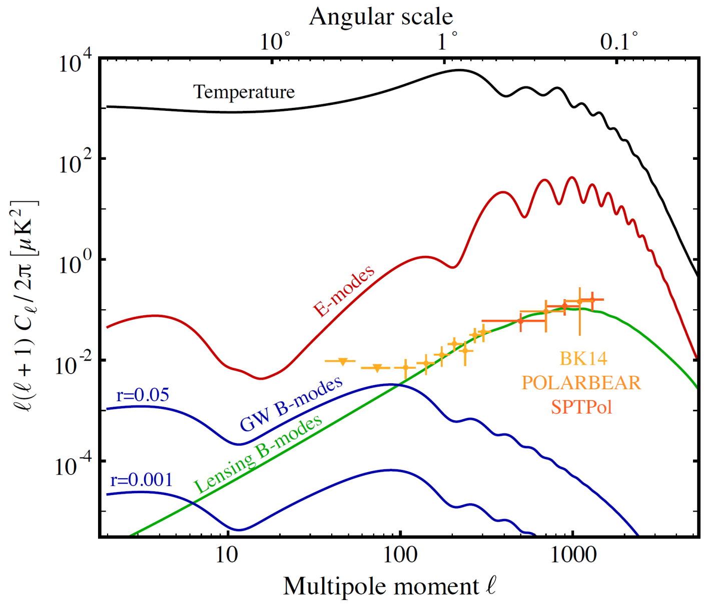

Figure 2.18: Theoretical prediction for the temperature (black), E-mode (red), and tensor B-mode (blue) power spectra. Primordial B-mode power spectra are show for two tensor-to-scalar ratios $r = 0.05$ and $r = 0.001$. The green curve shows the contribution to B-mode from gravitationally lensed E modes. Figure from [1].

on its own. The power spectrum of the lensing detection map is sensitive to physics that governs how structure grows, such as dark energy, modified gravity and the masses of the neutrinos.

Gravitational lensing can be measured because the statistical properties of the primordial (unlensed) CMB are well-known. In addition, B-mode lensing is the only contribution to the B-mode signal at small angular scales; it is non-degenerate. Therefore, the distortions caused by gravitational lensing can be used to make a map of the gravitational potentials in the foregrounds.

2.2.7 The inhomogeneous Universe

In the next chapter, we will introduce a cosmological Boltzmann code (`CMBAns`), which is used to compute the CMB angular power spectra from the anisotropies in the distributions of photons and matter. The name Boltzmann comes from the Boltzmann equation that describes the statistical behavior of a thermodynamic system not in a state of equilibrium. It can be used to compute how physical quantities (energy & momentum) change. So far, we have worked with the flat FRLW metric:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right) \tag{2.102}$$

However, as we know from the CMB, the Universe is not perfectly smooth, even on large scales. We need cosmological perturbation theory to describe a non-smooth Universe. We assume that the matter fluids (baryons, photons) had some small perturbations. For example, the perturbations for density, pressure, and velocity field have the form:

$$\begin{aligned}
\rho(t,\vec{x}) &= \bar{\rho}(t) + \delta\rho(t,\vec{x}) \\
P(t,\vec{x}) &= \bar{P}(t) + \delta P(t,\vec{x}) \\
\vec{v}(t,\vec{x}) &= \overline{\vec{v}}(t) + \delta\vec{v}(t,\vec{x}),
\end{aligned} \tag{2.103}$$

where the bar denotes the average quantity (i.e. $\bar{\rho}(t) \propto 1/a^3$ is the matter density in a perfectly smooth Universe), and $\delta\rho(t,\vec{x})$ denotes the perturbation. We assume higher order perturbations are negligible. Having defined the perturbations, we want to know how these perturbations evolve. A big overdensity will create a deep gravitational well to attract matter until radiation pressure resists it. We can describe this process with the distribution function for each species f_i and the perturbation to the distribution function

$$f_i(t,\mathbf{x},\mathbf{p}) = \overline{f_i}(t,\mathbf{p}) + \delta f_i(t,\mathbf{x},\mathbf{p}). \tag{2.104}$$

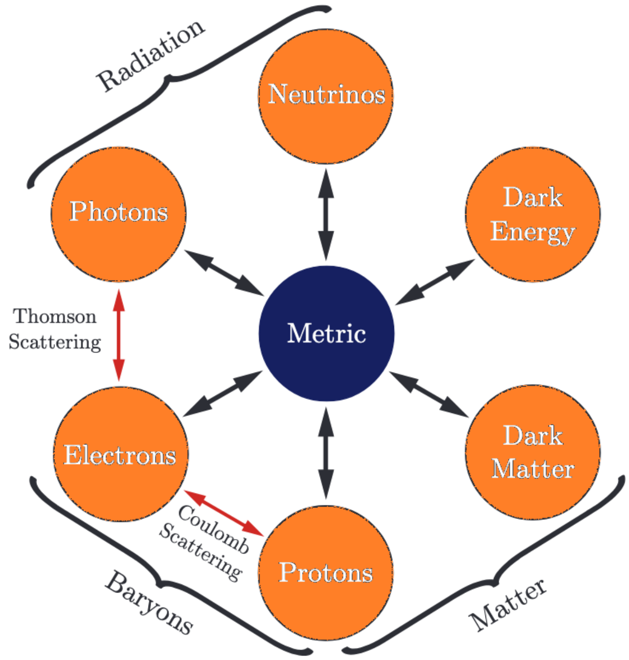

Figure 2.19: The interaction between different components of the Universe. The connections are described by the Boltzmann and Einstein equations. Protons and electrons are tightly coupled through Coulomb scattering and we can treat them as a single component: baryons. We do not consider perturbations to dark energy and it only enters the background metric.

The distribution function describes all the properties of a collection of particles, such as density and pressure. In addition, we need to take into account the interactions between different species, using the Boltzmann equation. Perturbing the energy density and pressure also perturbs the energy-momentum tensor on the right hand side of the Einstein equation, so the metric will also get perturbed. This process is summarized in Figure 2.19.

2.2.8 Metric perturbations and the Boltzmann equation

Since the energy density and pressure is perturbed, there is a corresponding perturbation in the energy-momentum tensor: $T_{\mu\nu} = \bar{T}_{\mu\nu} + \delta T_{\mu\nu}$. The metric also gets perturbed: $g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu}$, where $\bar{g}_{\mu\nu}$ is the unperturbed FLRW metric. We can decompose the perturbed metric into scalar, vector, and tensor components. The Einstein equations for scalars, vectors, and tensors are independent at the first order and can be treated separately. Scalar perturbations are associated with density fluctuations. Vector perturbations are not produced by inflation. Tensor perturbations, or gravitational waves, are an important predictor of inflation. The goal is to derive a set of equations that determines how structures formed in the early Universe. The starting point is the Boltzmann equation, which describes how the distribution function of some particle species evolves with time:

$$\frac{\mathrm{d}f}{\mathrm{d}\lambda} = C[f], \tag{2.105}$$

where λ is an affine parameter along a trajectory. The term on the right hand side is the collision term, which describes how particles are moved from one phase-space element to another. Collisions include scattering, pair creation, annihilation, and particle decay.

We perturb the distribution function f_i for each species (baryons, radiation, and cold dark matter). For each species, we evaluate the Boltzmann equation $\frac{\mathrm{d}f}{\mathrm{d}\lambda} = C[f]$ where the right-hand term describes the interaction between different species. Two major interactions are Compton scattering $e^- + \gamma \rightleftharpoons e^- + \gamma$ and Coulomb scattering $e^- + p^+ \rightleftharpoons e^- + p^+$. The perturbation in each species' density and pressure will give us a perturbed energy-momentum tensor $\delta T_{\mu\nu}$, which in turn gives us the perturbed Einstein equation $\delta G_{\mu\nu} = 8\pi G \delta T_{\mu\nu}$. This will return a closed system of coupled differential equations for how the metric perturbation, the baryon density, baryon velocity, CDM density, CDM velocity, etc. evolve. After getting the photon and baryon temperature perturbations, we can expand them in terms of spherical harmonics and calculate the CMB angular spectrum. The mathematical details of this process are described in Ma Bertschinger (1995) [85].

Chapter 3

CMBAns

Most of content of this chapter is based on the paper: "Cosmic Microwave Background Anisotropy numerical solution (CMBAns) I: An introduction to C_ℓ calculation" published in JCAP on May 2020. The full version is given here: https://iopscience.iop.org/article/10.1088/1475-7516/2020/05/006 .

This chapter will introduce `CMBAns`, which is a cosmological Boltzmann code for computing the CMB angular power spectra for a flat Universe. It is developed in the `C` language with modularity in mind, from which other researchers can develop their own independent packages without understanding the source code. CMBAns currently does not support calculations of the matter power spectrum that is used in measurements of the LSS by galaxy redshift surveys and intensity mapping. It also does not include secondary anisotropies from CMB lensing. For more detail on the modularity of the code, please refer to Appendix C.

3.1 Introduction to CMBAns

Since the discovery of the Cosmic Microwave Background (CMB) by Penzias and Wilson, the CMB has become an invaluable probe for understanding the physics in the early Universe. Several cosmological theories, proposed in the past, failed to explain the origin of the CMB. Hence, they were rejected as feasible cosmological theories. Others, like Big Bang cosmology, with some assumptions, provide a more complete explanation of the origin of CMB radiation. These models later, after several theoretical modifications, were accepted as the standard cosmological models. Hence, the discovery of the CMB marked the path for the birth of standard cosmology.

The precision of CMB observations has improved over the years. In the past decade, several ground-based and space-based experiments like WMAP, Planck, BICEP, ACT etc. have measured the CMB temperature anisotropies to an exquisite precision. Future experiments like SPT-3G, Simons Observatory and CMB-S4 will provide even better measurements of CMB temperature and polarization [1]. To analyze this influx of data and to test different cosmological models, we also need more accurate Boltzmann codes.

The theory of cosmological perturbations for standard model cosmology was first developed by Lifshitz [81] and later was reviewed by many others [82]. The subsequent research works are summarized in review articles [69, 97], in books and in theses [133, 129]. In electromagnetism, the electric field **E** and the magnetic field **B** can be expressed in terms of a scalar field ϕ and a vector field **A**, which together form a gauge, that have to satisfy certain conditions. The physics of electromagnetism does not depend on a particular choice of gauge (gauge invariance). Similarly, in cosmological perturbation theory, there are many choices of coordinates, or gauges. The physics must be invariant under a general coordinate transformation (gauge transformation). Lifshitz used the synchronous gauge for formulating the linear perturbation theory, since synchronous gauge makes some perturbation variables identically zero. Later, Bardeen and others developed the perturbation theory in the conformal Newtonian gauge due to some complications with the synchronous gauge, such as the appearance of the coordinate singularity [69, 20] etc. The conformal gauge is more frequently used for analytical calculations of the cosmological perturbation equations. However, the synchronous gauges are preferred for the numerical calculations due to the stability issues [94, 60].

The Boltzmann codes have been in use in cosmology for a long time to calculate the CMB angular power spectrum. The first such code provided in the public domain was `COSMICS` [86], written by Ma and Bertschinger. Later, Seljac and Zaldarriaga developed `CMBFAST` [121, 138], in which the line-of-sight integration method was used to make the power spectrum calculation faster. Since then, several packages utilizing Boltzmann codes, such as `CAMB` [74], `CMBEasy` [50], `CLASS` [71, 24, 72, 73], `PyCosmo` [113] etc., have come into existence. In this paper, we describe a new Boltzmann code, called `CMBAns` (Cosmic Microwave Background Anisotropy numerical solution). The package is based on `CMBFAST` and was initially developed in 2010 for a variety of CMB studies [37, 40, 42].

There are three principal motivations behind developing `CMBAns`. First of all, in future CMB missions, the precision of the CMB measurements will improve drastically. Hence, the Boltzmann packages should be able to calculate the CMB power spectrum very accurately up to high multipoles. Secondly, different Markov Chain Monte Carlo (MCMC) packages, such as `CosmoMC` [77, 76], `SCoPE` [41], `AnalyzeThis` [51], etc., which are often used to estimate the cosmological parameters, typically require 10^3 - 10^4 evaluations of Boltzmann codes. Therefore, the Boltzmann code should be able to calculate the CMB power spectrum fast and efficiently. Thirdly, most of the present Boltzmann codes follow a monolithic architecture design and are not modular. Therefore, it is difficult to add any new feature in the package and the functions cannot be used independently. Users cannot write their own packages and use existing functions without an extensive knowledge of the entire source code. To overcome this limitation, `CLASS` code introduced a modular architecture. `CMBAns` is also written in a modular format. It consists of several stand-alone codes, as well as some user-defined functions that users can use to write their codes.

`CMBAns` solves the linear Boltzmann equations for different constituents of the Universe and thereafter uses the line-of-sight integration approach to calculate the source terms and the brightness fluctuations. These are then convolved with the primordial power spectrum to get the CMB angular power spectrum. `CMBAns` can calculate the cosmological power spectrum for different dark energy models (both perturbed and unperturbed). Poulin et al. [110] proposed the existence of *early dark energy* that behaves like a cosmological constant at early times before recombination ($z \gtrsim 3000$), but whose energy density decreases as the Universe expands. This has the effect of increasing the CMB-inferred value of H_0 while leaving the later evolution of the Universe unchanged. This can potentially resolve the Hubble tension. `CMBAns` comes with a MATLAB GUI, where the Hubble parameter of the Universe can be visually modified as a function of redshift. `CMBAns` translates the modified Hubble parameter into the dark energy equation of state (EOS) and then computes the CMB power spectrum for that particular model [42].

In addition, `CMBAns` is capable of calculating the temperature and polarization anisotropies with contribution from massive neutrinos (along with baryons). Neutrinos were produced in large numbers in the high temperatures of the early universe, and they left distinctive imprints in the CMB and on the large-scale structure. Recent neutrino oscillation experiment shows that neutrinos have tiny masses. However, the overall magnitude of the masses is not known and it is currently an active field of research. Determining the neutrinos' mass is one of the key science goals of the CMB-S4 and Simons Observatory [1, 4]. In `CMBAns`, users can specify massive neutrino density and the number of neutrino species.

In this chapter, I will describe the mathematical foundations for the first three chapters of `CMBAns` – the part I was responsible for. I will describe the overall picture of the power spectrum calculation process in the section 3.2. I will also present the conformal time calculations between any two eras in the Universe and then describe different recombination processes, baryon temperature calculation, sound speed, optical depth and visibility. I will show `CMBAns` output power spectra for different initial perturbation conditions. The final section is for the conclusion.

3.2 Summary

The step-by-step big picture of the power spectrum calculation is as follows:

1. Calculate the distance using conformal time from $z = 0$ to $z = 10^8$, a region deep inside the radiation dominated era (See the right panel in Figure 3.1). The initial conditions are also set during this era.

2. Calculate the temperature field as a function of line-of-sight distance. The current monopole temperature of the CMB is 2.7 K with spatial variation of about 50 μK. This temperature will be different at different redshifts. To find it, we need to calculate the baryon temperature at each scale factor. In the beginning, when baryons and radiation are coupled, the baryon and radiation temperatures are the same. After decoupling, the baryon's temperature slowly falls compared to the radiation temperature. The baryons in the Universe were subjected to various sources of heating and cooling. This is done in section 3.4.4.

3. With the baryon temperature solved, we can calculate the ionization fraction as a function of conformal time. We can then calculate the number density of free electrons, which is used to calculate the optical depth and the visibility. The total ionization fraction is given by Equation 3.40.

4. We calculate the visibility along the line of sight. The visibility describes how opaque the gas is to photons. The visibility function at any conformal time τ is defined as $g = \dot{\kappa} \exp(-\kappa)$, where κ is the optical depth. In Figure 3.8, we show the visibility function vs the scale factor.

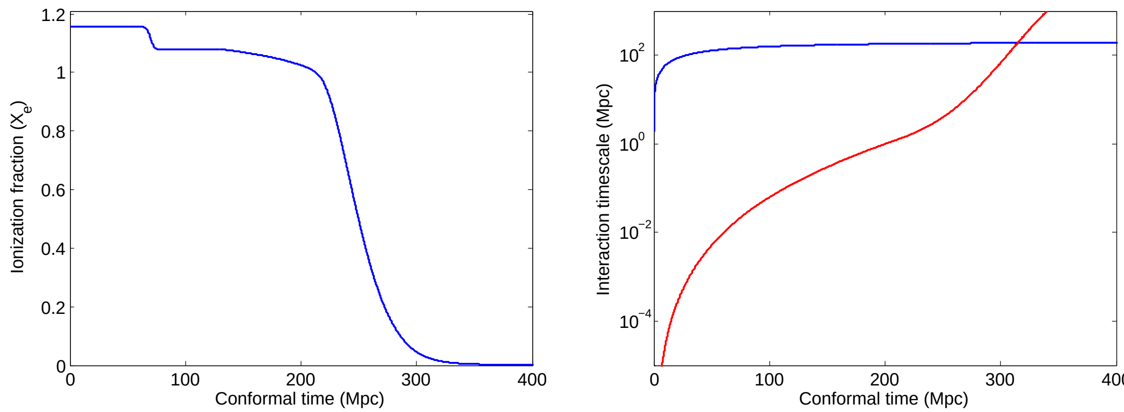

Figure 3.1: Left : The ionization fraction $x_e = n_e/n_H$ from the `recfast` recombination routine (a module inside `CMBAns`). Right : Change of the photon baryon interaction time scale $\left(\tau_c = (an_e\sigma_T)^{-1}\right)$ over time is plotted in red and the time scale at which the modes in the super-Hubble scale evolve ($\tau_H = a/\dot{a}$) is plotted in blue. In the region where $\tau_c \ll \tau_H$, the photons and baryons are tightly coupled to each other. We choose conformal time to be zero at redshift 10^8.

5. Solve the perturbation equations in the scalar and tensor components in the metric, Einstein tensor and stress-energy tensor numerically. We evolve the conservation equations ($\delta T^{\mu\nu}_{;\mu} = 0$) for different components of the Universe, which progress independently except before decoupling, when the baryons and photons evolve together as a single fluid. We consider the density and velocity perturbations (δ and θ) for cold dark matter (CDM), massless and massive neutrinos, photons, baryons, and dark energy.

6. Calculate the source functions $S_T(k, \tau)$ and $S_P(k, \tau)$ for the temperature and polarization. The source terms are the products of the density fluctuation with the visibility and are functions of both the magnitude of k and the conformal time τ. They are used to calculate the rotationally invariant primordial power spectrum $P^s(k)$. The primordial power spectrum for the anisotropic inflation model is taken as

$$P^s(\vec{k}) = P^s(k)\left[\sum_{lm} g_{lm}(k)Y_{lm}(\hat{k})\right] \tag{3.1}$$

where the g_{lm}'s are the spherical harmonic coefficients of g and $g_{00} = 1$. The source terms describe the density fluctuation in directions perpendicular to and parallel to the line of sight.

7. Solve the intensity and polarization fluctuations, Δ_T and Δ_P, using the Boltzmann equations below:

$$\frac{\partial \Delta_T}{\partial \tau} + ik\mu\Delta_T + \frac{2}{3}\dot{h} + \frac{4}{3}(\dot{h} + 6\dot{\eta})P_2(\mu) = \left(\frac{\partial \Delta_T}{\partial \tau}\right)_C \tag{3.2}$$

$$\frac{\partial \Delta_P}{\partial \tau} + ik\mu\Delta_P = \left(\frac{\partial \Delta_P}{\partial \tau}\right)_C, \tag{3.3}$$

where the terms on the right hand side are the collision terms due to Compton scattering. (They correspond to Equation 4.44 and 4.45 in the original paper.) The brightness fluctuation functions for temperature and E mode polarization as a function of ℓ, after solving the Boltzmann equations, are given by

$$\Delta_{Tl}(k) = \int_0^{\tau_0} \mathrm{d}\tau\, S_T(k,\tau)j_l(x)\,, \qquad \Delta_{El}(k) = \sqrt{\frac{(l+2)!}{(l-2)!}}\int_0^{\tau_0} \mathrm{d}\tau\, S_P(k,\tau)j_l(x) \tag{3.4}$$

8. The scalar power spectrum can be calculated as

$$C_l^{XX} = (4\pi)^2 \int k^2\, \mathrm{d}k\, P^s(k)[\Delta^s_{Xl}(k)]^2\,, \qquad C_l^{TE} = (4\pi)^2 \int k^2 \mathrm{d}k\, P^s(k)\,\Delta^s_{Tl}\,\Delta^s_{El}\,, \tag{3.5}$$

where, X can be T or E. The angular power spectrum of the CMB is the convolution of the initial fluctuations, $P(k)$, i.e. the primordial power spectrum set by inflation, and the transition of this fluctuation through different phases of the Universe, which is described by the integral of the source functions.

The power spectrum for the tensor perturbations can be calculated in a similar manner. However, as the tensor fluctuations are spin 2 quantities, we get an extra $\sqrt{\frac{(l-2)!}{(l+2)!}}$ term in the brightness fluctuation functions. The brightness fluctuation functions for the tensor perturbation are

$$\Delta^t_{Tl}(k) = \sqrt{\frac{(l-2)!}{(l+2)!}} \int_0^{\tau_0} \mathrm{d}\tau \, S^t_T(k,\tau) j_l(x) \, , \qquad \Delta^t_{E,Bl}(k) = \int_0^{\tau_0} \mathrm{d}\tau \, S^t_{E,B}(k,\tau) j_l(x) \, , \tag{3.6}$$

The brightness fluctuation functions can be convolved with the primordial tensor power spectrum to get

$$C_l^{tXX} = (4\pi)^2 \int k^2 \, \mathrm{d}k \, P^t(k) [\Delta^t_{Xl}(k)]^2 \, , \qquad C_l^{tXY} = (4\pi)^2 \int k^2 \mathrm{d}k \, P^t(k) \, \Delta^t_{Xl} \, \Delta^t_{Yl} \, , \tag{3.7}$$

where $(X,Y) \in (T,E,B)$.

3.3 Conformal time calculation

In cosmology, the redshift z is often used for measuring time. However, for numerical calculation of the perturbation equations, line-of-sight integration, etc., the conformal time plays an important role. It is straight-forward to calculate the conformal time under the assumption of a matter-dominated or dark energy-dominated Universe.

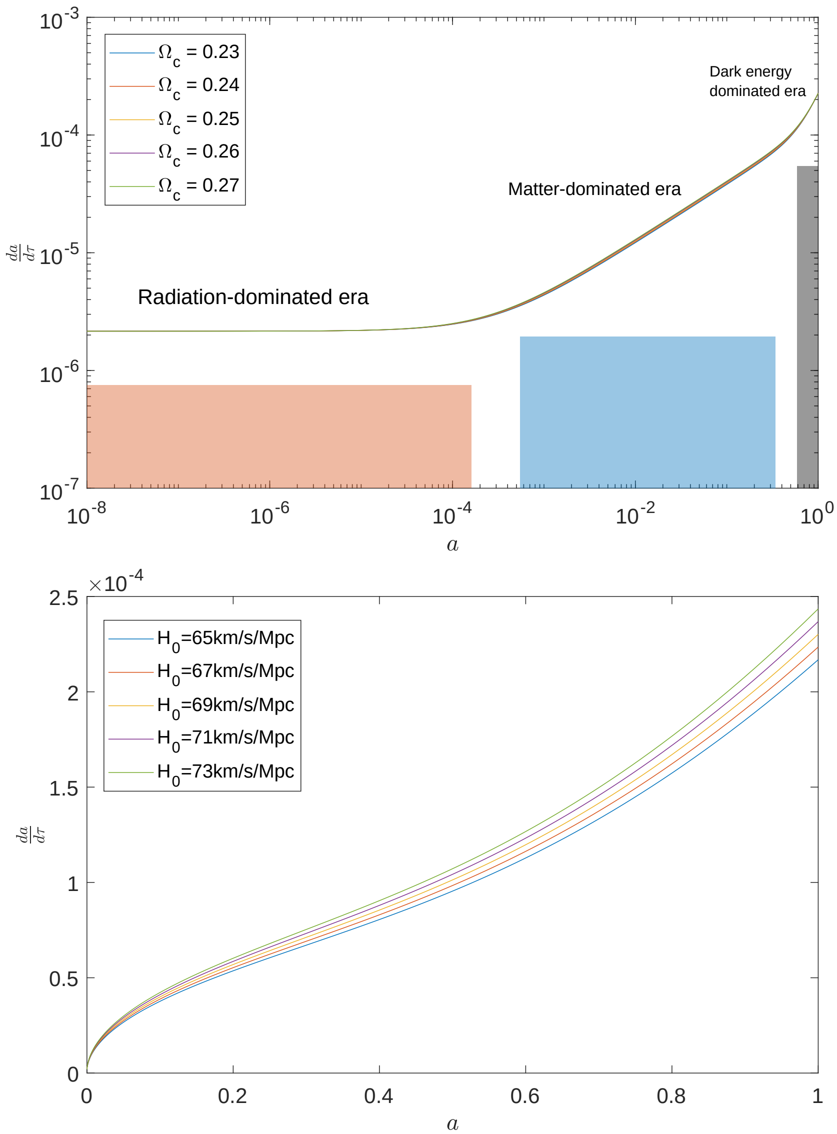

Figure 3.2: Plot of $\frac{\mathrm{d}a}{\mathrm{d}\tau}$ for different Ω_c (top) and H_0 (bottom). The radiation dominated, matter dominated, and the dark energy dominated eras are clearly shown in the top plot. As shown in Eq. 3.10, $\frac{\mathrm{d}a}{\mathrm{d}\tau}$ is constant in the radiation dominated era (orange), varies as $a^{\frac{1}{2}}$ in the matter dominated era (blue) and varies as a^2 in the dark energy dominated era (gray).

However, in the presence of all the components of the Universe, the calculations can be complicated and an analytical solution may not exist.

We denote the conformal time as τ. The Hubble parameter $H(\tau)$ is defined as

$$H^2(\tau) = \left(\frac{1}{a^2}\frac{\mathrm{d}a}{\mathrm{d}\tau}\right)^2 , \tag{3.8}$$

where a is the scale factor. From the FLRW equation, we can write the Hubble parameter as

$$\frac{H(\tau)^2}{H_0^2} = \Omega_{0,m}a^{-3} + \Omega_{0,\gamma}a^{-4} + \Omega_{0,\nu}a^{-4} + \Omega_{\nu_m} + \Omega_d\,. \tag{3.9}$$

The above two equations give

$$\frac{\mathrm{d}a}{\mathrm{d}\tau} = \sqrt{a^4 H_0^2\left(\Omega_{0,m}a^{-3} + \Omega_{0,\gamma}a^{-4} + \Omega_{0,\nu}a^{-4} + \Omega_{\nu_m} + \Omega_d\right)}\,, \tag{3.10}$$

where $\Omega_{0,m}$, $\Omega_{0,\gamma}$, and $\Omega_{0\nu}$ are the density parameters for present-day matter (which includes both cold dark matter and baryonic mater), photons, and massless neutrinos, respectively. The density parameter of massive neutrinos and dark energy at a scale factor a are Ω_{ν_m} and Ω_d, respectively. The density parameters are defined as the ratios of the respective densities over the critical density:

$$\Omega_{0,m} = \frac{\rho_{0,m}}{\rho_{cr}}, \qquad \Omega_{0,\gamma} = \frac{\rho_{0,\gamma}}{\rho_{cr}}, \qquad \Omega_{0,\nu} = \frac{\rho_{0,\nu}}{\rho_{cr}}, \qquad \Omega_{\nu_m} = \frac{\rho_{\nu_m}}{\rho_{cr}} \qquad \Omega_d = \frac{\rho_d}{\rho_{cr}}$$

where the critical density ρ_{cr} is given by $\rho_{cr} = \dfrac{3H_0^2}{8\pi G}$. The densities of matter and radiation at any era are scaled as a^{-3}, a^{-4}, respectively, with their densities at the present era. For calculating the density of the massive neutrinos, we need to use the Fermi-Dirac statistics. For a Λ dark energy model, the density of the dark energy will be constant. However, for any other dark energy model, we need to calculate the density variation from its equation of state (eos). Figure 3.2 shows the plot of $\frac{da}{d\tau}$ for different values Ω_c (left) and H_0 as calculated by `CMBAns`.

3.3.1 Matter density

The first term in Eq. 3.10 can be calculated by evaluating $aH_0^2\Omega_{0,m}$. As the CDM and baryon density parameters, $\Omega_{0,c}$ and $\Omega_{0,b}$ are the input parameters, we can calculate $\Omega_{0,m} = \Omega_{0,c}+\Omega_{0,b}$. H_0 is also an input parameter, but its unit is km/s/Mpc. In `CMBAns`, we use Mpc as the unit for both the spatial and temporal dimensions. In order to convert the Hubble parameter in Mpc^{-1}, we multiply H_0 with $1/c^2 = 1.11265 \times 10^{-11}$ $(\mathrm{km/s})^{-2}$.

3.3.2 Radiation density for photons

The radiation density consists of two components: photon density ρ_γ and the massless (relativistic) neutrino density ρ_ν. The photon number density as a function of frequency can be derived from the Planck radiation law:

$$n_\gamma(\nu)\,\mathrm{d}\nu = \frac{8\pi\nu^2\,\mathrm{d}\nu}{e^{h\nu/k_B T_0} - 1}\,, \tag{3.11}$$

where k_B is the Boltzmann constant, h is the Planck constant, and T_0 is the current CMB temperature. The photon energy density can be calculated as

$$\rho_{0,\gamma}c^2 = \int_0^\infty h\nu n_\gamma(\nu)\,\mathrm{d}\nu = a_B T_0^4\,, \tag{3.12}$$

where $a_B = \frac{8\pi^5 k_B^4}{15h^3c^3} = 7.56577 \times 10^{-16}\ \ \mathtt{Jm}^{-3}\mathtt{K}^{-4}$ is the radiation constant. We also know that

$$\rho_{cr} = \frac{3H_0^2}{8\pi G} = 1.87847 \times 10^{-30}\, H_0^2\ \ \mathtt{kg\,m}^{-3}(\mathtt{km/sec/Mpc})^{-2}. \tag{3.13}$$

Therefore, the second term in Eq. 3.10 can be calculated by evaluating $H_0^2\Omega_{0,\gamma}$ as follows:

$$\Omega_{0,\gamma}H_0^2 = \frac{\rho_{0,\gamma}}{\rho_{cr}}H_0^2 = \frac{a_B}{c^2\rho_{cr}}T_0^4 = 4.98613 \times 10^{-14} \times T_0^4\ \mathtt{Mpc}^{-2}\ . \tag{3.14}$$

3.3.3 Radiation density for massless neutrinos

Massless neutrinos follow Fermi-Dirac statistics with neutrino temperature T_ν. The distribution function is given by

$$n_\nu(\nu)\,\mathrm{d}\nu = \frac{8\pi\nu^2\,\mathrm{d}\nu}{e^{h\nu/k_BT_\nu} + 1}\ . \tag{3.15}$$

We can calculate the radiation density of the massless neutrinos as

$$\rho_{0,\nu}c^2 = \int_0^\infty h\nu n_\nu(\nu)\,\mathrm{d}\nu = \left(\frac{7}{8}\right)a_BT_\nu^4\ . \tag{3.16}$$

For relating the temperatures between photons and neutrinos, consider the era before neutrino and photon decoupling. In that ultra high energy regime, because photons and neutrinos were coupled, the medium in which they existed had a fixed temperature. Other species in the medium were electrons (2 spin states), positrons (2 spin states), neutrinos (1 spin state for each of the three generations), and antineutrinos (1 spin states for each of the three generations). Shortly after the photon and neutrino decoupling, the temperature drops below the electron mass, and the forward reaction $e^+ + e^- \longleftrightarrow \gamma + \gamma$ (annihilation) becomes strongly favored. This heats up the photons. We can assume that this entropy transfer did not affect the neutrinos because they were already completely decoupled. Using entropy conservation of the electromagnetic plasma, we can calculate the change in the photon temperature before and after $e^\pm$ annihilation. This gives [47]

$$\frac{T_\nu}{T_0} = \left(\frac{4}{11}\right)^{1/3} .$$

The neutrino density is related to the photon density by

$$\rho_{0,\nu} = N_{\text{eff}} \left(\frac{7}{8}\right) \left(\frac{4}{11}\right)^{4/3} \rho_{0,\gamma} \ ,$$

where N_{eff} is the effective number of neutrinos. Theoretically, there are 3 neutrino families. However, due to non-instantaneous decoupling and QED effects, etc., the effective neutrino density will be slightly higher then this value. This can be accounted for by considering $N_{\text{eff}} > 3$. Considering a general framework for neutrino decoupling, it can be shown that for non-instantaneous neutrino decoupling, $N_{\text{eff}} \approx 3.034$. In addition, the QED effects contribute about $\Delta N_{\text{eff}} \approx 0.011$. Assuming these two effects can be added linearly, the final value of $N_{\text{eff}} \approx 3.045$ [48, 49, 88, 58, 54, 117].

Therefore, the third term in Eq. 3.10 can be calculated as

$$\Omega_{0,\nu} H_0^2 = \frac{\rho_{0,\nu}}{\rho_{cr}} H_0^2 = N_{\text{eff}} \frac{7}{8} \left(\frac{4}{11}\right)^{4/3} \frac{a_B}{c^2 \rho_{cr}} T_0^4 = 1.1324 \times N_{\text{eff}} \times 10^{-14} \times T_0^4 \ \texttt{Mpc}^{-2}. \quad (3.17)$$

3.3.4 Radiation density for massive neutrinos

In the standard model of particle physics, neutrinos are massless. However, different experiments point toward a small nonzero mass for the neutrinos. For massive neutrinos, the Fermi-Dirac distribution function contains a mass term, and it is not analytically integrable. Therefore, to get the density ρ_{ν_m} at any given redshift, the distribution function must be integrated numerically.

Assuming that all the neutrino species have equal mass, the mass of the neutrinos is given by

$$m_{\nu_m} = \frac{\rho_{0,\nu_m}}{N_{\text{eff}} n_{0,\nu_m}} = \frac{\Omega_{0,\nu_m}}{N_{\text{eff}}} \frac{\rho_{cr}}{n_{0,\nu_m}} , \tag{3.18}$$

where N_{eff} is the effective number of neutrinos. ρ_{0,ν_m} and n_{0,ν_m} are the massive neutrino density and number density at the present time, respectively. ρ_{cr} is the critical density.

The neutrino number density can be calculated by integrating the Fermi-Dirac distribution function:

$$n_{\nu_m} = \frac{8\pi}{h^3} \int_0^\infty \frac{p^2 \mathrm{d}p}{\exp(\sqrt{p^2c^2 + m^2c^4}/k_b T_{\nu_m}) + 1} . \tag{3.19}$$

For neutrinos, $pc \gg mc^2$, and we can ignore the term mc^2 in the above equation. This simplifies to

$$n_{\nu_m} = \frac{8\pi}{h^3} \int_0^\infty \frac{p^2 \mathrm{d}p}{\exp(pc/k_b T_{\nu_m}) + 1} = \frac{8\pi}{h^3 c^3} k_b^3 T_{\nu_m}^3 \int_0^\infty \frac{\xi^2 \mathrm{d}\xi}{e^\xi + 1} = \frac{8\pi c^3}{h^3} k_b^3 T_{\nu_m}^3 \zeta(3)\Gamma(3) , \tag{3.20}$$

where $\zeta(3)$ is the Riemann Zeta function and $\Gamma(3)$ is the Gamma function. $\Gamma(3) = 2! = 2$.

The density and pressure of massive neutrinos at any given redshift can be written as

$$\rho = \frac{8\pi}{h^3c^3}k_B^4 T_{\nu_m}^4 \int_0^\infty q^2 f(q)\epsilon(q)\,\mathrm{d}q\,, \tag{3.21}$$

$$P = \frac{8\pi}{h^3c^3}k_B^4 T_{\nu_m}^4 \int_0^\infty q^2 f(q)\frac{q^2}{3\epsilon}\,\mathrm{d}q\,, \tag{3.22}$$

where $q = apc$ and,

$$\epsilon = \frac{a}{k_B T_{\nu_m}}\sqrt{m_{\nu_m}^2 c^4 + (pc)^2}\,. \tag{3.23}$$

Here, in Eq. 3.22 the factor of 3 comes because we consider 3 spatial dimensions[1]. Simple rearrangements of the above equations give us the massive neutrino density and pressure in terms of the massless neutrino density, as

$$\rho = \left(\frac{7}{8}\right) a_B T_\nu^4 \rho_{\mathrm{dl}} = \left(\frac{7}{8}\right) a_B T_{0,\nu}^4 a^{-4} \rho_{\mathrm{dl}} = \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} a_B T_0^4 a^{-4} \rho_{\mathrm{dl}}\,, \tag{3.24}$$

$$P = \left(\frac{7}{8}\right) a_B T_\nu^4 p_{\mathrm{dl}} = \left(\frac{7}{8}\right) a_B T_{0,\nu}^4 a^{-4} P_{\mathrm{dl}} = \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} a_B T_0^4 a^{-4} P_{\mathrm{dl}}\,. \tag{3.25}$$

Here, ρ_{dl} and p_{dl} are dimensionless density and pressure and are expressed as

$$\rho_{\mathrm{dl}} = \frac{1}{\Upsilon}\int_0^\infty q^2 f(q)\epsilon(q)\,\mathrm{d}q\,, \tag{3.26}$$

$$P_{\mathrm{dl}} = \frac{1}{\Upsilon}\int_0^\infty q^2 f(q)\frac{q^2}{3\epsilon}\,\mathrm{d}q\,. \tag{3.27}$$

[1]For an ideal gas, the pressure can be found by $nmv^2/3$. n is the number density of the gas molecules, v is the velocity, and m is the mass of each molecule. The factor of 3 arises because we have considered 3 spatial dimensions and we consider that the velocity distribution of the gas is isotropic, i.e. $v_x^2 = v_y^2 = v_z^2 = v^2/3$. Eq. 3.22 can also be derived in a similar way, where q corresponds to the momentum.

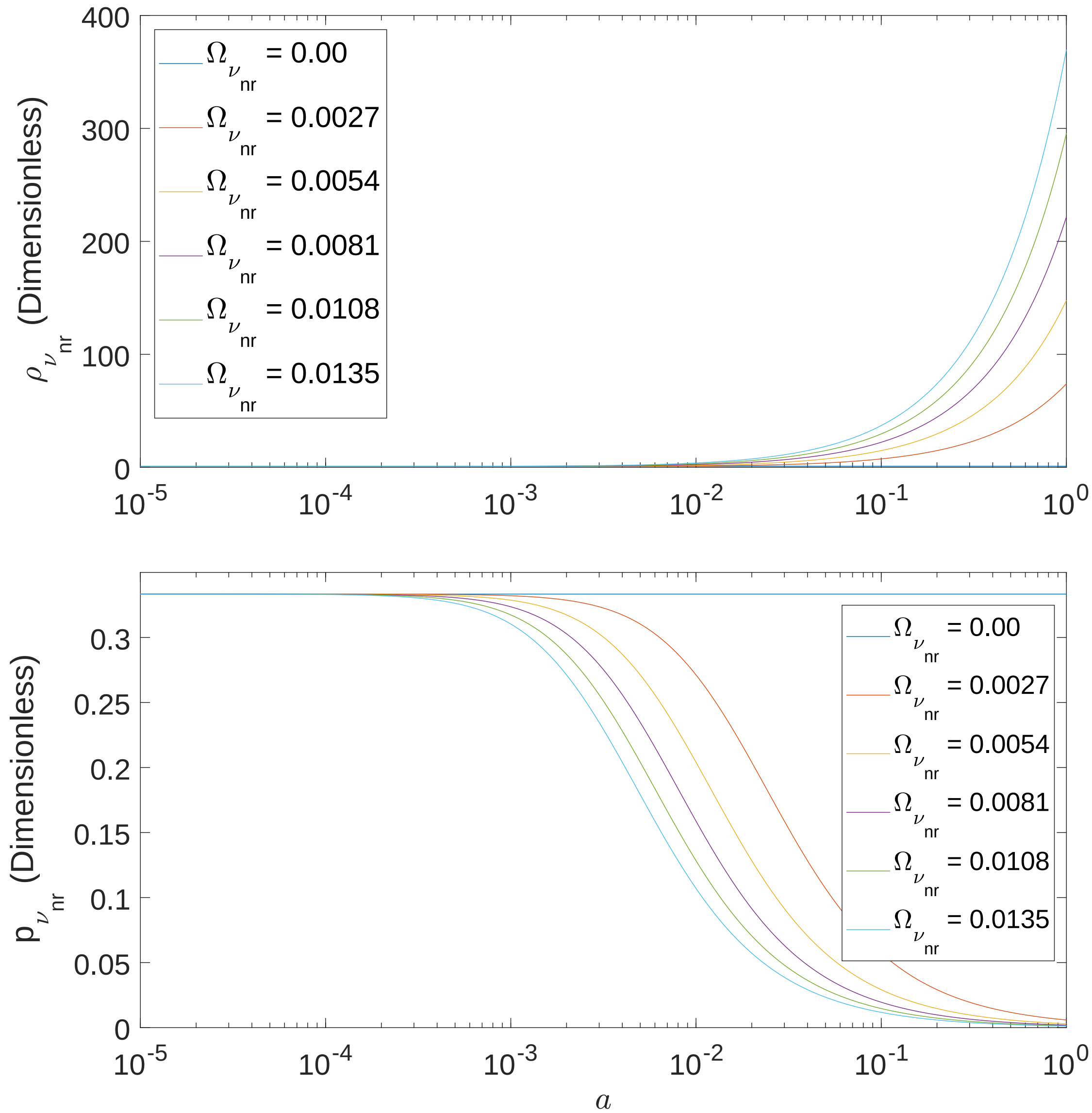

Figure 3.3: Dimensionless neutrino density ρ_{DL} (top) and pressure p_{DL} (bottom), given by Eq. 3.26 and Eq. 3.27, respectively, for different massive neutrino density parameters. For reference, the massless neutrinos are shown by the dark blue curve.

where $\Upsilon = \frac{7}{8}\frac{\pi^4}{15}$ [2]. In Fig. 3.3, we plot the dimensionless density and pressure for massive neutrinos for different density parameters, Ω_{ν_m} (note that $\sum m_{\nu_m}/93.14\texttt{eV} = \Omega_{\nu_m} h^2$, where h is the the Hubble parameter in units of 100 km/s/Mpc [89]. In the early Universe, where the temperature is high, this inequality goes the other way $pc \ll mc^2$, the neutrinos behave like massless particles and $\rho_{\rm dl} \to 1$ and $P_{\rm dl} \to \frac{1}{3}$. However, later, where mc^2 dominates, the massive neutrinos start behaving like matter particles and $P_{\rm dl} \to 0$ and $\rho_{\rm dl} \propto a$, i.e. the actual density of the massive neutrinos goes as a^{-3}.

3.3.5 Contribution from dark energy

The last term in Eq. 3.10 is the contribution from the dark energy. We can use the approximation $\Omega_{0,d} \approx 1 - \Omega_{0,m}$ (since $\Omega_{0,\gamma}$, $\Omega_{0,\nu}$, Ω_{ν_m} are of the order of 10^{-5}). For a ΛCDM model, the equation of state for dark energy is $w_d = -1$. However, several dark energy models have been proposed over the years based on a single scalar field, a mixture of multiple scalar fields (e.g. quintessence [112], K-essence [16, 33, 17], tachyon [106, 18], and dilatonic models [67]), massive vector fields [70, 25], etc. For different dark energy models, the equation of state for dark energy may vary as a function of the scale factor, i.e. $w_d(a)$. In such cases we can write the generalized form of Ω_d as

$$\Omega_d = \Omega_{0,d} \exp\left(-3 \int_1^a \frac{\mathrm{d}a}{a} \left[1 - w(a)\right] \right). \tag{3.28}$$

`CMBAns` is capable of handling both constant w_d or varying equation of state, $w_d(a)$, models of dark energy. Presently, there is tension between the Hubble parameter measured using the CMB (Planck data) and using local measurements (supernova

[2]Note that for calculating Υ we need the Bose-Einstein integration formula, $\int_0^\infty \frac{\xi^3 \mathrm{d}\xi}{e^\xi+1} = \frac{\pi^4}{15}$.

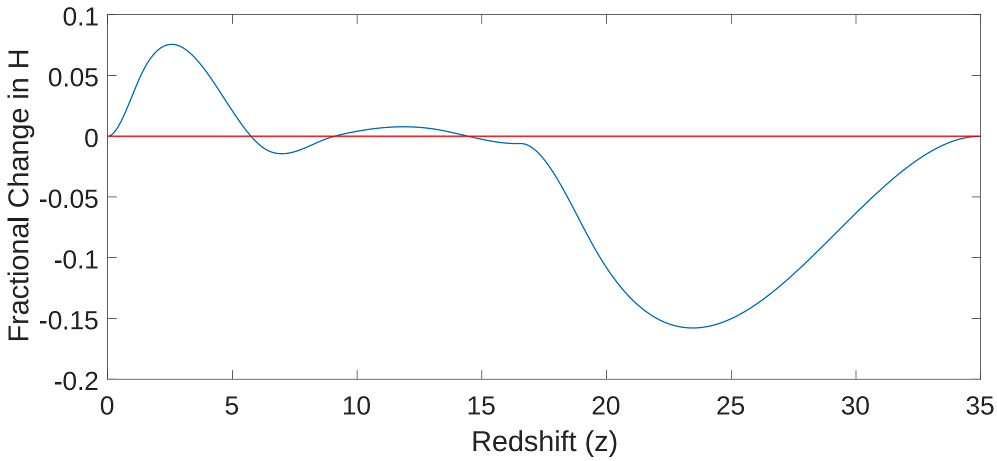

Figure 3.4: Fractional change in the Hubble parameter ($f(z) = \Delta H(z)/H_\Lambda(z)$) is shown here. We try to keep the distance to the last scattering surface to be constant. At low redshift we see a bump feature and at high redshift we see a dip feature.

data). The CMB-derived value is lower than the local measurements. Astronomers are trying to model the Hubble parameter as a function of redshift and modify the dark energy accordingly. In `CMBAns`, we have added modules which allow users to provide the Hubble parameter as a function of redshift using a Matlab-GUI input. `CMBAns` translates the Hubble parameter into the equation of state of dark energy [42] and calculates the resulting CMB power spectrum. In addition, there is the integrated Sachs-Wolfe (ISW) effect that is not part of the primordial power spectrum but it can cause the CMB power spectrum to appear uneven. It is caused by CMB photons traversing a time-varying linear gravitational potential, making the CMB gravitationally redshifted. `CMBAns` provides the primordial, late time (ISW) and the interference term between the primordial and late time ISW, along with the full power spectrum to show the effect of the late time expansion history of the Universe on the CMB power spectrum.

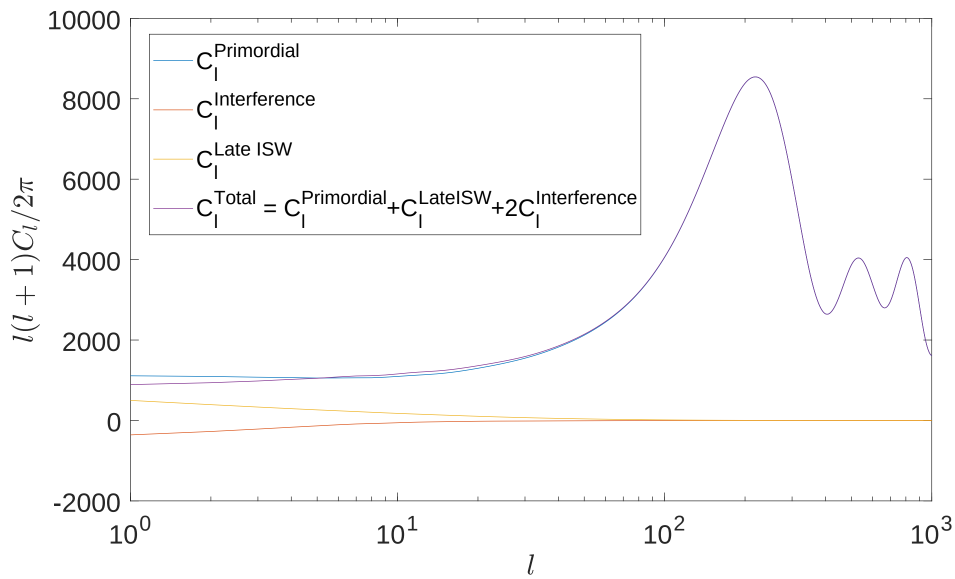

Figure 3.5: Several contributions to the CMB angular power spectrum for the Hubble parameter described by Fig. 3.4. We show the primordial (Sachs Wolf, velocity term, and the early ISW term) and the ISW term separately in this plot. The interference term is the cross term between the primordial and the ISW brightness fluctuation functions. However, this particular $f(z)$ is interesting because this shape gives a negative ISW contribution to the CMB temperature power spectrum.

Fig. 3.2 shows the variation of $\frac{\mathrm{d}a}{\mathrm{d}\tau}$ as a function of scale factor for different values of Ω_c and H_0 for the standard ΛCDM model. The conformal time between two given redshifts can be calculated by numerically integrating Eq. 3.10. For various dark energy models the shape of the CMB power spectrum changes. In Fig. 3.4 and Fig. 3.5 we demonstrate the special feature of `CMBAns` where we can adjust the deviation in the Hubble parameter to be different from the standard ΛCDM Hubble parameter ($f(z) = H(z)/H_\Lambda(z)$) as a function of redshift using a Matlab GUI input. The deviation that we have selected for this illustration is shown in Fig. 3.4. We try to keep the distance to the last scattering surface to be constant. The temperature power spectrum that we get for this particular deviation of Hubble parameter is shown in Fig. 3.5. An interesting fact for such kind of deviation is that the ISW part provides a very small, sometimes even negative, contribution to the C_l^{TT}. For this particular illustration, we have not considered any re-ionization. This is because the purpose of the illustration is to show the particular form of the late time ISW effect, which is not a well known phenomenon. As the polarization part doesn't have any ISW contribution, there will be no such effect on the polarization power spectra [42]. Without the GUI input, exploring such a model would have been immensely difficult.

3.4 Recombination and Reionization

For calculating the baryon sound speed, optical depth, and visibility function, we need to calculate the recombination and the reionization process very accurately. `CMBAns` provides various functions for calculating the recombination, including the Saha equation, the Peebles equation, `recfast`, or `CosmoRec` method.

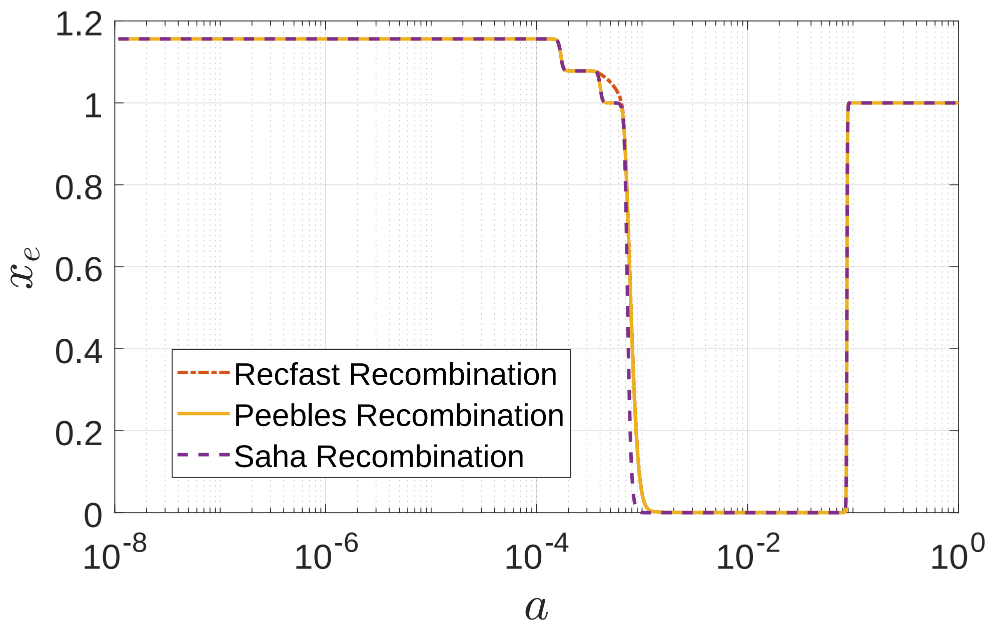

Figure 3.6: Ionization fractions for Saha, Peebles and `recfast` recombination processes are shown as a function of the scale factor. The reionization is considered as a step function. The first step on the left is for $He^{++} \rightarrow He^{+}$. The second step is for $He^{+} \rightarrow He$ recombination. Vertical axis is x_e^{tot} as defined by Eq. 3.39.

3.4.1 Saha Equation

The Saha equation provides a very rough estimate of the recombination epoch. It assumes the recombination reaction $p + e^- \longleftrightarrow H + \gamma$ is fast enough that it proceeds near thermal equilibrium, i.e. it ignores the expansion of the Universe. According to the Saha equation,

$$n_{\rm H}\frac{x_e^2}{1-x_e} = \left(\frac{k_B m_e T_b}{2\pi\hbar^2}\right)^{3/2} e^{-B_1/k_B T_b}\,, \tag{3.29}$$

where x_e is the hydrogen ionization fraction. $n_{\rm H}$ is the number density of the hydrogen atoms, i.e. $n_H = n_{1s}+n_p$, where n_{1s} and n_p are the number density of neutral hydrogen and ionized hydrogen, respectively. $B_1 = m_e e^4/(2\hbar^2) = 13.6\,\texttt{eV}$ is the ionization potential of the hydrogen atom. T_b is the baryon temperature.

The hydrogen number density can be calculated as

$$n_{\rm H} = n_b\,(1-Y_{\rm He}) = \frac{\rho_b}{m_{\rm H}}\,(1-Y_{\rm He}) = \frac{3}{8\pi G}\Omega_{0,b}a^{-3}H_0^2\frac{(1-Y_{\rm He})}{m_H}, \tag{3.30}$$

where $Y_{\rm He}$ is the helium fraction after Big Bang nucleosynthesis. In Fig. 3.6 we show the recombination result using the Saha equation. The plot shows that the recombination of hydrogen is almost instantaneous. For the helium recombination, we separately use the Saha equation, given by Eq. 3.37.

3.4.2 Peebles' Recombination

Peebles' equation provides a very accurate estimate of the recombination history of hydrogen. The calculations are done using effective three-level atom calculations. Peebles' formalism is based on the assumptions that

- Direct recombinations to the ground state of hydrogen are very inefficient: each such event leads to a photon with energy greater than $13.6\,\mathtt{eV}$, which almost immediately re-ionizes a neighboring hydrogen atom. Electrons therefore only efficiently recombine to the excited states of hydrogen, from which they cascade very quickly down to the first excited state, with principal quantum number $n = 2$.

- From the first excited state, electrons can reach the ground state $n = 1$ through two pathways:

  1. Decay from the 2p state by emitting a Lyman-α photon. This photon will almost always be reabsorbed by another hydrogen atom in its ground state. However, cosmological redshifting systematically decreases the photon frequency, and hence there is a small chance that it escapes reabsorption if it gets redshifted far enough from the Lyman-α line resonant frequency before encountering another hydrogen atom.

  2. Decay from the $2s$ to $1s$ state, which is forbidden with a single transition and only possible using through a double transition. The rate of this transition is very slow, $8.22\,\mathtt{s}^{-1}$. It is however competitive with the slow rate of Lyman-α escape in producing ground-state hydrogen.

- Atoms in the first excited state may also be re-ionized by the ambient CMB photons before they reach the ground state, as if the recombination to the excited state did not happen in the first place. To account for this possibility, Peebles defines the factor C as the probability that an atom in the first excited state reaches the ground state through either of the two pathways described above before being photo-ionized.

Accounting for these processes, the recombination history is then described by the differential effect [107]

$$\frac{\mathrm{d}x_{\mathrm{e}}}{\mathrm{d}t} = -aC\left(\alpha^{(2)}(T_b)n_{\mathrm{p}}x_e - 4(1-x_{\mathrm{e}})\beta(T_b)e^{-E_{21}/T}\right), \tag{3.31}$$

where

$$\beta(T_b) = \left(\frac{m_e k_{\mathrm{B}} T_b}{2\pi\hbar^2}\right)^{3/2} e^{-B_1/k_{\mathrm{B}}T_b}\,\alpha^{(2)}(T_b). \tag{3.32}$$

The recombination rate to excited states [86] is taken as

$$\alpha^{(2)}(T_b) = \frac{64\pi}{(27\pi)^{1/2}}\frac{e^4}{m_e^2c^3}\left(\frac{k_{\mathrm{B}}T_b}{B_1}\right)^{-1/2}\phi_2(T_b)\ , \qquad \phi_2(T_b) \approx 0.448\ \ln\left(\frac{B_1}{k_{\mathrm{B}}T_b}\right)\ . \tag{3.33}$$

This expression for $\phi_2(T_b)$ provides a good approximation at low temperature. At high temperature this expression underestimates ϕ_2, but the amount is negligible. For $T_b > B_1/k_{\mathrm{B}} = 1.58\times10^5\,\mathtt{K}$, we set $\phi_2 = 0$.

$$C = \frac{\Lambda_\alpha + \Lambda_{2s\to1s}}{\Lambda_\alpha + \Lambda_{2s\to1s} + \beta^{(2)}(T_b)} \tag{3.34}$$

where

$$\beta^{(2)}(T_b) = \beta(T_b)e^{+hc/\lambda_\alpha k_B T_b}\ , \qquad \Lambda_\alpha = \frac{8\pi\dot{a}}{a^2\lambda_\alpha^3 n_{1s}}\,. \tag{3.35}$$

$\lambda_\alpha = \frac{8\pi\hbar c}{3B_1} = 1.21567\times10^{-7}\mathtt{m}$, is the wavelength for Lyman-α emission. Over-dot represents the derivative with respect to the conformal time. $\Lambda_{2s\to1s}$ is the rate of hydrogen double transition from $2s$ to $1s$. $\Lambda_{2s\to1s} = 8.227\mathtt{s}^{-1} = 8.4678\times10^{14}\mathtt{Mpc}^{-1}$.

$$\begin{aligned}
\Lambda_{2s\to 1s}/\Lambda_\alpha &= \frac{\Lambda_{2s\to 1s}\lambda_\alpha^3 a^2 n_{1s}}{8\pi\dot{a}} = \frac{\Lambda_{2s\to 1s}\lambda_\alpha^3(1-x_e)a^3 n_H}{8\pi\dot{a}a} = (1-x_e)(1-Y_{He})\frac{\Lambda_{2s\to 1s}\lambda_\alpha^3}{8\pi\dot{a}a}\frac{a^3\rho_m}{m_H} \\
&= \Lambda_{2s\to 1s}\left(\frac{\lambda_\alpha^3}{8\pi}\frac{3}{8\pi G}\frac{1}{m_H}\right)\frac{(1-x_e)}{\dot{a}a}(1-Y_{He})\Omega_{m0}H_0^2 \\
&= \left(8.4678\times 10^{14}\right)\times\left(8.0230194\times 10^{-26}\right)\frac{(1-x_e)}{\dot{a}a}(1-Y_{He})\Omega_{m0}H_0^2. \qquad (3.36)
\end{aligned}$$

Similarly, $\beta^{(2)}(T_b)/\Lambda_\alpha$ can be calculated using

$$\frac{\beta^{(2)}(T_b)}{\Lambda_\alpha} = T_b\phi_2(T_b)\mathcal{K}e^{-0.25T_{ion}/T_b}\left(8.0230194\times 10^{-26}\right)\frac{(1-x_e)}{\dot{a}a}(1-Y_{He})\Omega_{m0}H_0^2,$$

where $\mathcal{K} = \left(\frac{64\pi}{(27\pi)^{1/2}}\frac{e^4}{m_e^2c^3}\left(\frac{k_\mathrm{B}}{B_1}\right)^{-1/2}\left(\frac{m_e k_\mathrm{B}}{2\pi\hbar^2}\right)^{3/2}\right) = 5.13\times 10^{18}$. Here, H_0 is in `km/sec/MPc` unit, and $\dot{a}$ has units of $\texttt{MPc}^{-1}$. The numerical values are converted to match these units.

Helium Recombination

For calculating the He recombination, we use the Saha Equation [86].

$$\frac{n_e x_{n+1}}{x_n} = \frac{2g_{n+1}}{g_n}\left(\frac{m_e k_B T_b}{2\pi\hbar^2}\right)^{3/2} e^{-\chi_n/k_B T_b}\,, \qquad (3.37)$$

where $n \in (0,1)$, and $x_0 = 1 - x_1 - x_2$. The helium ionization fractions $x_1 = n\left(\mathrm{He}^+\right)/n\left(\mathrm{He}\right)$ and $x_2 = n\left(\mathrm{He}^{++}\right)/n\left(\mathrm{He}\right)$, where $n\left(\mathrm{He}\right)$ is the total number density of helium nuclei. n_e is the free electron number density. $g_0 = g_1 = 1$ and $g_2 = 2$. $\frac{\chi_1}{k_B} = T_1^{ion} = 2.855\times 10^5\,\texttt{K}$ and $\frac{\chi_2}{k_B} = T_2^{ion} = 6.313\times 10^5\,\texttt{K}$ are the first and second ionization temperatures of He.

3.4.3 Recfast, CosmoRec

Peebles' three-level atom model accounts for the most important physical processes. However, these approximations may lead to errors on the predicted recombination history at a level as high as 10%. This can also alter the temperature and polarization power spectra up to $3-5\%$ at high multipoles. Several research groups have revisited the details and proposed different models like `recfast`[3][119, 120], `CosmoRec`[4][35, 8, 36, 128, 57, 115], `HyRec`[5][7] etc. These packages can calculate the recombination history up to 0.1% accuracy. We use the available `CosmoRec` code in `CMBAns` as a default case. However, users can choose to use the Saha, Peebles or `recfast` routines, which are also available in `CMBAns`. The other packages can also be easily added to `CMBAns` or run separately. In the later case, the ionization fraction, and baryon temperature can be stored in a file as a function of scale factor and passed to `CMBAns`.

In Fig. 3.6, we show the ionization fraction computed by `CMBAns` for different recombination methods. We use a smooth reionization, where we join an ionization fraction before and after the reionization using a tanh(...) function. In Fig. 3.7, we show the differences between `recfast`, `recfast++` and `CosmoRec` recombination. This small change in the ionization fraction can change the C_l at high multipoles.

3.4.4 Calculating baryon temperature

For calculating the ionization fraction during recombination, we need the baryon temperature at each scale factor. The rate of change of the baryon temperature can be calculated as (check Appendix **??**)

[3] https://www.cfa.harvard.edu/~sasselov/rec/
[4] http://www.jb.man.ac.uk/~jchluba/Science/CosmoRec/Welcome.html
[5] https://cosmo.nyu.edu/yacine/hyrec/hyrec.html

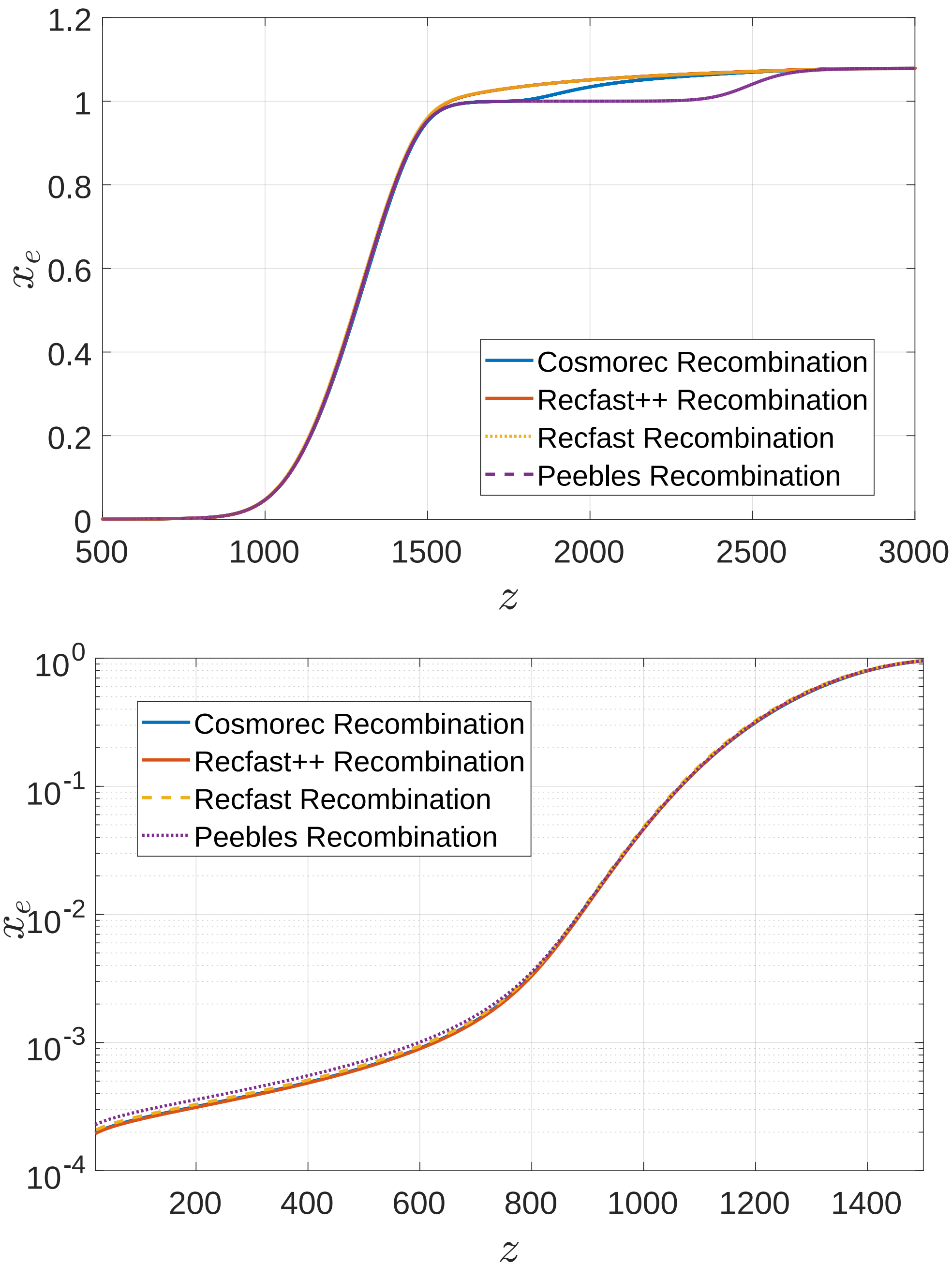

Figure 3.7: Comparison between the ionization fractions from different modern recombination routines: `recfast`, `recfast++` and `CosmoRec`. For `CosmoRec`, we choose the dark matter annihilation efficiency to be 10^{-24} `eV/sec` and all the other parameters are set to default settings. Top: Ionization fraction is plotted with a linear scale to show the He^+ recombination. Bottom: Ionization fraction is plotted with a log scale to amplify the effect at low redshift after the H^+ recombination.

$$\dot{T}_b = -2\left(\frac{\dot{a}}{a}\right)T_b + \frac{8\pi^2}{45}\frac{k_B^4}{c^4\hbar^3}\frac{\sigma_T T_\gamma^4}{m_e} f_e\left(T_\gamma - T_b\right) \,, \tag{3.38}$$

where σ_T is the Thomson scattering cross section. f_e is given by

$$f_e = \frac{\left(1 - Y_{He}\right) x_e^{tot}}{1 - \frac{3}{4}Y_{He} + (1 - Y_{He})x_e^{tot}} \,. \tag{3.39}$$

x_e^{tot} is the total ionization fraction and is given by

$$x_e^{tot} = x_e + \frac{1}{4}Y_{He}\frac{(x_1 + 2x_2)}{(1.0 - Y_{He})} \,. \tag{3.40}$$

The constant term in Eq. 3.38 is given by $\frac{8\pi^2}{45}\frac{k_B^4}{c^4\hbar^3}\frac{\sigma_T}{m_e} = 5.0515 \times 10^{-8}\mathtt{K}^{-4}\mathtt{Mpc}^{-1}$. We can see that the baryon temperature depends on the ionization fraction of the electrons. Therefore, we need to jointly evaluate the baryon temperature and ionization fraction. The temperature of the photons at any era is $T_\gamma = a^{-1}T_{0\gamma}$. We can consider $T_b = T_\gamma$ before recombination (in the tight coupling era), and we can use it as the initial condition for solving Eq. 3.38.

3.4.5 Baryon sound speed, optical depth and visibility

Calculating the baryon acoustic oscillations requires the speed of sound in the plasma, c_s. If we consider the plasma as a single fluid, then the pressure, density and the temperature of the fluid will be related as $P_b = \frac{k_B}{m}\rho_b T_b$. We can calculate the sound speed in the plasma as

$$c_s^2 = \left.\frac{\mathrm{d}P_b}{\mathrm{d}\rho_b}\right|_{adiabatic} = \frac{k_B T_b}{m}\left(1 - \frac{1}{3}\frac{\mathrm{d}(\ln T_b)}{\mathrm{d}(\ln a)}\right)$$
$$= \frac{k_B T_b}{m_p}\left[1.0 - \frac{3}{4}Y_{He} + (1.0 - Y_{He})x_e^{tot}\right]\left(1 - \frac{1}{3}\frac{\mathrm{d}(\ln T_b)}{\mathrm{d}(\ln a)}\right) \quad (3.41)$$

Here m is the mean molecular weight of the fluid, and m_p is the mass of a proton[6]. The mean molecular weight is calculated assuming the fluid contains free electrons, H, H^+, He, He^+, and He^{++}. Here, one should note that a more accurate formulation of the sound speed was proposed by [75], and is used in `CAMB` and `CLASS`. We are in the process of implementing it in `CMBAns`.

The optical depth from the present time (τ_0) to any conformal time τ is given by

$$\kappa = \int_{\tau}^{\tau_0} a n_e \sigma_T \mathrm{d}\tau = \int_{\tau}^{\tau_0}\left(\frac{H_0^2 c^2}{8\pi G}\right)\left(\frac{\Omega_b}{m_H a^2}\right)\sigma_T(1 - Y_{He})\mathrm{d}\tau \ . \quad (3.42)$$

The visibility function at any conformal time τ can be calculated as $g = \dot{\kappa}\exp(-\kappa)$. In Fig. 3.8 we show the visibility function vs the scale factor. The visibility function is nonzero only during the recombination and reionization processes. The change in the visibility function is significantly smaller during reionization than during recombination. To show both on the same plot, we multiply the reionization part by 100.

This is the end of my contribution to the original paper. For more details on the rest of the calculation of scalar and tensor perturbations, and the corresponding scalar and tensor power spectra, please refer to the original paper. Figures 3.9 and 3.10 show

[6]For all our calculations, we consider the mass of H and $H^+ = m_p$, and the mass of He, He^+, $He^{++} = 4m_p$, i.e. we consider that the mass of electron is negligible and the mass of the proton and neutron are the same.

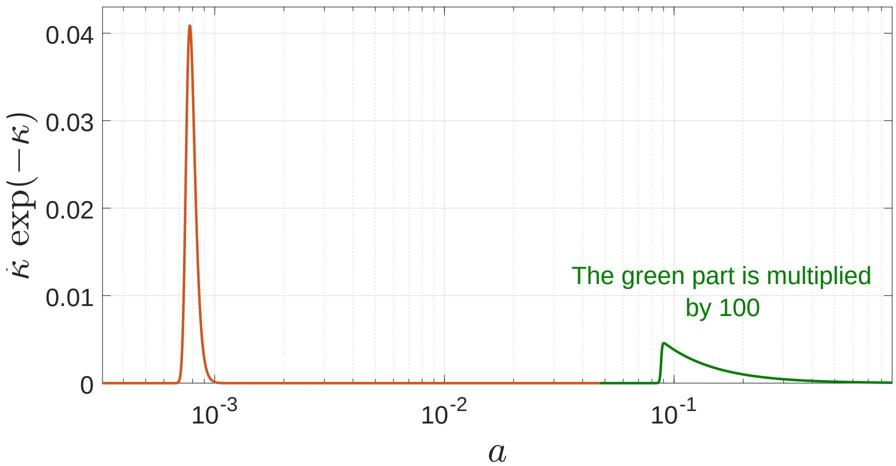

Figure 3.8: Plot of the visibility function ($g = \dot{\kappa}\exp(-\kappa)$) as a function of the scale factor. The green section of the plot is multiplied by 100 for displaying it on the same plot.

the scalar and tensor power spectra for different types of initial conditions as calculated by `CMBAns`.

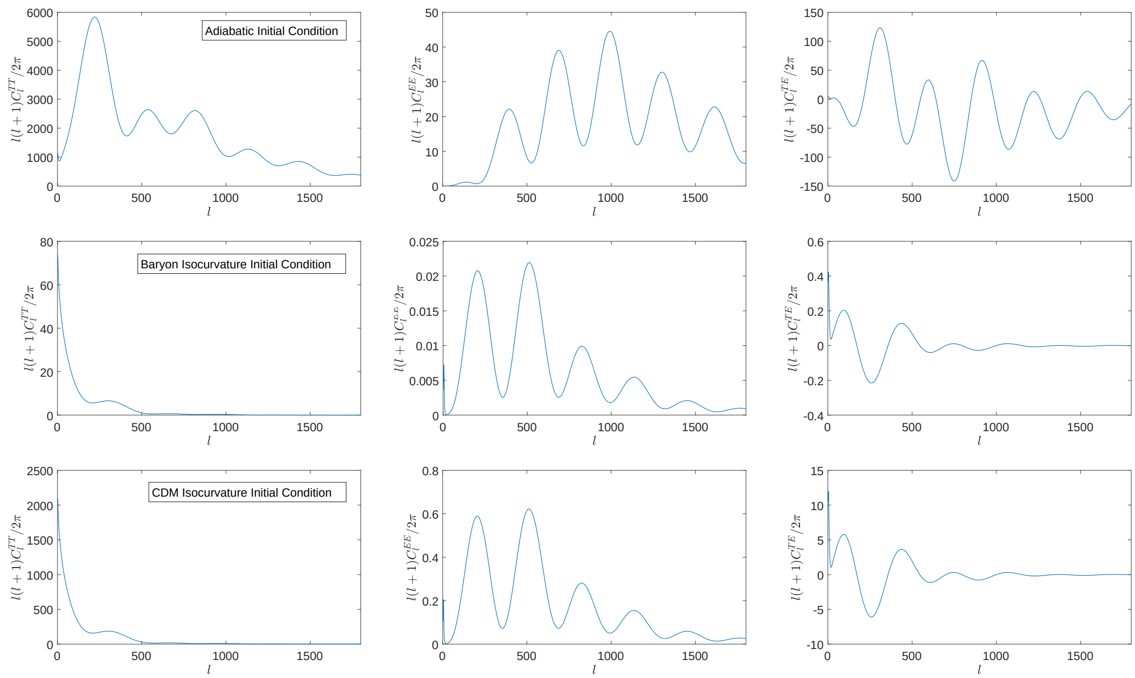

Figure 3.9: These plots show the unlensed CMB scalar power spectra (C_l) for adiabatic (top row), baryon isocurvature (middle row), and CDM isocurvature (bottom row) initial conditions. We use $\Omega_b h^2 = 0.0223$, $\Omega_b h^2 = 0.1188$, $h = 67.74$ km/s/Mpc, $n_s = 0.9667$, $\kappa = 0.08$. The plots show that the isocurvature CMB power spectrum decays at high l.

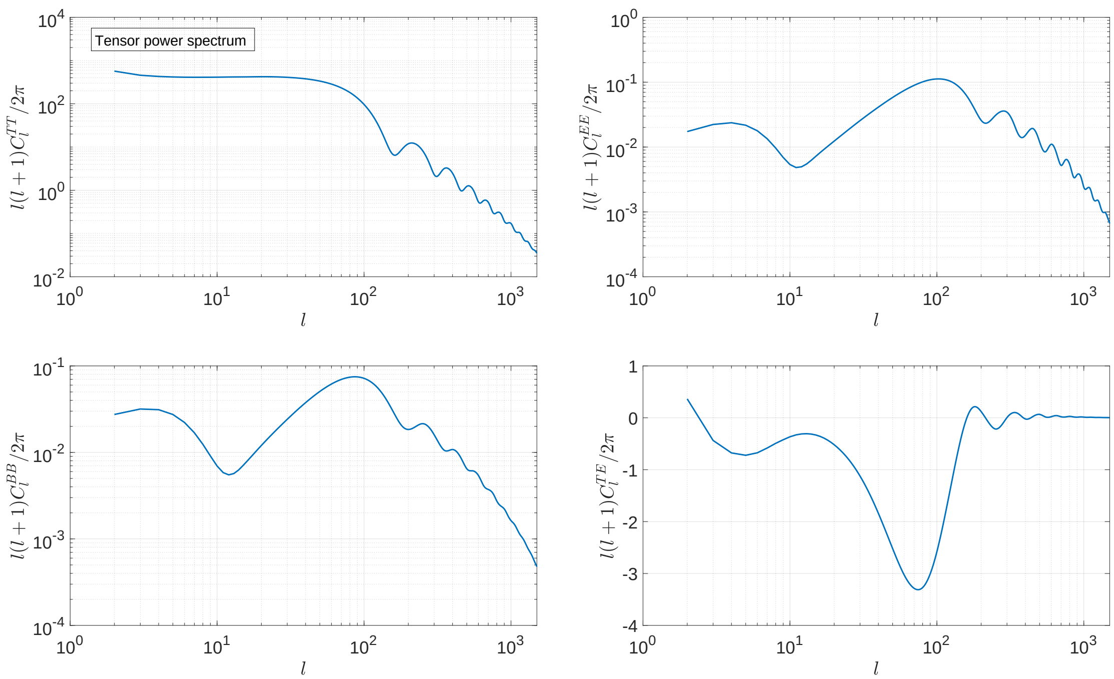

Figure 3.10: These plots show the unlensed CMB tensor power spectra (C_l^t). We use $\Omega_b h^2 = 0.0223$, $\Omega_b h^2 = 0.1188$, $h = 67.74$ km/s/Mpc, $n_t = 0.04$, $\kappa = 0.08$. As C_l^{TE} has negative values we plot the y-axis in linear scale.

3.5 Conclusion

CMBAns allows fast and accurate calculation of the CMB power spectrum for a flat ($\Omega_k = 0$) background cosmology. The lensing calculations and the comparison of the results with other Boltzmann packages like CAMB, CLASS etc. are not discussed in this paper and are left for future work.

We use the C programming language for CMBAns. However, to make the program object oriented, we use the concept of class from C++. A similar technique is also used in CLASS code. Several stand alone modules, such as calculating the recombination history, power spectrum evolution with different cosmological parameters, Bessel function calculations, etc. are provided with the package. However, users are not limited to what already comes with the program. The influx of precision CMB data means that CMB modeling tools must evolve quickly. Modularity, an important feature of CMBAns, offers a way to solve this problem. The modularity of CMBAns offers a lot of flexibility and lets users quickly expand the functionalities of the package to include new cosmological models by simply writing a new module or classes using the functionality already provided in CMBAns.

Chapter 4

Introduction to Intensity Mapping

In this chapter, we introduce neutral hydrogen (HI) intensity mapping. The distribution of HI roughly follows the distribution of stars and galaxies, which are biased tracers of dark matter. Instead of resolving individual galaxies with traditional optical observations, intensity mapping performs low angular resolution observations in three-dimensions and thus can in principle, rapidly survey very large cosmological volumes from the present day to the dark ages. We will also introduce the optimal region in k-space for avoiding strong foregrounds for intensity mapping, called the EoR window.

In contrast to the CMB where the primary anisotropies arise from a thin shell centered at $z = 1100$, intensity mapping can probe a much bigger volume of the Universe. For the CMB, the number of modes with a comoving wave number $k \equiv 2\pi/\lambda$ between k and $k+\mathrm{d}k$ is $\mathrm{d}N_{\mathrm{CMB}} = \pi k \mathrm{d}k\,[\mathcal{A}/(2\pi)^2]$, where $\mathcal{A} = D^2\mathrm{d}\Omega$, D is the comoving distance to the surface of last scattering, and $\mathrm{d}\Omega$ is the solid angle of the sky survey. Intensity mapping has an advantage over 2-dimensional CMB mapping: it can also sample the sky at different redshifts (along the line of sight) from the dark ages to the present day. The number of accessible modes is $\mathrm{d}N_{\mathrm{3D}} = 2\pi k^2 dk\, [\mathcal{V}/(2\pi)^3]$. In practice, foreground contamination and instrumental mode-mixing will restrict the number of available modes. This is discussed later in the EoR window section. Nonetheless, 3D intensity mapping can still provide access to a higher number of modes than 2D CMB observation. The Planck measurement offers roughly $N_{\mathrm{modes}}^{\mathrm{CMB}} \sim 2\ell_{\mathrm{max}}^2 \sim 10^7$ modes. The 21 cm emission can access to $\ell_{\mathrm{max}} \sim 10^6$ and up to 10^4 independent redshift slices, so in principle $N_{\mathrm{modes}}^{21\ \mathrm{cm}} \sim 10^{16}$. Figure 4.1 compares the number of available modes for different sky surveys, including the CMB, galaxy surveys, and 21 cm surveys.

4.1 The Inhomogeneous Universe

In the previous chapter, we have shown how a perfectly homogeneous Universe evolves, with small perturbations in the early Universe leaving traces as anisotropy in the CMB signal. These perturbations later grow to form structures in the Universe. We can observe large inhomogeneities in our local Universe, such as galaxy clusters, groups of galaxies, galaxies, and voids that form the cosmic web. On smaller scales, we have different galactic components, galaxy halos, interstellar medium, stars, planets, etc. As stated in the previous chapter, the very early Universe and the Universe at the time of the CMB were much more homogeneous, with average relative inhomogeneities of $\sim 10^{-5}$. On the contrary, in our local Universe, the average relative inhomogeneities is several orders of magnitude higher, with the typical relative inhomogeneity necessary to form a dark matter halo being around 200. In order to understand structure formation, we will need to consider inhomogeneities in the primordial density field. The primordial inhomogeneities are not well understood, but most inflationary models assume a Gaussian random primordial field, which is consistent with currrent observations

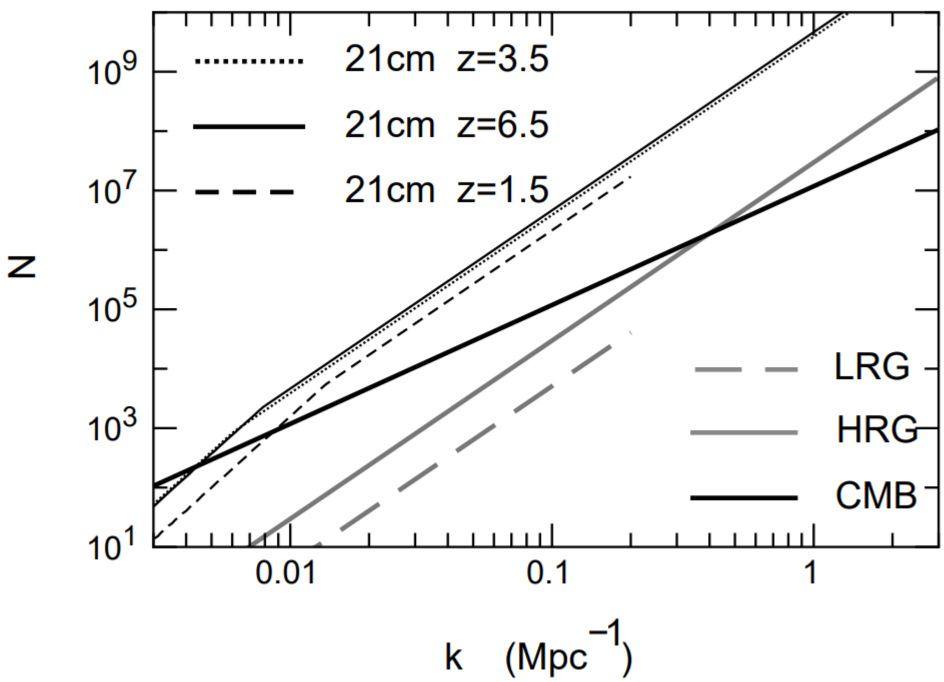

Figure 4.1: The number of available modes, N, within a k bin of width $\Delta k = k/10$ centered on k for different cosmological surveys. The thick dashed grey line (LRG) corresponds to the spectroscopic sample of SDSS. The thick solid line (HRG) corresponds to a future spectroscopic survey at $2.5 < z < 3.5$ covering 1000 square degrees with a co-moving galaxy density equal to the LRG sample. The thick dark line corresponds to a CMB data set with 65% sky coverage. The thin lines show the number of modes accessible in a 21 cm survey (including the constraints by foreground removal) covering 65% of the sky within a redshift range spanning a factor of 3 in (1 + z), and centered on z = 1.5, 3.5 and 6.5. Figure from Loeb & Wyithe (2008) [84]

.

[5]. These random Gaussian initial conditions eventually grow into the non-linear structures that we are currently seeing in the local Universe.

4.1.1 Linear Perturbation Theory

We will discuss the time evolution of the matter density field using linear perturbation theory. We define the matter overdensity field as

$$\delta(\mathbf{x},t) = \frac{\rho(\mathbf{x},t)}{\bar{\rho}(t)} - 1 \tag{4.1}$$

in which $\rho(\mathbf{x},t)$ is the matter field density at $\mathbf{x}$ and time t, $\bar{\rho}$ is the average matter field density over all space. We can see that the overdensity field has to satisfy $-1 \leq \delta(\mathbf{x},t) < \infty$. We can treat δ as a perturbation since it is small at early times and large scales.

To describe the evolution of the matter overdensity field $\delta(\mathbf{x},t)$, we will need a few assumptions. First, the dark matter only interacts with other types of matter gravitationally. Second, the scales are smaller than the Hubble horizon ($\lambda \ll H^{-1}$) beyond which general relativity effects can become significant. Third, we will use the Newtonian expansion, i.e. the peculiar velocities of the particles are not relativistic ($v_p \ll c$). This means the dark matter is cold, so it is massive enough that it is slow moving, as opposed to warm or hot dark matter. Lastly, we assume a matter-dominated Universe where $\Omega_m = 1$

For a perfect fluid, the continuity, Euler's and Poisson's equations in terms of $\delta(\mathbf{x},t)$ can be written as

$$\begin{aligned}
\dot{\delta} + \frac{1}{a}\nabla \cdot \mathbf{v} = 0 \\
\dot{\mathbf{v}} + H_{\mathrm{v}} + \frac{\nabla P}{\rho a} + \frac{1}{a}\nabla\Phi = 0 \\
\nabla^2\Phi = 4\pi G a^2 \bar{\rho}\delta,
\end{aligned} \tag{4.2}$$

where Φ is the gravitational potential, P is the pressure, and the spatial derivatives are taken with respect to the comoving coordinates x. The three equations above can be combined into one second-order differential equation in $\delta(\mathbf{x}, t)$:

$$\frac{\partial^2\delta}{\partial t^2} + 2H\frac{\partial\delta}{\partial t} - 4\pi G\bar{\rho}\delta = \frac{\nabla^2 P}{\bar{\rho}a^2}. \tag{4.3}$$

Assuming that the fluid is pressureless, this equation becomes

$$\frac{\partial^2\delta}{\partial t^2} + 2H\frac{\partial\delta}{\partial t} = 4\pi G\bar{\rho}\delta. \tag{4.4}$$

The term on the right is the gravitational source term, and the second term on the left is a friction term that describes the growth of perturbations due to Hubble expansion. For linear perturbations, we can assume

$$\delta(\mathbf{x}, t) = D(a)\delta_0(\mathbf{x}), \tag{4.5}$$

where $\delta_0(\mathbf{x})$ is the initial overdensity field, and $D(a)$ is the linear growth factor. In the matter dominated era, the solution for a given initial field is given by two independent growth factors:

$$\begin{aligned}
D_+(t) \propto a(t), \\
D_-(t) \propto a(t)^{-3/2},
\end{aligned} \tag{4.6}$$

where $D_+(t)$ and $D_-(t)$ are the growing and decaying modes, respectively. The final solution is a linear combination of those two modes:

$$\delta(\mathbf{x}, t) = A(\vec{x})D_+(t) + B(\vec{x})D_-(t). \tag{4.7}$$

In the full ΛCDM Universe where $\Omega_m < 1$, the linear growth factor is given by:

$$D_+(z) = \frac{5}{2}\Omega_m H(z) \int_z^\infty \frac{1+z}{H(z)^3} \mathrm{d}z. \tag{4.8}$$

The solution describes how overdense regions ($\delta > 0$) will become more overdense, and the underdense regions ($\delta > 0$) will become more underdense. The linear perturbation breaks down when δ becomes larger than 1.

4.1.2 Correlation function

The perturbation theory cannot predict where the overdense regions are, but they can predict how matter clusters together. In order to model the clustering structure of observed matter, with tracers being galaxies or gas, we will use the correlation function of the overdensity field, which is assumed to be a Gaussian random field with zero mean. We need a correlation function to describe and measure the density perturbations as a function of their separation. The correlation function $\xi(r)$ is defined as the excess probability of finding a pair of galaxies at a given separation $r_{12} = |\mathbf{x}_1 - \mathbf{x}_2|$:

$$dP = n^2 \delta V_1 \delta V_2 \left[1 + \xi\left(r_{12}\right)\right], \tag{4.9}$$

where n is the average number density and δV_i is the volume element. The correlation is invariant to translations (independent of $\mathbf{x}$) because of homogeneity and rotations (independent of the direction of $\mathbf{r}_{12}$) because of isotropy, so ξ must only depend on the relative separation r_{12}. The two-point correlation function is defined as

$$\xi(r_{12}) = \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2)\rangle. \tag{4.10}$$

The majority of autocorrelation information is contained in the two-point correlation function. For highly non-linear structures, caused by deviation from random Gaussian assumptions or non-linear gravitational collapse, we may need higher n-point correlation function

4.1.3 Power spectrum

Similar to the CMB, to derive the power spectrum, we will work with the overdensity field $\delta(\mathbf{x})$ in Fourier space. We define $\delta(\mathbf{k})$ to be the Fourier dual to $\delta(\mathbf{x})$ as follows:

$$\begin{aligned} \delta(\mathbf{k}) &= \mathcal{F}[\delta(\mathbf{x})] = \int \delta(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}^3 x \\ \delta(\mathbf{x}) &= \mathcal{F}^{-1}[\delta(\mathbf{k})] = \frac{1}{(2\pi)^3} \int \delta(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}^3 k, \end{aligned} \tag{4.11}$$

where $k = 2\pi/\lambda$. The power spectrum $P(k)$ of the overdensity field is then defined as

$$\langle \delta\left(\mathbf{k}_1\right) \delta^*\left(\mathbf{k}_2\right)\rangle \equiv (2\pi)^3 \delta_{\mathrm{D}}^3\left(\mathbf{k}_1 + \mathbf{k}_2\right) P\left(k_1\right), \tag{4.12}$$

where the average $\langle\rangle$, in theory, is done over multiple realizations of the Universe. However, we only have one observed instance of the Universe. Therefore, similar to the CMB, there is a fundamental uncertainty in the knowledge of the power spectrum. This uncertainty is the cosmic variance. It can also be shown that the power spectrum is a Fourier transform of the correlation function

$$
\begin{aligned}
\xi(r) &= \frac{1}{(2\pi)^3}\int P(k)e^{-i\mathbf{k}\cdot\mathbf{x}}\mathrm{d}^3k \\
P(k) &= \int \xi(r)e^{i\mathbf{k}\cdot\mathbf{x}}\mathrm{d}^3r.
\end{aligned}
\tag{4.13}
$$

We are only concerned about the magnitude k of $\mathbf{k}$ because of the isotropy assumption. In reality, any sky survey will be done over a large but finite volume V of the Universe, and equation 4.11 and 4.12 becomes:

$$
\begin{aligned}
\delta(\mathbf{k}) &= \int_V \delta(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}\mathrm{d}^3x \\
\frac{\langle \delta\left(\mathbf{k}_1\right)\delta^*\left(\mathbf{k}_2\right)\rangle}{V} &= \delta_{\mathbf{k}_1,\mathbf{k}_2}P\left(k_1\right).
\end{aligned}
\tag{4.14}
$$

From the first equation, we see that the Fourier coefficient of the overdensity is given by $a_{\mathbf{k}} = \delta_{\mathbf{k}}/V$. The second equation tells us that the power spectrum $P(k)$ is the product of the variance of Fourier coefficients with the survey's volume. The variance of Fourier coefficients varies as $\langle a_{\mathbf{k}_1} a_{\mathbf{k}_2}\rangle \propto 1/V$. This means the larger the survey's volume, the more samples we have of the smaller Fourier modes.

It is conventional in the literature to talk about the dimensionaless power spectrum $\Delta^2(k)$ instead of the traditional power spectrum $P(k)$:

$$
\Delta^2(k) = \frac{k^3P(k)}{2\pi^2}
\tag{4.15}
$$

where the Dirac delta becomes the Kronecker delta in the second equation. The density of a sphere of radius R placed randomly in the Universe will be equal to the average density of the Universe, so its average overdensity is zero, but the variance of the overdensity is given by

$$
\sigma^2(R) \approx \Delta^2\left(\frac{1}{R}\right).
\tag{4.16}
$$

It follows that this variance increases as the radius R of the sphere gets smaller, since $\Delta^2(k)$ is higher at large values of k (i.e. at smaller scales).

4.1.4 Linear matter power spectrum

From the previous section, it is shown that linear perturbation theory can describe the evolution of the overdensity field in the matter dominated Universe. As described in previous chapters and Figure 2.5, the Universe went through three phases: radiation-dominated era during the early times, matter-dominated era after photon decoupling, and the current dark energy dominated era. To describe the matter power spectrum in each of these phases, we need to consider the interaction between matter and other components such as radiation and dark energy and evolve the full fluid equation. All of these components will contribute to structure formation. During the radiation-dominated era, the baryons are tightly coupled to the photons and this prevented the growth of structures. In addition, dark matter perturbations continued to grow during this era. After recombination, baryons were attracted to the gravitational wells created by dark matter overdensities and formed galaxies and large scale structures. The perturbations due to the sound waves in the photon-baryon fluid leave an imprint on the power spectrum.

We need to relate the matter power spectrum $P_{\mathrm{m}}(k) \propto \langle|\delta(k)|^2\rangle$ to the gravitational potential perturbation power spectrum $P_\Phi(k) \propto \langle|\Phi(k)|^2\rangle$. This is done by using the Poisson equation to relate the gravitational potential and the matter overdensity field as $\delta(k) \propto k^2\Phi(k)$. Thus, the predicted power spectrum follows

$$P_{\mathrm{m}}(k, z) \propto k^{n_s}, \tag{4.17}$$

where the spectral index n_s is one of the main cosmological parameters in the ΛCDM model. The value for n_s from current Planck CMB experiment is 0.968 ± 0.006 [5], a slight deviation from unity due to inflation.

We also need to consider the effect of the growth of the Hubble horizon ($\propto H^{-1}$) on different modes. At very early times, the relevant modes are outside the Hubble horizon. As the modes enter the horizon, they will start to grow with growth rates that depend whether we are in the radiation or matter dominated eras. This is characterized by the transfer function $T(k, z)$. The final linear power spectrum has the following form:

$$P_{\mathrm{m}}(k, z) \propto D^2(z) T^2(k, z) k^{n_{\mathrm{s}}}. \tag{4.18}$$

The shape of the linear matter power spectrum is very sensitive to cosmological parameters. We can solve for the power spectrum by including all different cosmological components using numerical Boltzmann codes such as `CAMB` [78], `CLASS` [24], or `CMBAns` [39].

In Figure 4.2, we show the linear matter power spectrum at $z = 0$ obtained from different cosmological probes, with the best fit model in solid black line [6]. We see that most of the cosmological measurements over a wide range of scales, or wavenumber k, agree very well with the theoretical prediction from the standard ΛCDM model. We can also measure the power spectrum at redshifts other than $z = 0$ with galaxy redshift surveys and weak gravitational lensing studies.

4.1.5 Galaxy power spectrum

In practice, we do not really observe the total matter power spectrum, which contains information on baryons and dark matter together. Most astronomical observations

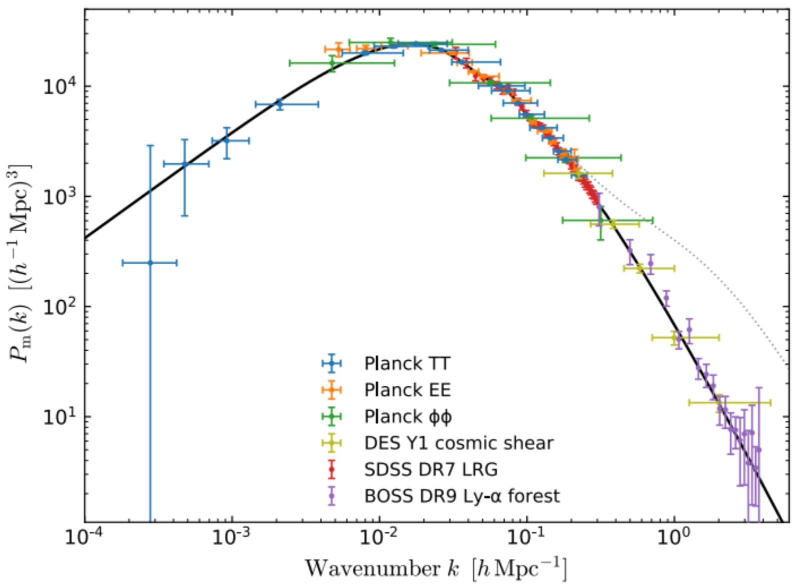

Figure 4.2: Linear matter power spectrum at $z = 0$ inferred from different cosmological probes with the best fit model (solid black line) and the impact of non-linear clustering at $z = 0$ (dotted black line). Plot from Planck 2018 result [6].

observe the tracers of the underlying matter fields. These matter tracers are galaxies, gas in emission or absorption lines in quasar spectra, etc. We do not yet know the exact relationship between the tracers and the underlying matter field, but we can follow a simple assumption that the galaxies' power spectrum is proportional to the matter power spectrum by a factor called the linear galaxy bias b_g:

$$P_{\rm g}(k, z) = b_{\rm g}^2(z) P_{\rm m}(k, z). \tag{4.19}$$

This linear approximation works well at large scales. However, at smaller scales, the effect of non-linearity becomes important and this approximation no longer holds true.

4.1.6 Redshift space distortions

According to Hubble's Law, objects recede from us at speeds which are proportional to their distance from us. In real observations, we usually measure the redshift of astrophysical objects instead of the real space comoving coordinate $\mathbf{x}$. In addition, they also have peculiar velocities, which are the components of an object's velocity that deviate from the Hubble flow. If the object has a peculiar velocity that is pointing toward us, we will measure a lower redshift and it would appear that the object is closer to us than it actually is. The opposite is true when the object is moving away from us. This Doppler shift causes a redshift-space distortion in which the spatial distribution of galaxies appears distorted when their positions are plotted as a function of their redshift rather than as a function of their distance.

On smaller scales (higher k modes), such as around dark matter halos, the effect of the random peculiar velocities is more substantial. This randomness will decrease the correlation between two points at these scales, so the power at high k is small. On the contrary, the random peculiar velocities create less of an effect at larger scales (lower k modes) and the overall flow of matter is more important, for example, the flow of galaxies into overdense regions to form galaxy clusters. The two-point correlation is thus higher at larger scales (low k) compared to smaller scales (high k).

Next, we need to know how the power spectrum is different in redshift space. Differentiating the proper distance $\mathbf{r}(t) = a(t)\mathbf{x}(t)$ with respect to time, we get

$$\mathbf{v}(t) = \frac{\mathrm{d}\mathbf{r}(t)}{\mathrm{d}t} = H(t)\mathbf{r}(t) + a(t)\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = H(t)\mathbf{r}(t) + \mathbf{u}(t), \tag{4.20}$$

where $\mathbf{u}(t) = a(t)\dfrac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}$ is the peculiar velocity field. In the radial direction, $v_z = Hr_z + u_z = H(r_z+u_z/H)$, where $\hat{\mathbf{z}}$ points in the radial direction. We will use a new coordinate system in redshift space that takes into account the peculiar velocity

$$\mathbf{s} = \mathbf{r} + u_{\hat{\mathbf{z}}}\hat{\mathbf{z}}/H. \tag{4.21}$$

where the letter s denotes quantity in redshift space and the letter r denotes quantity in real space. Using the continuity equation 4.2 and assuming the peculiar vector field is curl-free, the velocity field has the form

$$v_{\mathbf{k}} = -i\frac{aH}{k}f\delta_{\mathbf{k}}, \tag{4.22}$$

where the growth rate f is defined by

$$f \equiv \frac{d\ln\delta}{d\ln a}. \tag{4.23}$$

The number of particles in both redshift and real space must be conserved. The redshift space overdensity is related to the real space overdensity by

$$\delta_s(\mathbf{k}) = \left(1 + \beta(z)\mu^2\right)\delta_r(\mathbf{k}). \tag{4.24}$$

where $\mu = \mathbf{k}_{\|}/\mathbf{k}$ is the cosine of the angle between $\mathbf{k}$ and the line-of-sight direction, $\beta = f/b$ in which b is the bias factor of the tracer. All in all, the matter power spectrum in redshift space is given by [66, 59]:

$$P_s(k,\mu,z) = \left(1 + \beta(z)\mu^2\right)^2 P_r(k,z). \tag{4.25}$$

At smaller scales, non-linear effects becomes more pronounced. In addition, random peculiar velocities along the line-of-sight on smaller scales flatten the power spectrum in the radial directions. The galaxy distribution is elongated in the redshift space that points directly back at the observer. This is known as the Fingers of God effect. A Gaussian streaming model is used to account for this [59]:

$$F\left(k, \mu, \Sigma_{\mathrm{s}}\right) = e^{-\left(k\mu\Sigma_{\mathrm{s}}\right)^2}, \tag{4.26}$$

where Σ_s is the streaming scale that describes the random peculiar velocities along the line-of-sight direction which contribute to the suppression of the power spectrum at smaller scales. Putting this altogether, the power spectrum with redshift has the form:

$$P_s(k, \mu, z) = \left(1 + \beta(z)\mu^2\right)^2 F\left(k, \mu, \Sigma_{\mathrm{s}}\right) P_r(k, z). \tag{4.27}$$

The redshift space power spectrum contains anisotropy, in contrast to the real space power spectrum. This information can be used to measured the growth rate f. The power spectrum can be decomposed into Legendre polynomials $L_\ell(\mu)$:

$$\begin{aligned} P(k, \mu, z) &= \sum_{\ell=0}^{\infty} P_\ell(k, z) L_\ell(\mu) \\ P_\ell(k, z) &= \frac{2\ell + 1}{2} \int_{-1}^{1} P(k, \mu, z) L_\ell(\mu) \mathrm{d}\mu. \end{aligned} \tag{4.28}$$

The correlation function can then be expressed as

$$\xi_\ell(r) = \frac{i^\ell}{2\pi^2} \int k^3 \mathrm{d}(\log k) P_\ell(k) j_\ell(kr), \tag{4.29}$$

where j_ℓ is the spherical Bessel function of ℓ^{th} order. All the odd ℓ terms are zero. Without the Fingers of God effect, the only nonzero terms are the $\ell = 0$ (monopole),

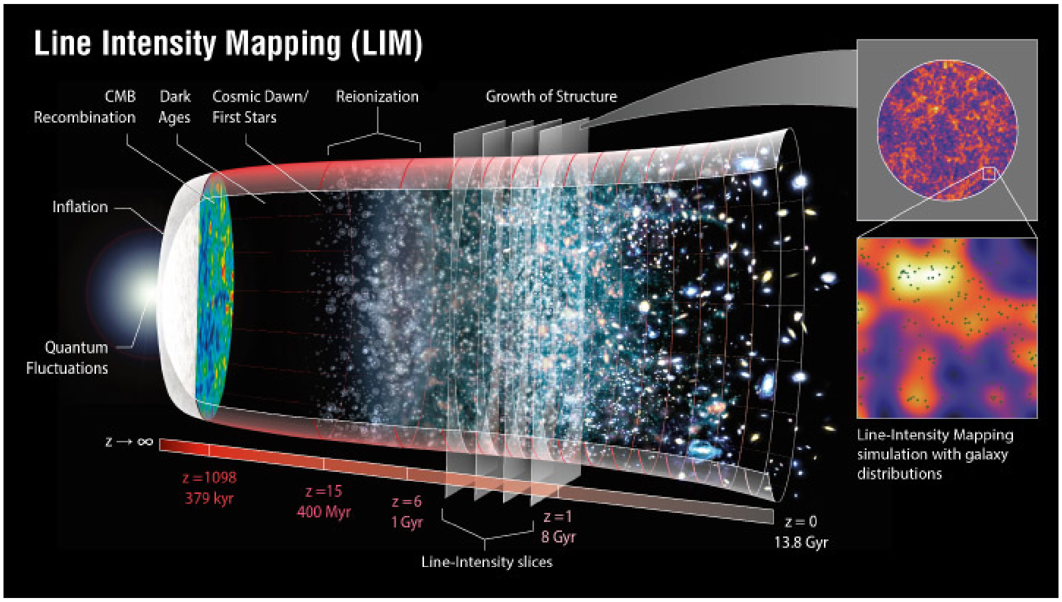

Figure 4.3: Line Intensity Mapping taking snapshot slices of the growth of structure at different redshifts. Plot from the NASA/LAMBA archive.

$\ell = 2$ (quadrupole), and $\ell = 4$. With this effect taken into account, the higher order even ℓ terms are no longer zero.

4.2 Hydrogen

Hydrogen is the most abundant element in the Universe, and we can study the distribution of hydrogen as a tracer of matter clustering into galaxies, star formation, and the phase structure of the interstellar medium. Following the decoupling of CMB photons, the Universe was dark and transparent. There were few ionization sources to create more photons, and the existing photons were free to propagate. This period is know as the Dark Ages. Small quantum fluctuations from inflation generated seeds for gravitational wells, or over-densities of matter. When these overdensities grew sufficiently large, fusion turned on, thus creating stars and galaxies. This happens at redshift $700 < z < 20$. Radiation from these new bright sources reionized the hydrogen residing in the intergalactic medium (IGM). This era is referred to as the Epoch of Reionization (EoR) ($6 < z < 20$). A timeline of this process is shown in Figure 4.3.

The early Universe contains many phenomena which are not well understood. By looking at the highly redshifted 21 cm emission from neutral hydrogen, we could further understand the physics of the early Universe, such as the nature of dark matter, dark energy and inflation. The 21 cm (or 1420 MHz) radiation comes from the transition between the two levels of the hydrogen in the ground state, very slightly split by the interaction between the electron spin and the nuclear spin, and we could use this radiation to map out the distribution of the neutral hydrogen, which strongly correlates with dark matter halos in the ΛCDM model. The 21 cm radiation penetrates the dust clouds, and this is an advantage over observations in visible light. This tomographic mapping of the redshifted hyperfine transition of neutral hydrogen (HI) is known as intensity mapping, which has the potential to map out the large-scale structure of the Universe over a wide redshift range.

Unlike the cosmic microwave background (CMB), which only provides a two - dimensional map of the the last scattering surface, HI intensity mapping can in principle be used to make three-dimensional maps of matter at all redshifts up to $z \approx 50$, even before galaxies formed, albeit with some complications in practice. By looking at the collective emission from many galaxies or galaxy clusters without resolving individual galaxies, HI intensity mapping can map the large scale structure. In addition, the redshifted 21 cm line can be used to map the Universe at different redshifts by tuning the receivers to a range of frequencies. At lower redshift, 21 cm tomography enables intensity mapping of self-shielded gas within galaxies, allowing for precise measurements of baryon acoustic oscillations (BAO), which can be used as a standard ruler for measuring cosmological distance. At higher redshifts, 21 cm tomography allows us to probe the cosmic dawn, the time at which the first stars and black holes were formed at the beginning of the Epoch of Reionization (EoR). Both the sky-averaged 21 cm brightness temperature and its fluctuations store important information about these processes. In principle, 21 cm intensity mapping allows us to probe most of the observable Universe from the dark ages until the present day, as shown in Figure 4.4. Furthermore, intensity mapping also complements CMB observation by mapping the effect of gravitational lensing on the CMB. The gravitational lensing potential can mix the E-modes polarization into B-modes. By mapping the deflection potential with intensity mapping, we can derive an expectation for the lensed B-modes from the measured E-modes and the transfer between E- and B-modes. In principle, the lensed B-mode can be subtracted from the CMB polarization maps to search for B-modes from gravitational waves with the hope to learn more about the inflationary period at the very beginning of the Universe.

Observation of the HI distribution can constrain the history of the EoR, star formation, galaxy assembly, and the statistical properties of large scale structure, from which we can study the nature of dark energy, dark matter, and the inflationary origin of the Universe. The power spectrum of intensity mapping contains cosmological information on the matter distribution and helps us understand galaxy evolution by tracing the HI content of galaxies at different redshifts and the scale-dependence of HI clustering. The influx of astronomical data in the last decade has greatly increased our understanding of these phenomena. Indeed, galaxy redshift surveys, such as LSST (optical imaging), WFIRST (NIR imaging & slitless spectroscopy), and DESI (optical spectroscopy), will characterize the nature of dark energy with very high precision.

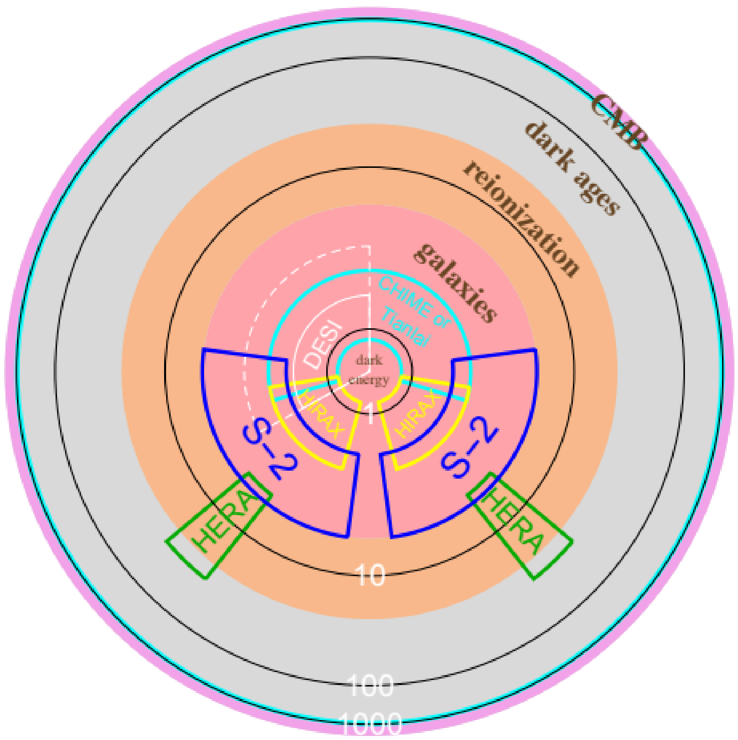

Figure 4.4: Plotted is a 2D representation of the observable Universe in which the area is proportional to the comoving volume and the distance from the center increases monotonically with distance from Earth. The volume that can be accessed using the 21 cm intensity mapping of neutral hydrogen is within the orange region and a little bit of the dark ages. The outermost thin shell at $z = 1100$ is the CMB. The enclosed cyan region is the survey volume of the CHIME and Tianlai 21 cm observatories. Figure from [14]

However, optical surveys are expensive to expand to a large volume. Additionally, these experiments leave a lot of the post-reionization Universe with redshift $z < 6$ uncovered; the new technique of intensity mapping offers the capability to fill in this void.

So far, the 21 cm signal has been detected with three intensity mapping instruments: the Green Bank Telescope (GBT) [29, 127, 93], the Parkes Observatory [13] and CHIME by cross-correlating intensity maps with galaxy redshift surveys.

4.2.1 The Epoch of Reionization

When the first galaxies were formed, they emitted ultraviolet (UV) radiation and ionized the sourrounding IGM. As a consequence, the intermediate regions of the IGM also became ionized and transparent to ultraviolet photons, which means that the surrounding hydrogen could no longer radiate 21 cm emission. This era is referred to as the Epoch of Reionization (EoR). This UV radiation continued to propagate beyond the over-densities where it originated, creating large spherical bubbles. These bubbles continued to grow until they overlapped with one another, and eventually they spread through the whole IGM. What is left today is a fully ionized IGM with small pockets of neutral gas residing in galaxies (HI clouds) or external to galaxies as part of intercloud gas. However, the details of this process are still not known, such as when reionization began, how long it took the IGM to fully ionize, and how reionization affected galaxy formation.

Prior CMB observations have placed constraints on the beginning of reionization. Since the CMB radiation is older (around 400,000 years after the Big Bang), it serves as a backlight to the EoR. Using a model in which reionization was instantaneous, and the optical depth can be translated into reionization redshift, WMAP constrained the

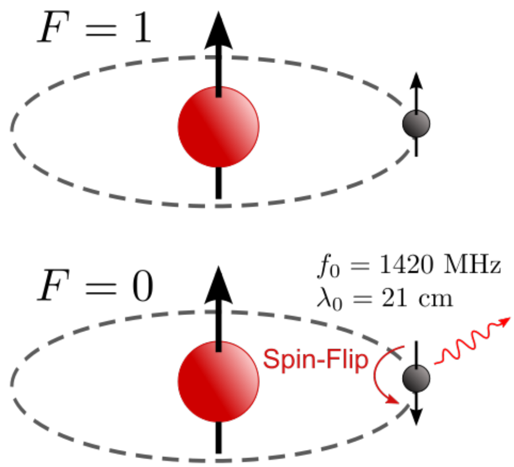

Figure 4.5: Hydrogen spin-flip emission. Both electron and proton spins are 1/2, so there are two possible states. The higher energy symmetric state (spins parallel) can can spontaneously transition to the lower energy anti-symmetric state (spin antiparallel), emitting a photon with wavelength 21 cm.

reionization redshift to $z_{re} = 10.6 \pm 1.1$ (Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. 2014). Results from the Planck satellite, which launched in 2009, constrained the redshift of reionization at $z_{re} = 8.8^{+1.7}_{-1.4}$. (Planck 2015 results. XIII. Cosmological parameters.)

Observations of highly redshifted quasars can also place upper bounds on the redshifts by which reionization is complete. Beardsly, Fan et al. (24) demonstrated that reionization must be completed by $z_{re} \approx 6$. From observing a steep decline in Lyman-α emitting galaxies, Choudhury et al. (17) set the reionization redshift in the range $6 \lesssim z_{re} \lesssim 8$. Future deep optical and infrared galaxy surveys, such as the James Webb Space Telescope, will allow us to conduct very sensitive studies to constrain reionization.

4.2.2 21 cm Intensity Mapping

The 21 cm emission is a very narrow spectral line and is created by the spontaneous transition between the two energy levels of the hydrogen atom in the 1s ground state, slightly split by the interaction between the nuclear spin and the electron spin, as shown in Figure 4.5. This is known as the hyperfine structure. The splitting energy is only 5.9×10^{-6} eV, only about four parts in ten million compared to the ground state energy -13.6 eV. This spontaneous emission has an extremely small transition rate of $2.9 \times 10^{-15} s^{-1}$, and a mean lifetime of the excited state of about 10 million years. However, due to the sheer amount of hydrogen in the Universe, we can observe this transition line.

By studying 21 cm intensity mapping, we can can a three dimensional, tomographic map of the Universe. The advantages of 21 cm intensity mapping are numerous. First, hydrogen makes up the majority of baryonic matter. Second, the hydrogen flipping transition is very narrow and well understood. Third, the observed redshift of this transition maps directly to the scale factor at the time of emission. With a model, this redshift can be converted to a line of site distance.

The expected 21 cm brightness temperature is on the order of a few mK, whereas the foreground signals can range from hundreds to thousands of Kelvin. It is difficult for early generation telescopes to produce high quality 21 cm maps to see the individual structures. However, we can we can infer the global properties of the 21 cm signal via the power spectrum, which encodes the cosmological information in its shape. Specifically, the 21 cm power spectrum allows us to measure the position of the BAOs, which can be used as a standard ruler to tells us about the scale of the Universe vs redshift. This in turn tells us about the time evolution of dark energy.

The Fourier transform of the spatial 21 cm brightness temperature is defined as

$$\widetilde{T}_{21}(\mathbf{k}) = \mathcal{F}[\delta T_{21}(\mathbf{r})] = \int \mathrm{d}^3\mathbf{r}\delta T_{21}(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}. \tag{4.30}$$

where the integral is over all space, and $\mathbf{k}$ is the three dimensional wavenumber Fourier dual to the position vector $\mathbf{r}$. The 21 cm power spectrum is defined as

$$\left\langle \widetilde{T}_{21}(\mathbf{k})\widetilde{T}^{*}_{21}(\mathbf{k}')\right\rangle \equiv (2\pi)^3\delta_D(\mathbf{k}-\mathbf{k}')\,P_{21}(\mathbf{k}), \tag{4.31}$$

where the angle bracket $\langle\rangle$ denotes the ensemble average over many realizations of the Universe, and δ_D is the Dirac delta function, but in reality we can observe only one realization of the Universe. Since the Universe is isotropic, the power spectrum only depends on the magnitude k ($P_{21}(\mathbf{k}) \longrightarrow P_{21}(k)$). In practice, we decompose k space into bands in $k_\parallel$ and $k_\perp$, and we can average the measured power spectrum in spherical shells along the line of sight axis in the full three-dimensional k space.

To extract information from 21 cm intensity mapping, we need to compare the observational data with the theoretical predictions. To linear order, the amplitude and shape of the 21 cm power spectrum at redshift z is proportional to the product of the HI bias $b_{\mathrm{HI}}(z)$ and its cosmic abundance $\Omega_{\mathrm{HI}}(z) = \rho_{\mathrm{HI}}(z)/\rho_c(z=0)$, where $\rho_{\mathrm{HI}}(z)$ is the mean HI density at redshift z and $\rho_c(z=0)$ is the critical density at $z=0$. The assumption here is that HI acts as a biased tracer of the dark matter distribution, and the 21 cm power spectrum is related to the linear matter power spectrum as follows [131]:

$$P_{\mathrm{HI}}(k,z,\mu) = \bar{T}_b(z)^2\left[\left(b_{\mathrm{HI}}(z)+f(z)\mu^2\right)^2 P_m(k) + P_{\mathrm{SN}}(z)\right], \tag{4.32}$$

where $b_{\mathrm{HI}}(z)$ is the HI bias factor, P_{SN} is the HI shot noitse, $\bar{T}_b$ is the mean brightness temperature increase relative to the CMB and is given by [22, 131]:

$$\bar{T}_b = 189\,\mathrm{mK}\,\frac{H_0(1+z)^2}{H(z)}\Omega_{\mathrm{HI}}(z)h \tag{4.33}$$

The meaning of the brightness temperature will be discussed in the next section.

4.2.3 HI spin temperature

As discussed in the previous section, we calculate the power spectrum with the brightness temperature, which is the difference in brightness temperature between the 21 cm signal and the CMB. We are also interested in the intrinsic temperature of the neutral hydrogen gas. The intensity of the 21 cm radiation is characterized by its excitation, or spin temperature T_{spin}. To detect the 21 cm emission of a neutral hydrogen cloud, its temperature must be out of equilibrium with the background CMB (see Figure 4.6). The neutral hydrogen cloud will contribute a net emission if $T_{\mathrm{spin}} > T_{\mathrm{CMB}}$ or a net absorption if $T_{\mathrm{spin}} < T_{\mathrm{CMB}}$. To quantify the amount of background radiation being transmitted through the cloud, we need to start with the radiative transfer equation:

$$\frac{\mathrm{d}I_v}{\mathrm{d}\tau_v} = -I_v + B_v(T_{\mathrm{spin}}), \tag{4.34}$$

where I_ν is the specific intensity (or brightness), which is the energy carried by the radiation per unit area, frequency, solid angle, and time from a specific direction $\hat{n}$ and has the unit Janskys per steradian (Jy/str). τ_ν is the optical depth through the cloud and B_ν is the Planck distribution function. We can also convert Jy/str to mK by

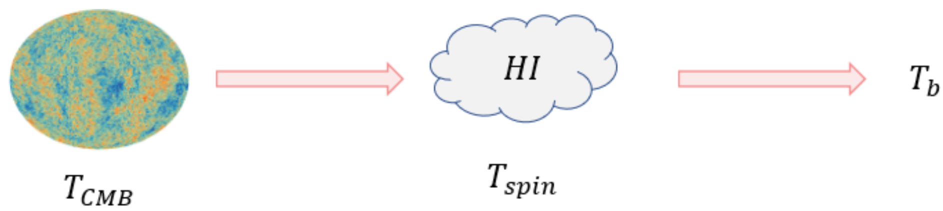

Figure 4.6: The CMB passing through a cloud of neutral hydrogen gas with spin temperature T_{spin} and emerging with the brightness temperature T_b which we measure.

$$1\text{mK} = 10^{-23}\frac{c^2\text{str}}{2\nu^2 k_B}\text{Jy/str}. \tag{4.35}$$

Integrating the above equation from 0 to τ_ν to solve for the specific intensity, we get

$$I_\nu = I_\nu^{\text{CMB}} e^{-\tau_\nu} + B_\nu(T_{\text{spin}})(1 - e^{-\tau_\nu}). \tag{4.36}$$

A specific intensity I_ν has a brightness temperature T_b, which is the equivalent blackbody temperature that would produce the same specific intensity $I_\nu = B_\nu(T_b)$. In the Rayleigh-Jeans low energy limit, the brightness temperature is related to the specific intensity by

$$B_\nu = \frac{2\nu^2 k_B T_b}{c^2} \longrightarrow T_b(\nu) = \frac{I_\nu c^2}{2k_B \nu^2}. \tag{4.37}$$

In terms of brightness temperature, equation 4.36 becomes

$$T_b(\nu) = T_{\text{CMB}}(\nu)e^{-\tau_\nu} + T_{\text{spin}}\left(1 - e^{-\tau_\nu}\right). \tag{4.38}$$

Therefore, to measure the brightness temperature, we need to know the optical depth τ_ν and the spin temperature T_{spin}. The first exponential term $e^{-\tau_\nu}$ gives the transmission probability of the background CMB radiation, and the term $(1 - e^{-\tau_\nu})$ gives the emission probability of the 21 cm photons in the HI cloud. In the special case where $T_{\text{spin}} = T_{\text{CMB}}$, because absorption equals emission at every frequency, the brightness temperature is exactly the CMB temperature, and this does not reveal any information about the cloud.

The brightness temperature at redshifts other than $z = 0$, due to the Hubble expansion, is

$$T_b(\nu) = \frac{T_{\text{CMB}}(\nu)}{(1+z)} e^{-\tau_\nu} + \frac{T_{\text{spin}}}{(1+z)} \left(1 - e^{-\tau_\nu}\right), \tag{4.39}$$

where τ_ν is the optical depth of the 21 cm frequency. We need to compute this optical depth τ_ν, which is the absorption coefficient integrated along the proper length of the photons' path. The hyperfine flipping transition is determined by the absorption rate of 21 cm photons ($|0\rangle \to |1\rangle$ transition), characterized by the Einstein coefficient B_{01}, and the stimulated emission of 21 cm photons ($|1\rangle \to |0\rangle$ transition), characterized by the Einstein coefficient B_{10}. These Einstein coefficients are given by [116]:

$$\begin{aligned} I_\nu B_{01} &= \frac{g_1}{g_0} B_{10} I_\nu \\ I_\nu B_{10} &= A_{10} \frac{\lambda^2 I_\nu}{2h\nu_{10}} \end{aligned} \tag{4.40}$$

where $\nu_{10} = 1420.4$ MHz is the frequency of the hyperfine 21 cm transition, and $A_{10} = 2.85 \times 10^{-15}\ \text{s}^{-1}$ is the spontaneous emission rate. The absorption cross section of the 21 cm line is

$$\sigma_\nu \equiv \sigma_{01}\phi(\nu) = \frac{3c^2 A_{10}}{8\pi\nu^2}\phi(\nu), \tag{4.41}$$

where $\phi(\nu)$ is the line profile of the 21 cm transition, which includes natural, thermal, turbulent, and velocity broadening. It is normalized such that $\int \phi(\nu)d\nu = 1$. Integrating this over the line of sight from the observer to the source to get the optical depth:

$$\tau_\nu = \int \sigma_{01} \left(1 - e^{-E_{10}/k_B T_{\rm spin}}\right) \phi(\nu) n_0 \, \mathrm{d}s, \tag{4.42}$$

where n_0 is the number density of the lower energy singlet state and ds is the proper length element and can be converted to the redshift by

$$\mathrm{d}s = a\mathrm{d}r = c\mathrm{d}t = c\frac{a}{\mathrm{d}a/\mathrm{d}t}\frac{\mathrm{d}a}{a} = -\frac{c\mathrm{d}z}{(1+z)H(z)}. \tag{4.43}$$

The result of the integral for the 21 cm optical depth is

$$\begin{aligned} \tau_{\nu_0} &= \frac{3}{32\pi}\frac{hc^3 A_{10}}{k_B T_{spin}\nu_0^2}\frac{\mathrm{x}_{HI} n_H}{(1+z)\left(\mathrm{d}v_\parallel/\mathrm{d}r_\parallel\right)} \\ &\approx 0.0092(1+\delta)(1+z)^{3/2}\frac{\mathrm{x}_{HI}}{T_{\rm spin}}\left[\frac{H(z)/(1+z)}{\mathrm{d}v_\parallel/\mathrm{d}r_\parallel}\right], \end{aligned} \tag{4.44}$$

where $(1+\delta)$ is the baryon fractional overdensity and d$v_\parallel$/d$r_\parallel$ is the gradient of the proper velocity along the line of sight. Another unknown quantity in Equation **??** is the spin temperature. For a system in thermal equilibrium, the relative occupation of two different energy states is given by

$$\frac{n_1}{n_0} = \frac{g_1}{g_0}\exp\frac{-\Delta E}{kT}, \tag{4.45}$$

where n_1 and n_0 are the number densities of electrons in the higher energy symmetric state and the lower energy symmetric state, respectively. In the context of 21 cm emission, this equation reduces to

$$\frac{n_1}{n_0} = 3 \exp\left(-\frac{T_*}{T_{\rm spin}}\right), \tag{4.46}$$

where $T_* = 0.0681$ K is the temperature corresponding to the 21 cm emission. We see that the spin temperature is related to the ratio of the number densities between the two energy levels, and this tells us about the intensity of the 21 cm radiation from a cloud of neutral hydrogen. However, we need to take into account different background radiation being transmitted through the cloud and quantify the emission and absorption within the cloud. Three main processes determine T_s: (1) The spin temperature T_s is coupled to the IGM kinetic temperature T_k, (2) coupling with high energy radiation (Lyman-α photons), and (3) absorption and induced emission of CMB photons. The spin temperature can be written as

$$T_{\rm spin} = \frac{T_{\rm CMB} + y_k T_k + y_\alpha T_\alpha}{1 + y_k + y_\alpha}, \tag{4.47}$$

where y_k and y_α are the kinetic and Lyman-α coupling terms, respectively. y_k is determined by collisional excitation with other hydrogen atoms, free electrons, and protons in the IGM. y_α is determined by the Lyman-α pumping mechanism (Wouthuysen–Field coupling), as neutral hydrogen absorbs and then re-emits Lyman-α photons, and may enter either of the two spin states. $T_{\rm spin}$ needs to be different from $T_{\rm CMB}$ in order for the 21 cm to be detected. This formula can be used to calculate the spin temperature, and the strength of each coupling term can be found in most physics textbooks.

Figure 4.7 shows the the evolution of spin, CMB, and gas temperatures as a function of redshift. At early times (redshift beyond $z \approx 200$), the spin temperature is coupled to the kinetic temperature of the hydrogen gas in the IGM and the CMB, and they all fall off as $(1 + z)$. During this period, the IGM gas is coupled to the CMB photons

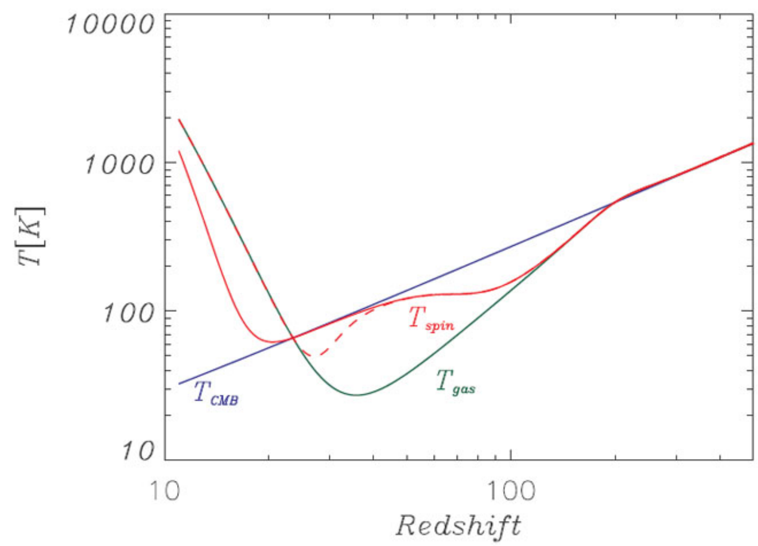

Figure 4.7: The evolution of spin, CMB, and gas temperatures (in red, blue, and green, respectively) as a function of redshift. Figure from [139].

via Compton scattering. At redshift $z \approx 200$, the gas temperature briefly decouples from the CMB temperature and decreases adiabatically as $(1+z)^2$ since the gas is non-relativistic. This is due to the Hubble expansion of the Universe and there are no heating sources. Once ionizing sources such as X-ray binaries form, the gas starts heating up at redshift below 30.

The spin temperature behavior is more complex. At redshift beyond 100, it is still coupled to the gas temperature due to electron collisions. However, the strength of this coupling decreases due to Hubble expansion, and the spin temperature trends toward the CMB temperature, creating an absorption region. At redshift below 15, the ionizing sources couple the spin temperature to the gas again and $T_{\text{spin}} \gg T_{\text{CMB}}$. In the middle where there is a dip, there are two possible scenarios. In the first scenario, the spin temperature couples with the gas temperature as it heats up beyond the CMB temperature (red solid line) in Figure 4.7. This is mostly due to emission. In the second scenario, the spin temperature couples with the gas temperature before it reaches the CMB temperature (red dashed line). This is due to initial absorption and later emission.

In actual observations, we are interested in the deviation of the HI signal from the CMB temperature, and the differential brightness temperature $\Delta T_b \equiv T_b - T_{\rm CMB}$ is what we need:

$$\Delta T_b = (28\,\mathrm{mK})\,(1+\delta)x_{\rm HI}\left(1 - \frac{T_{\rm CMB}}{T_{\rm spin}}\right)\left(\frac{\Omega_b h^2}{0.0223}\right)\sqrt{\left(\frac{1+z}{10}\right)\left(\frac{0.24}{\Omega_m}\right)}\left[\frac{H(z)/(1+z)}{\mathrm{d}v_\parallel/\mathrm{d}r_\parallel}\right] \tag{4.48}$$

where h is the Hubble constant, Ω_m and Ω_b are the matter and baryon density parameters, and $(1+\delta)$ is the fractional baryon overdensity. The term $(1 - T_{\rm CMB}/T_{\rm spin})$ is positive when the HI cloud is in emission and negative when it is in absorption. Usually we write T_b instead of ΔT_b when the context is understood (i.e. an increase in brightness temperature over the CMB). From this equation, we see that at high redshifts when the fraction of neutral hydrogen is high ($x_{\rm HI}$ is close to unity), ΔT_b probes the density fluctuation. On the other hand, at lower redshift, ΔT_b probes the neutral and ionized regions. Figure 4.8 shows the time evolution of the brightness temperature from the dark ages to the end of the ionization epoch. .

4.2.4 HI in galaxies halos

One prominent analytical model that describes the dark matter distribution is the halo model. It assumes that dark matter is partitioned over halo building blocks, which are spherical and have a density distribution that depends only on halo mass. In this model, dark matter halos permeate and surround individual galaxies as well as clusters of galaxies (see Figure 4.9). Dark matter halos above a certain mass threshold

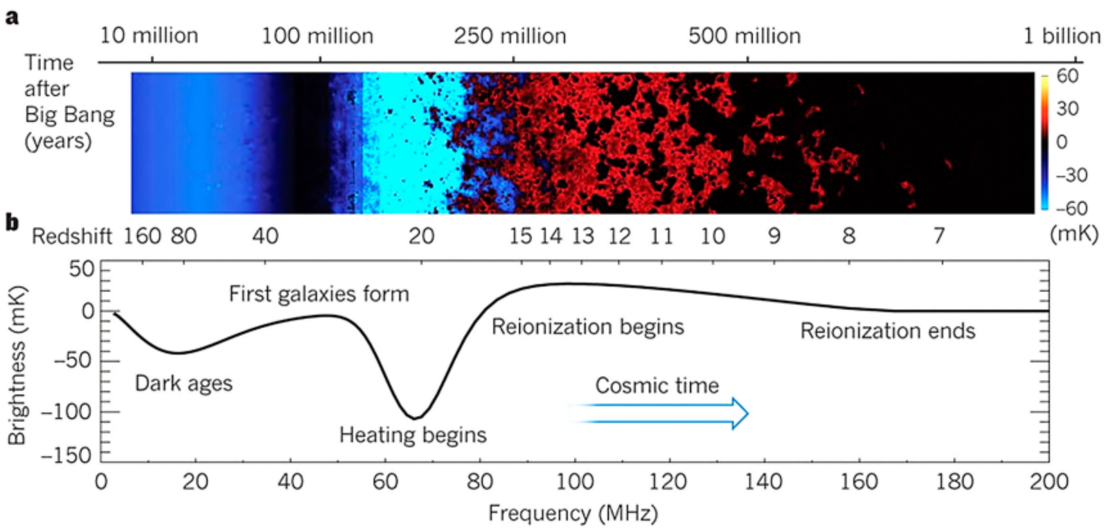

Figure 4.8: Reionization history: Time evolution of the brightness temperature from the dark ages through the end of the reionization epoch. The brightness temperatures in the dark ages and when heating begins are negative due to absorption. When the first galaxies form, the spin temperature decouples from the gas temperature. The brightness temperature is positive at $z \approx 13$ due to X-ray heating and then decreases due to reionization and finally becomes zero when most of the hydrogen is ionized. Figure from [111].

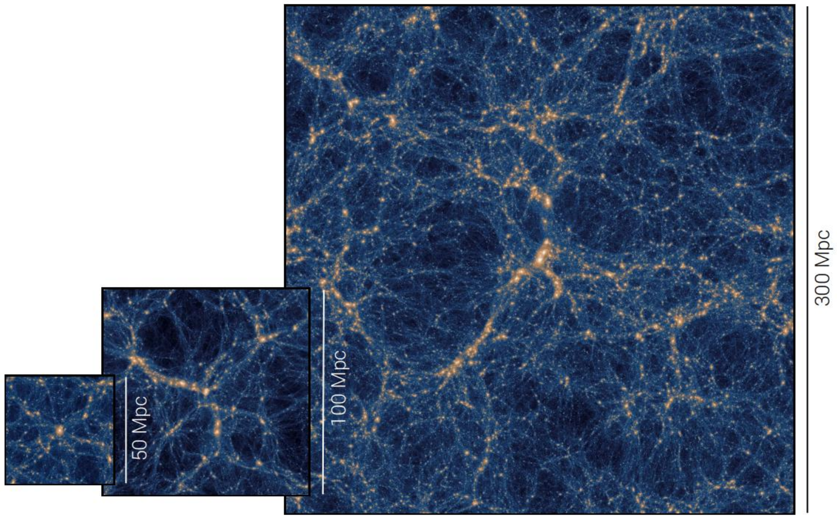

Figure 4.9: The large-scale structure from a ΛCDM gravo-magnetohydrodynamical simulations named Illustris at different scale from left to right [98]. The color indicates the dark matter density, with the bright spots being central or satellite galaxies.

have at least one central galaxy at the center. Higher mass halos contain additional satellite galaxies within the same halos. The clustering of galaxies coming from pairs of galaxies in separate halos (typically on the scale of 1-2 Mpc/h) is called the 2-halo term. Clustering of galaxies coming from the same central parent halo (on a scale less than 1 Mpc/h) is called the 1-halo term.

The halo model can predict the abundance and distribution of HI and the shot-noise. We will need a relationship between the total halo mass M_h and the total HI mass inside the halo M_{HI}, as well as the density distribution function of HI within each halo $\rho_{\mathrm{HI}}(r|M_h) = \rho_0 \exp(-r/r_s)$, where r_s is the scale radius of the halo.

We assume that the amount of HI in a halo depends only on its mass, and the relationship between M_{HI} and M_h is as follows [131, 28]:

$$M_{\mathrm{HI}}\left(M_h\right) = M_0 \left(\frac{M_h}{M_{\mathrm{min}}}\right)^{\alpha} \exp\left(-\frac{M_{\mathrm{min}}}{M_h}\right), \tag{4.49}$$

where α is a scaling factor between the total HI mass $M_{\rm HI}$ and the total halo mass M_h, and $M_{\rm min}$ is the minimum threshold mass cut-off for a halo to host HI. In addition, the mass density of HI is given by

$$\Omega_{\rm HI} = \frac{1}{\rho_{\rm c}} \int_0^\infty n\left(M_h\right) M_{\rm HI}\left(M_h\right) \mathrm{d}M_h, \tag{4.50}$$

where $n(M_h)$ is the halo mass function. The HI bias in the halo model is

$$b_{\rm HI}(z) = \frac{\int_0^\infty b\left(M_h, z\right) n\left(M_h, z\right) M_{\rm HI}\left(M_h, z\right) \mathrm{d}M_h}{\int_0^\infty n\left(M_h, z\right) M_{\rm HI}\left(M_h, z\right) \mathrm{d}M_h}, \tag{4.51}$$

where $b\left(M_h, z\right)$ is the bias of halos of mass M_h at redshift z. The halo model for the HI power spectrum at a given redshift is related to the 1-halo and 2-halo terms as follows [105, 131]:

$$\begin{aligned}
&P_{\rm HI}(k) = P_{\text{HI, 1-halo}}(k) + P_{\text{HI, 2-halo}}(k) \\
&P_{\text{HI, 1-halo}}(k) = F_2^0(k) \\
&P_{\text{HI, 2-halo}}(k) = P(k)\left[F_1^1(k)\right]^2 \\
&F_\beta^\alpha(k) \equiv \int_0^\infty n\left(M_h\right) b^\alpha\left(M_h\right)\left[\frac{M_{\rm HI}\left(M_h\right)}{\bar{\rho}_{\rm HI}} u_{\rm HI}\left(k|M_h\right)\right]^\beta \mathrm{d}M_h,
\end{aligned} \tag{4.52}$$

where $u_{\rm HI}\left(k|M_h\right)$ is the normalized Fourier transform of the HI density profile and is given by

$$u_{\rm HI}(k|M_h) = \frac{4\pi}{M_{\rm HI}(M_h)} \int_0^{R_{\rm v}} \rho_{\rm HI}(r) \frac{\sin kr}{kr} r^2 \ \mathrm{d}r. \tag{4.53}$$

At $k = 0$, $b_{\rm HI} = F_1^1(k = 0)$, and this means at large scales the HI power spectrum is mostly determined by the 2-halo term (the first term in Equation 4.32). In contrast, the normalized HI density profile is constant at low k, so the 1-halo term acts as a constant contribution to the total HI power spectrum. The discrete nature of the HI sources at the largest scales contribute to the shot noise $P_{\rm SN} = F_2^0(k=0)$ of the power spectrum in Equation 4.32. It has been shown that the shot noise is not a significant contribution to the power spectrum, regardless of the parameters $\alpha, M_{\rm min}$. This makes the HI power spectrum a fundamentally high signal-to-noise ratio measurement, barring other sources of noise. On the other hand, at high k, or smaller scale, the clustering of HI is determined by the 1-halo term, or the HI density profile. Current intensity mapping experiments can benefit from cross-correlation with galaxy surveys, such as ALFAFA [56], which is an extragalactic HI survey at redshift 0, and SDSS [53], which is an optical galaxy survey.

4.2.5 Foregrounds

Becase the 21 cm signal is notoriously weak, coupled with the fact that the foreground is 10^5 times larger, this requires extremely sensitive instruments to observe it. For extracting the 21 cm signal, we rely on the fact that the foreground emission is a smooth function of frequency, while the 21 cm spectrum has structure arising from the large scale distribution of matter along the line of sight. Instrumental effects can introduce structure into the otherwise smooth foregrounds. Specifically, the angular component of the antenna beam pattern is frequency dependent, and through a process called mode-mixing, introduces frequency structure into the smooth foregrounds which can be confused with cosmic 21 cm structure. In addition, even though the 21 cm signal is unpolarized, the bright foreground is partially polarized, and Faraday rotation in the interstellar medium creates additional spectral structure in the polarization signal. Furthermore, frequency-dependent instrumental leakage of the Stokes Q, V, U into I brings another type of complicated spectrum into the signal. Removing this mode-mixing effect requires a detailed understanding of the frequency-dependent antenna beam pattern as well as the gain and phase of the instrument's electronics by calibration. (Shaw et al. 2015 2011.05946) show by simulation of the CHIME interferometer that, in the presence of foreground, it is necessary to know that beamwidth of the antennas to 0.1% and the electronic gain to 1% within each minute of observation to recover the unbiased power spectrum of the HI signal.

The foreground wedge and the EoR window

Foreground cleaning will inevitably throw away some valuable information about the sky and reduce the sensivity to the 21 cm signal. As we are determined to measure the 21 cm power spectrum, we need to understand how the physical instrument interacts with the foregrounds and 21 cm signal. There is a limited region of the $k_\perp k_\parallel$ Fourier space that we can access. The lowest $k_\perp$ mode is determined by the shortest baseline of the interferometer, while the highest $k_\perp$ mode is determined by the longest baseline, as shown in Figure 4.10. Similarly, the upper limit for $k_\parallel$ is set by the spectral resolution of the array. The lower limit for $k_\parallel$ is determined by cosmic variance in theory, but in practice it is determined by bright and spectrally smooth foregrounds which contaminate the lowest $k_\parallel$. The foregrounds present in the 21 cm EoR measurements are synchrotron emission from our own Milky Way galaxy, galactic and extra-galactice free-free emissions, and point sources. These foregrounds are many orders of magnitude brighter than the 21 cm signal. In addition, these sources are spectrally smooth, so after the 3D Fourier transform, they only occupy a few low $k_\parallel$ modes.

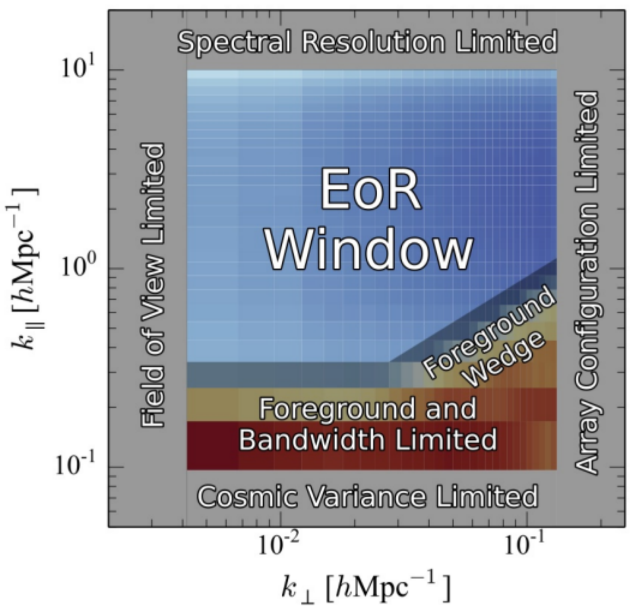

Figure 4.10: A schematic of the EoR window in the $k_\perp k_\parallel$ Fourier plane. The minimum and maximum $k_\perp$ modes accessible are determined by the shortest and longest baselines of the array, respectively. The maximum $k_\parallel$ mode is set by the frequency resolution of the array. The minimum $k_\parallel$ is determined by cosmic variance in principle, but in practice, it is determined by bright and spectrally smooth foregrounds which contaminate the lowest $k_\parallel$. These foregrounds out to higher $k_\parallel$ in a characteristic shape called the foreground wedge. The complimentary region is the EoR window, where a clean measurement of the 21 cm power spectrum is possible. cite: https://arxiv.org/pdf/1404.2596.pdf

The Foreground Wedge and the EoR window In radio interferometry, the baselines that sample specific $k_\perp$ have fixed absolute distances but these distances vary in the uv-plane as they are measured in terms of wavelength. This frequency-dependent baseline distances in the uv-plane introduce instrumental mode mixing. The effect is that the spectrally smooth foreground power leaks into higher $k_\parallel$, producing a wedge shaped region as seen in Figure 4.10. As a result, most of the large astrophysical foregrounds can be found in a wedge shaped region with $k_\parallel < \beta k_\perp$ for an experiment-dependent constant β. This is called the *foreground wedge*. The region outside of the foreground wedge (also known as the "EoR window"; this same terminology is used for observations in the post-EoR epoch.) remains mostly free of contamination and thus provides the best opportunity for measuring the cosmological 21 cm power spectrum during the Epoch of Reionization. The edge of the contaminated foreground wedge is given by

$$k_\parallel = \frac{H_0 D_c E(z) \theta_0}{c(1+z)} k_\perp, \tag{4.54}$$

where $E(z) = \sqrt{\Omega_m (1+z)^3 + \Omega_\Lambda}$, and D_c is the co-moving line-of-sight distance, θ_0 is the beam characteristic width. Equation from https://arxiv.org/pdf/1404.2596.pdf. We see that the contaminated region is proportional to the baseline distance (or $k_\perp$) and bounded by a line with a slope proportional to θ_0. In theory, the maximum is set by the horizon ($\theta_{0,max} = \pi/2$), but in practice this is determined by the primary beam of the array. In addition, real sources in the sky do contain some spectral structure, and together with the frequency-depedent beam shape of the instrument, blur out the edge of the foreground wedge. Therefore, merely looking in the EoR window does not ensure that we are probing a region free of foreground signals. Foreground cleaning, together with a better understanding of the instrument's beams and wedge mode mixing, are essential to getting accurate measurements of the underlying HI signal.

Chapter 5

Interferometry with the Tianlai Pathfinder Array

Chapter 5 will introduce intensity mapping with the Tianlai Pathfinder Array, which is an interferometric radio telescope designed to map the density of neutral hydrogen in the post-EoR epoch. Inteferometry allows the use of multiple small antennas instead of a large one. We first introduce the fundamentals of radio interferometry. We will then describe the instrumental response, stability, calibration, and the data pipeline for the Tianlai Dish Pathfinder Array. Finally, we discuss future plans and forecast the level of HI that can be obtained with the array. Some materials from this chapter can be found in the paper "The Tianlai Dish Pathfinder Array: design, operation and performance of a prototype transit radio interferometer" published in MNRAS in 2021 [135].

5.1 Fundamentals of Radio Interferometry

Radio telescopes can observe neutral hydrogen by detecting its 21 cm electromagnetic radiation. Traditional single-dish radio telescopes, such as the Green Bank, Parkes, Arecibo, or FAST telescopes, use parabolic metal mirrors that focus the incoming radiation into a radio receiver. The dish parabola coherently adds all electromagnetic radiation coming from a given direction. The bigger the diameter of the dish parabola, the higher is its angular resolution. Given the diameter of the dish is D, the resolution scales as λ/D, where $\lambda = \lambda_0(1+z)$ is the redshifted wavelength from the original λ_0. Since the radio wavelengths are long compared to optical signals, radio telescopes need large dish diameters to achieve a fine resolution.

Instead of a large single dish parabolic reflector that combines the signal optically, it is also possible to add the signals electronically. This class of telescopes is called interferometers. Each interferometer consists of multiple smaller antennas, and signals can be combined electronically to synthesize a dish whose aperture equals the largest separation between individual dishes. This concept is known as aperture synthesis. Since the signal from every pair of elements needs to be correlated, the total number of correlations (auto and cross correlations) grows as $N(N+1)/2 \sim N^2$.

The fundamental quantities measured by an interferometer are the visibilities. They are the cross-correlation of electric fields from all pairs of antennas. The visibility between antenna i and j, at frequency ν, is defined as

$$V_{ij}(\nu, \tau) \equiv \langle E_i(\nu, \tau) E_j^*(\nu, \tau) \rangle_t, \tag{5.1}$$

where the angle bracket denotes the time-average. The time dependence is due to the rotation of the Earth, thermal noise, gain fluctuations, cable length contraction or expansion, and so on. Most intensity mapping interferometers, like Tianlai, observe the sky by drift-scanning, so the time average is taken over a period of time that is small compared to the time over which the sky moves through the beam. Given the sky intensity $I(\boldsymbol{\theta}, \nu)$ that depends on angular position and frequency, the response of the interferometer can be expressed as

$$V_{ij}(\nu) = \int A_{ij}(\boldsymbol{\theta}, \nu) I(\boldsymbol{\theta}, \nu) e^{-2\pi i \nu (\mathbf{r}_i - \mathbf{r}_j)\cdot \boldsymbol{\theta}/c} \, \mathrm{d}^2\boldsymbol{\theta}, \tag{5.2}$$

where $A_{ij}(\boldsymbol{\theta}, \nu)$ is the primary beam response of antenna pair i and j, both as a function of angular position and frequency. $\mathbf{r}_i$ and $\mathbf{r}_j$ are position vectors of antennas i and j, respectively. In interferometry, we care about the baseline vector, defined as the difference in position between antenna pair $\mathbf{b}_{ij} \equiv \mathbf{r}_i - \mathbf{r}_j$, rather than the absolute positions of the antennas themselves. We note that the above equation is the Fourier transform of the product of the sky intensity and the primary beam response. This is the principle for radio interferometry. If the interferometer can sample a sufficiently large amount of data in the Fourier domain, then we can do an inverse Fourier transform to create images of the sky. This method provides an alternative to a large, single aperture telescope. Since we are dealing with the Fourier transform of the sky, a small baseline $\mathbf{b}_{ij}$ will capture the large-scale structure and a large baseline $\mathbf{b}_{ij}$ will capture the small-scale structure. In a practical interferometer, we have a mix of small and large baselines to sufficiently image the sky.

For simplicity, if we assume $A_{ij}(\boldsymbol{\theta}, \nu) = 1$ for all (i, j) pairs and $\boldsymbol{\theta}, \nu$, then Equation 5.2 becomes

$$V_{ij}(\nu) = \int I(\boldsymbol{\theta}, \nu) e^{-2\pi i \nu \mathbf{b}_{ij}\cdot \boldsymbol{\theta}/c} \, \mathrm{d}^2\boldsymbol{\theta}. \tag{5.3}$$

We define the Fourier dual to $\boldsymbol{\theta}$ as the baseline length in terms of the number of wavelengths: $\boldsymbol{u} = \boldsymbol{b}/\lambda$, where λ is the observed wavelength. $\boldsymbol{u}$ is now a frequency dependent quantity, and is usually denoted as a vector quantity $\boldsymbol{u} = (u, v, w)$. In the typical case that the array antennas lie on the same plane, $\boldsymbol{u} = (u, v)$. This is known as the uv-plane. The Fourier transform of the sky intensity is given by the coherence function

$$\widetilde{I}(\boldsymbol{u}, \nu) = \int I(\boldsymbol{\theta}, \nu) e^{-2\pi i \boldsymbol{u}\cdot\boldsymbol{\theta}} \, \mathrm{d}^2\boldsymbol{\theta}. \tag{5.4}$$

Reintroducing the primary beam, the visibility is a convolution of the pair-wise primary beam with the coherence function

$$V_{ij}(\nu) = \int \widetilde{A}_{ij} \left(\mathbf{u} - \mathbf{u}_{ij}, \nu\right) \widetilde{I}(\mathbf{u}, \nu) \, \mathrm{d}^2\mathbf{u}. \tag{5.5}$$

The pair of antennas integrates over the uv space, and thus the visibility takes samples in the Fourier transformed domain of the sky intensity (the uv-plane), modulated by the antenna response. The visibility is approximately the coherence function at the location of the baseline:

$$V_{ij}(\nu) \approx \widetilde{I}'(\boldsymbol{u}, \nu), \tag{5.6}$$

where the prime denotes the measured values. The interferometer is not sensitive to the sky signal directly but to the Fourier transform of the sky. In general, the visibility is a complex 2D function. At a time t and frequency ν, one baseline pointing in a specific direction will provide an unique (u, v) point in the Fourier space, which corresponds to one sample of the complex visibility funciton. For a driftscan array, due to the rotation of the Earth, the visibility samples elliptical tracks in the uv-plane.

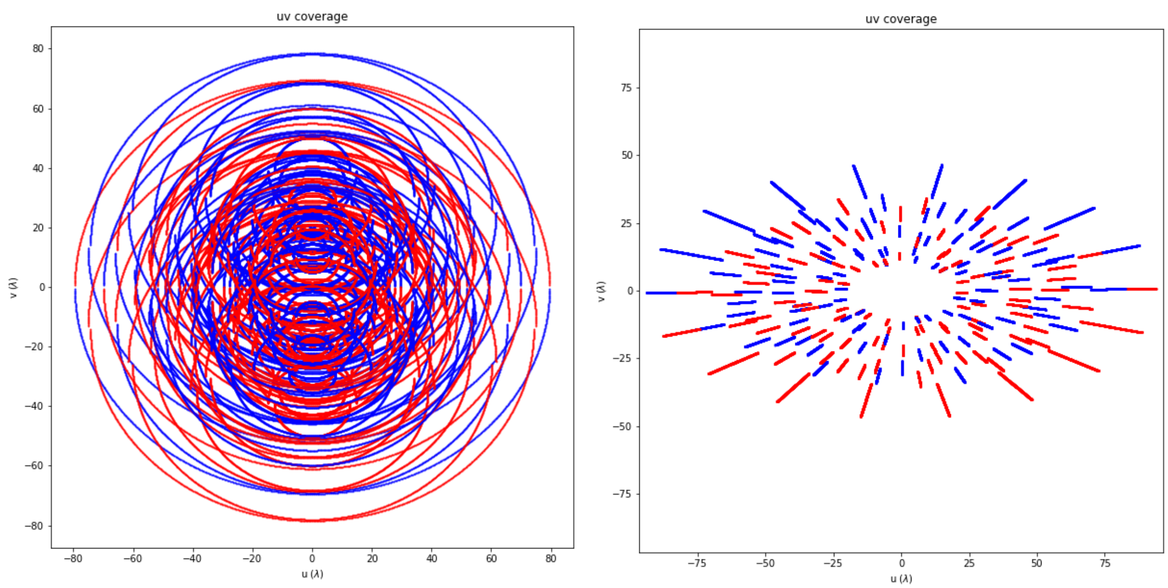

Figure 5.1: The uv-coverage map of the Tianlai array. Left: Samples collected by the array over a period of 12 hours at the lowest frequency bin 687 MHz. Red and blue tracks are complex conjugate of the same baseline (a baseline and its reversed baseline). Right: Samples collected by the array at one specific time over the whole frequency band 687 MHz to 812 MHz.The large tracks sample finer details in the sky, while the smaller tracks sample the larger scale structure.

With a set of measured visibilities in the uv-plane, we can do an inversion to the Fourier transform to approximate the intensity $\widetilde{I}(\boldsymbol{u}, \nu)$. We don't need full coverage of the uv-plane to do so, and this is also what makes interferometry a powerful tool to observe the sky:

$$I'(\boldsymbol{\theta}, \nu) \equiv \mathcal{FT}^{-1}\left[\widetilde{I}'\left(\mathbf{u}_{ij}, \nu\right)\right] \approx \sum_{ij} V_{ij}(\nu) e^{2\pi i \mathbf{u}_{ij} \cdot \boldsymbol{\theta}}. \tag{5.7}$$

A discrete inverse Fourier transform will give more weight to the points on the uv-plane that are more heavily sampled. The weight of each point can be normalized by dividing by the number of samples at such point. Since interferometers don't have a filled uv-plane, but only a sparse sampling of the plane, the measured intensity I' (also known as the dirty image) is a convolution of the sky intensity I with a point spread function (PSF):

$$I' = \mathrm{PSF} * I \tag{5.8}$$

where the PSF of the interferometric array is the Fourier transform of the visibility domain sampling function S:

$$\mathrm{PSF} \leftrightarrow S(i, v). \tag{5.9}$$

We want to measure the brightness temperature $I(\mathbf{r})$ (in mK) within a given region, where $\mathbf{r}$ is the normal Cartesian coordinates (x, y, z) and has units of Mpc. We choose the z direction to be parallel to the line-of-sight, so $z \equiv r_{\parallel}$. x and y are on the plane $\mathbf{r}_{\perp}$, perpendicular to the line-of-sight. The relationship between $\mathbf{r}$ and the angles on the sky $\boldsymbol{\theta}$ and frequency ν is as follows [96]:

$$
\begin{aligned}
x &= D_M(z)\theta_x \\
y &= D_M(z)\theta_y \\
\Delta z &= \frac{c(1+z)^2\Delta\nu}{H_0\nu_{21}\left[\Omega_m(1+z)^3+\Omega_k(1+z)^2+\Omega_\Lambda\right]^{1/2}},
\end{aligned} \tag{5.10}
$$

where $D_M(z)$ is the comoving distance to the observed redshift and ν_{21} is the redshifted frequency of the 21 cm emission. Having transformed Equation 5.7 into $\mathbf{r}$ coordinates, we can further transform the intensity in the wavenumber $\mathbf{k}$ space, where $\mathbf{k}$ is the Fourier dual to the real position $\mathbf{r}$ and has units of Mpc^{-1}:

$$
\widetilde{I}'(\mathbf{k}) \equiv \mathcal{FT}\left[I'(\mathbf{r})\right] = \int \mathbf{r} I'(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} \mathrm{d}^3. \tag{5.11}
$$

The transformed intensity has units of mK Mpc^3. Since the baseline coordinates $\mathbf{u}$ are the Fourier dual to $\boldsymbol{\theta} \propto \mathbf{r}_\perp$, which is itself a Fourier dual to $\mathbf{k}_\perp$, we have the proportion:

$$
\mathbf{k}_\perp = \frac{2\pi}{D_M(z)}\mathbf{u}. \tag{5.12}
$$

This means a large baseline will sample large $\mathbf{k}$ modes, and vice versa. The power spectrum can be calculated from $\widetilde{I}'(\mathbf{k})$ [96]:

$$
\left\langle \left|\widetilde{I}'(\mathbf{k})\right|^2 \right\rangle = \frac{1}{(2\pi)^3}\int \mathrm{d}^3\mathbf{k}' P\left(\mathbf{k}'\right)\left|\widetilde{A}\left(\mathbf{k}-\mathbf{k}'\right)\right|^2. \tag{5.13}
$$

If the primary beam is sharply peaked in $\mathbf{k}$-space, the power spectrum is approximately constant in the integral, so

$$
P'(\mathbf{k}) = \frac{\left\langle \left|\widetilde{I}'(\mathbf{k})\right|^2 \right\rangle}{\frac{1}{(2\pi)^3}\int \mathrm{d}^3\mathbf{k}'\left|\widetilde{A}\left(\mathbf{k}-\mathbf{k}'\right)\right|^2}, \tag{5.14}
$$

in which the units of the power spectrum are mK^2 Mpc^3. Using Parseval's theorem, we can approximate the denominator as

$$\frac{1}{(2\pi)^3}\int \mathrm{d}^3\mathbf{k}' \left|\widetilde{A}\left(\mathbf{k}-\mathbf{k}'\right)\right|^2 = \int \mathrm{d}^3\mathbf{r}' \left|A\left(\mathbf{r}-\mathbf{r}'\right)\right|^2 \approx D_M(z)^2\Omega\Delta D, \tag{5.15}$$

where Ω and ΔD are the solid angle of the primary beam and the extent of the observation in the line-of-sight direction, respectively.

5.2 The Tianlai Pathfinder Array

The Tianlai Project is a 21 cm intensity mapping survey of the northern sky aimed at characterizing the equation of state of the dark energy by measuring the baryon acoustic oscillation (BAO) features in the large scale structure power spectrum as a function of redshift. There are currently six 21 cm intensity mapping experiments with dedicated instrumentation that are either under construction or observing: CHIME and CHORD in Canada, HIRAX in South Africa, Tianlai in China, OWFA in India, and BINGO from UK Brazil. At the current time, the Tianlai Project is in the Pathfinder stage, which is the technology verification stage. It has two arrays: one array consists of 16 circular, on-axis dish antennas, with dual-polarization feed antennas in each dish. The other array consists of two large cylinder reflector antennas, each with 96 dual-polarization feeds. These complementary array designs were chosen to test different approaches to 21 cm intensity mapping. Both arrays are located in a radio quiet site ($44°9'$N, $91°48'$E) in Honglixia, Balikun County, Xinjiang Autonomous Region in northwest China. They perform drift scans of the sky at constant declination. To minimize radio interference (RFI), the station housing the digital electronics, including the digital correlator, is located 5.8 km (11.8 km by road) away from the two arrays.

The array and the digital electronics station are connected by a 8 km power line and optical fiber cable.

The driftscan strategy allows for large sky coverage with a simple and inexpensive instrument design. However, unlike tracking instruments, which can calibrate continuously on bright sources in or near the field they are mapping, driftscan instruments such as Tianlai must wait for bright sources to pass through the field of view to calibrate or resort to calibrating on dimmer sources.

For both the dish and cylinder arrays, the reflectors, the feed antennas, and the amplifiers are designed to operate from 400 MHz to 1430 MHz, corresponding to redshift $2.55 \geq z \geq -0.01$. The instrument can operate in any RF bandwidth of ~100 MHz by tuning the local oscillator frequency in the receivers and replacing the bandpass filters. Currently, the Pathfinder arrays operate in the frequency range 700-800 MHz, corresponding to redshift $1.03 \geq z \geq 0.78$ in 512 frequency channels ($\delta\nu = 244\,\mathrm{kHz}$, $\delta z = 0.0002$). This chapter only discusses the dish array.

The Tianlai Dish Array

The Tianlai Dish Array consists of 16 on-axis dishes, each with a diameter of 6 m. Each dish is mounted on an Alt-Azimuth mount and contains a dual linear polarization feed antenna. One polarization axis is parallel to the altitude axis (horizontal, H, parallel to the ground)), and the other is orthogonal to that axis (vertical, V). The motors which control the dishes can steer the dishes to any direction, but once pointed, the dishes operate in drift scan mode. The design parameters of the Tianlai dish antenna are summarized in Table 5.1.

The dishes are tightly packed in a circle, as shown in Figure 5.2. The dishes are arranged in two concentric circles of radius 8.8 m and 17.6 m around a central dish. The

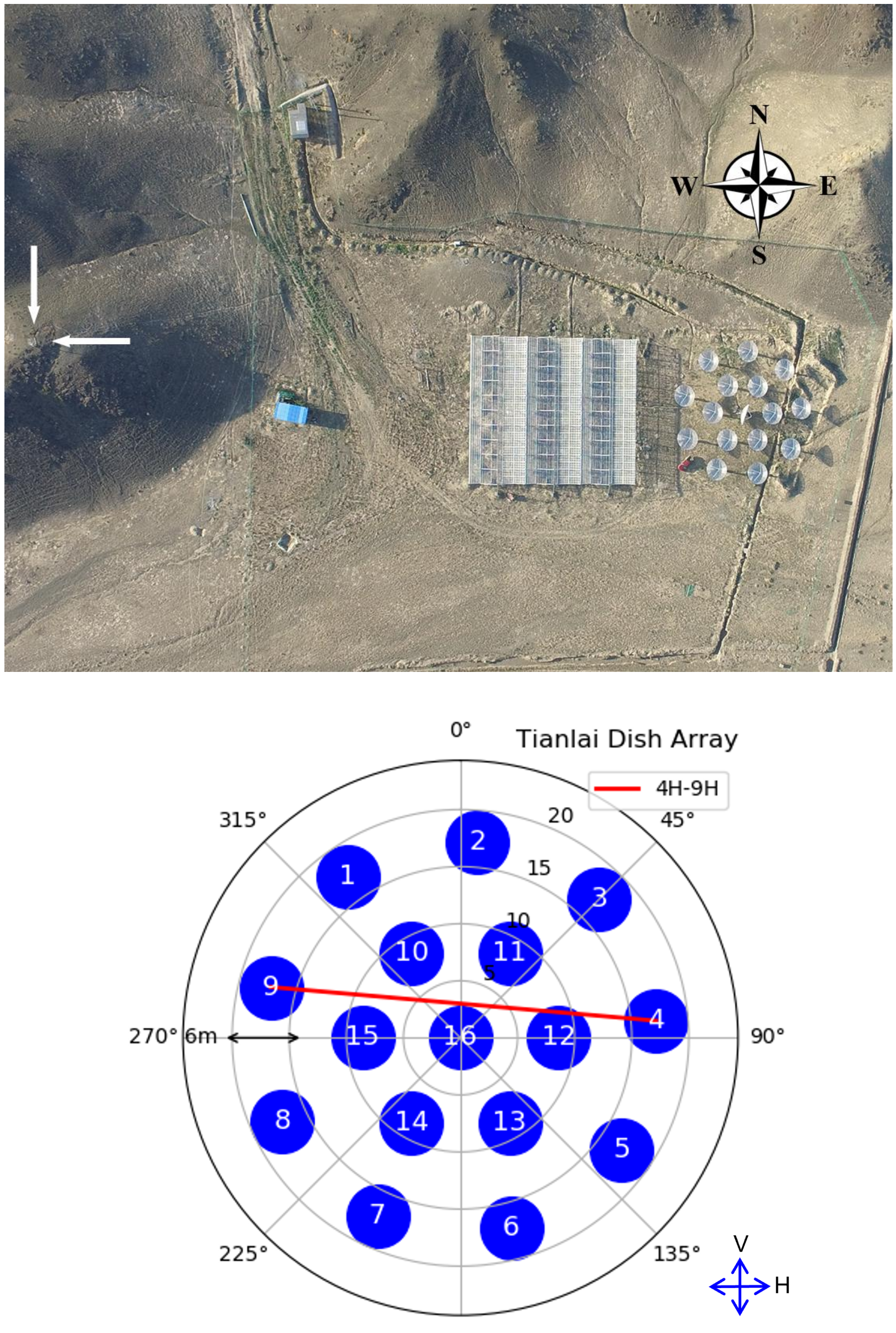

Figure 5.2: Top: A top view photograph of the Tianlai arrays, which consist of the Dish Array and the Cylinder Array. The photo was taken with a drone at a height of 280 m above the ground. The arrays saw first light in 2016. The position of the calibration noise source (CNS) is indicated by the white arrows on the left. Bottom: A schematic diagram of the dish array. Shown in red is a baseline in which the H polarization of dish 4 is correlated with the H polarization of dish 9.

arrangement of baselines allows for a moderate coverage of the uv-plane (see Figure 5.1).

The low noise amplifiers (LNA) are mounted to the back of the receivers and are designed to have low noise temperature (about 47 K at room temperature). The RF signals are transmitted through 15 m long coaxial cables to optical transmitters located underneath the antennas, where they are converted to optical signals via amplitude modulation and transmitted via 8 km optical fiber to the station house. The optical signals are then converted back to RF signals there. The RF signals are shifted to an intermediate frequency (IF) in the range 135-235 MHz before they are sent to the digital system. A summary of this process in shown in Figure 5.3.

The IF outputs from the RF analog system are then fed into an FPGA-based correlator, which consists of three FPGA boards: one for control and two processing boards for signal sampling and processing. The Analog to Digital Converters (ADC) in the processing boards sample the RF signals at a rate of 250 MSPS and a sampling length of 14 bits. The FPGAs in the two processing boards compute the cross-correlations in the time series data. The output visibility data contains 528 visibilities (32 auto-correlations and 496 cross-correlations) for 512 frequency channels. The visibilities are averaged over an integration period of 1 second. The full data rate from the correlator is approximate 175 GB/day. Finally, the visibility data are sent to a storage server and saved in `HDF5` format.

To compensate for the phase variation along the 8 km long optical cables, we can use absolute calibration using bright astronomical sources. However, there are not enough bright sources on the sky for small aperture arrays like the Tianlai Dish Array to have at least one in the primary beam at all times. To overcome this challenge, we have a dedicated calibration noise source (CNS) located on top of a hill nearby (see

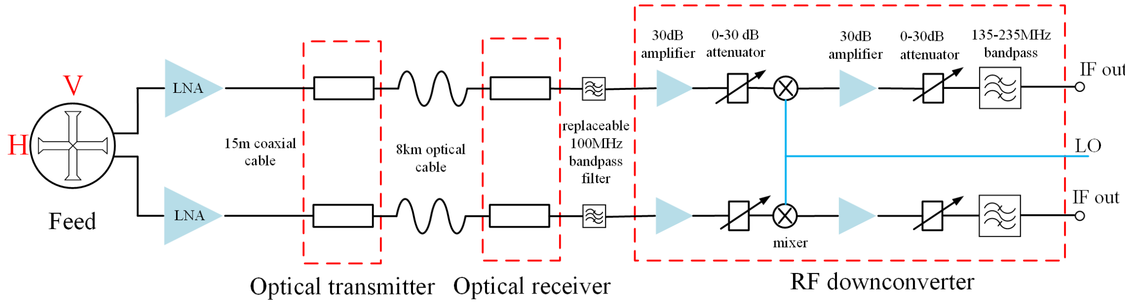

Figure 5.3: Schematic of the RF analog system.

Reflector diameter	6 m
Antenna mount	Alt-Az pedestal
f/D	0.37
Feed illumination angle	68°
Surface roughness (design)	$\lambda/50$ at 21 cm
Altitude angle	8° to 88.5°
Azimuth angle	±360°
Rotation speed of Az axis	$0.002 \sim 1^\circ/\mathrm{s}$
Rotation speed of Alt axis	$0.002 \sim 0.5^\circ/\mathrm{s}$
Acceleration	$1^\circ/\mathrm{s}^2$
Gain(design)	29.4+20log(f/700 MHz) dBi
Total mass	800 kg

Table 5.1: Design parameters of a Tianlai dish antenna.

Figure 5.2) to provide relative calibration. A broadband RF noise generator is placed in a thermally controlled environment and is supplied with a regulated DC power to ensure the stability of the RF amplitude. The on-off timing of the CNS is controlled by a clock signal carried by optical fiber located in the station house 8 km away. More details about calibration will be discussed further.

Observations

The Tianlai Dish Array has collected about 6200 hours of observational data, of which 5700 hours are drift scans directly targeted at the North Celestial Pole (NCP).

Data Set	Date	Calibration Sources	Targets	Length (hours)
Data 201605-06	May 2016	None	Cygnus A	72
CygnusANP 20170812	Aug 2017	Cygnus A	North Pole	67
CasAs 20171017	Oct 2017	None	North Pole	147
CasAs 20171026	Oct 2017	None	Cassiopeia A	290
3srcNP 20180101	Jan 2018	3C48, Cassiopeia A, M1	North Pole	241
2srcNP 20180112	Jan 2018	3C48, M1	North Pole	97
IC443NP 20180323	Mar 2018	IC443	North Pole	181
M87NP 20180407	Apr 2018	M87	North Pole	90
2srcNP 20180416	Apr 2018	IC443, M87	North Pole	142
3srcNP 20181212	Dec 2018	Cassiopeia A, 3C48, M1	North Pole	757
1DaySun 20190113	Jan 2019	None	Sun	48
3srcNP 20190128	Jan 2019	Cassiopeia A, 3C48, M1	North Pole	741
3srcNP 20190228	Feb 2019	3C123, M1, IC443	North Pole	764
3srcNP 20190402	Apr 2019	M1, IC443, 3C273	North Pole	522
3srcNP 20190611	Jun 2019	M87, Hercules A, Cygnus A	North Pole	737
3srcNP 20190830	Aug 2019	3C400, Cygnus A, Cassiopeia A	North Pole	924
3srcNP 20191022	Oct 2019	3C400, Cygnus A, Cassiopeia A	North Pole	302

Table 5.2: Observation log for the Tianlai Dish Array from 2016 to late 2019.

The details of each observation are shown in Table 5.2. Before each NCP observation, the antennas are pointed at one or more bright astronomical source(s) for calibration.

During each observation, the CNS is turned on and off periodically. In 2017, the CNS was turned on for 20 seconds, followed by 220 seconds off, repeatedly. The fraction of noise-on time is $\sim 8.33\%$. In 2018, the noise-on fraction was reduced to $\sim 1.67\%$: the CNS was turned on for 4 seconds, followed by 236 seconds off, repeatedly.

5.2.1 Instrument and Beam Pattern

Since the astrophysical foregrounds are several orders of magnitude larger than the HI signal, and instrumental effects can introduce erroneous structures into the HI signal, it is essential to have a very good knowledge of the frequency-dependent beam pattern of the antennas.

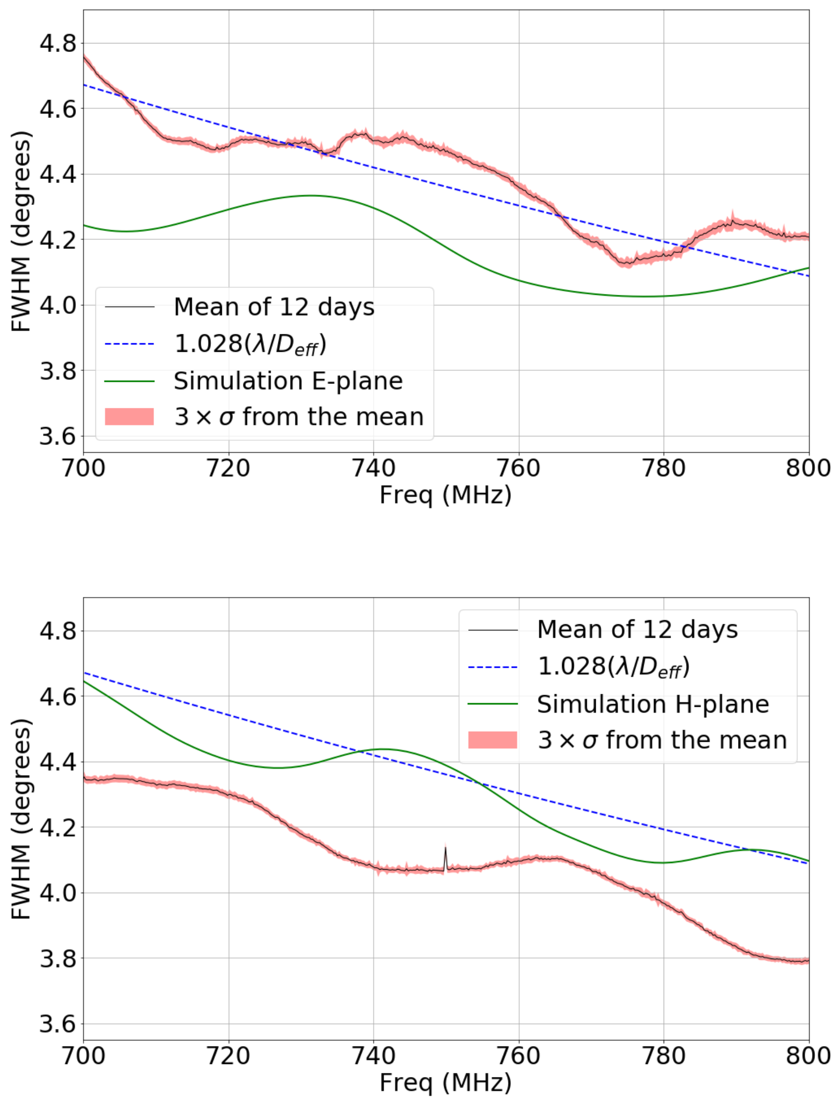

Figure 5.4: Plot of the mean FWHM of the main beam vs. frequency using daily transits of Cas A. The top figure shows a typical horizontal baseline (4H-9H), which primarily measures the E-plane of the antenna, while the lower figure shows a vertical baseline (4V-9V), primarily measuring the H-plane. The black line shows the mean in each 244-kHz wide frequency bin over the period of 12 days and the red band shows 3 times the standard deviation in FWHM for each day. The solid green line shows the expected FWHM from electromagnetic simulation. The dotted blue line shows the FWHM of a uniformly-illuminated Airy disk.

We have performed extensive electromagnetic simulations of the feed antenna and the dish reflector using `CST Studio Suite`. The beam pattern of the array has also been scanned with an unmanned aerial vehicle (UAV). The UAV, outfitted with a broadband noise source and flown in the farfield of the dish array, measured the profile of the mainbeam and the sidelobes at different RF power levels and frequencies. Details of the electromagnetic simulations and the UAV measurements can be found in [141].

We also used the transit of Cassiopeia A (Cas A) to characterize the main beam as a function of frequency. The Cas A signal is not bright enough to assess the beam's sidelobes. Figure 5.4 shows the full-width half-maximum (FWHM) of one cut through the beam pattern for a typical baseline as a function of frequency. Since a baseline consists of two dishes, the measured beam pattern is the geometric mean of the patterns of those two dishes. The beam pattern is measured in the E-W direction repeatedly over a period of 12 days (starting on 2017/10/26) by observing the transit of Cas A. The absolute magnitude of the visibility versus time is fitted to a Gaussian shape for each transit. The mean and standard deviation of the FWHM fit is shown in the figure (black line and red band, respectively). We can see that daily variation of the FWHM is less than 1%. The electromagnetic simulation of the FWHM is also shown in the plot (solid green line). The dotted blue line shows diffraction limited circular aperture, which equals $1.028\lambda/D_{\text{eff}}$ with $D_{\text{eff}} = 0.9D$ of the actual $D = 6\,\text{m}$ and the 1.028 prefactor comes from the FWHM of an Airy pattern from a uniformly illuminated disk. We also see a standing wave pattern in the plot, which appears in both simulated and measured data.

The uniformity of the beam widths for the dish arrays also needs to be characterized, since we want the beam width to remain consistent across all antennas. We use the transit of M1 to measure the FWHM of 118 H-H baselines (excluding auto-correlation

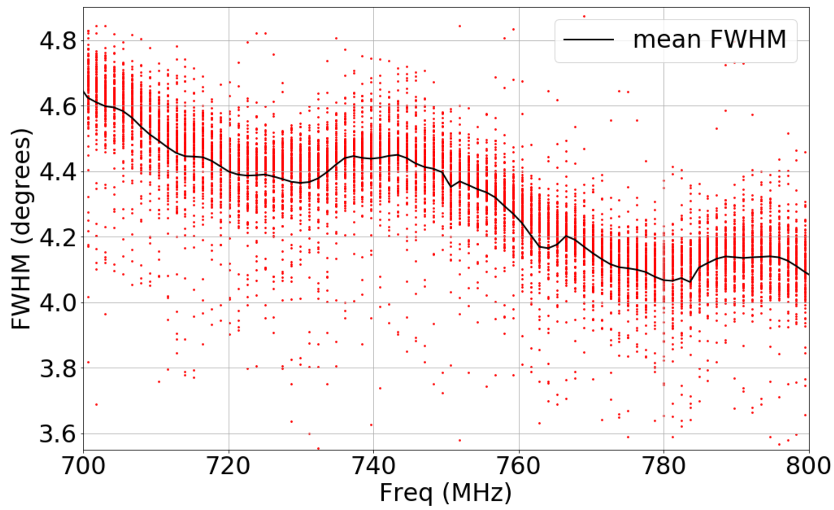

Figure 5.5: The black line shows the mean FWHM vs. frequency for 118 H-H baselines during a transit of M1 on 2018/01/02

and faulty baselines) as a function of frequency. The result is shown in Figure 5.5. We see that the beam widths are fairly consistent with about a few percent variation.

We also use the transit data for Cas A on 2016/10/30 to quantify the pointing accuracy of the dishes in the E-W direction. Figure 5.6 shows the variation in the absolute amplitude of the visibilities during the same observation interval. 120 cross-correlation visibilities are shown in different colors. Each curve is the binned average of 40 middle frequency channels of width 244 kHz each, for a total frequency range from 742.6172 MHz to 752.1387 MHz. The black curve shows the predicted response from a 5.4m diameter dish with simple Gaussian beam with no sidelobe.

5.2.2 Gain stability

As described in the previous section, we can use the CNS to calibrate the electronic phase drift and the transits of point sources to get absolute gains and calibrations.

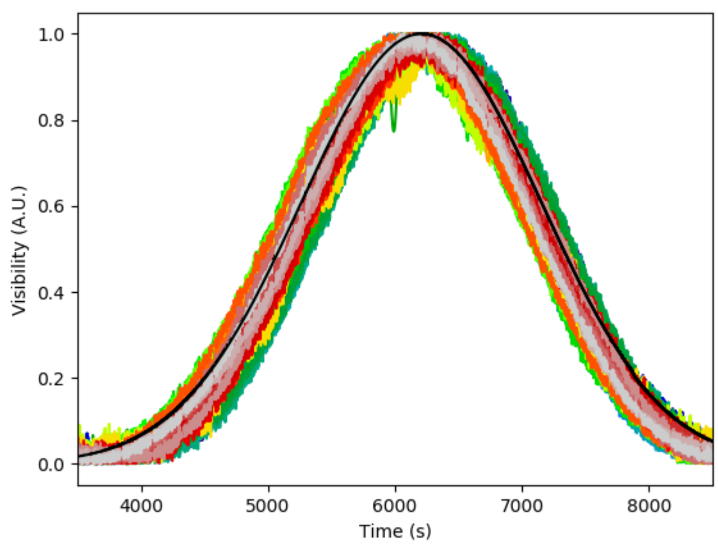

Figure 5.6: Variations of the absolute magnitude of the cross-correlation visibilities vs. time (in seconds) during a transit of Cas A. All the un-calibrated cross-correlations have been renormalized to unity at their respective maxima. The green dip at around 6000 s is due to an artifact from the CNS interpolation.

We are also interested in the gain stability, i.e. how the gain varies with time. In this section, we will characterize the gain stability of the instrument by studying its response to a strong point source (Cas A). In this observation, the whole array was pointed at a fixed declination of 58.8 degrees over a period of 12 days, and we measured the variations in the absolute amplitude and phase of Cas A transits through the meridian.

We look at the peak response of the array during Cas A transits and plot the uncalibrated amplitude and phase for all frequency channels in Figure 5.11, where different colors represent different days. Since the array is pointed directly at the declination of Cas A, this is a measurement through the main beam and thus there is minimal side lobe structure. We see that both the amplitude and the phase of the gain are quite stable over time.

In Figure 5.7, we show the phase calibration with the CNS before and after using `nscal` over a period of 11 days. The phase calibration is performed at the peak response during Cas A transits, and the calibration reduces the phase variation significantly over most baselines. Work is still in progress to improve the efficacy of the algorithm.

5.2.3 Sensitivity versus integration time

Figure 5.8 shows the level of visibility fluctuations due to receiver noise, sky signal, etc. The plot shows the overlapping Allan variance versus integration time. Allan variance is used to estimate the stability due to noise and other systematic errors. As expected for white noise, at low integration time τ, the variance is high. At higher τ, it decreases since the noise averages out, as the noise should integrate down as $1/\sqrt{\tau}$. However, at τ beyond 300 seconds, the variance starts to increase again due to the rotation of the sky. This is important in choosing the ideal length of time over which we can bin the visibility in the time direction. To get to the level of the HI signal

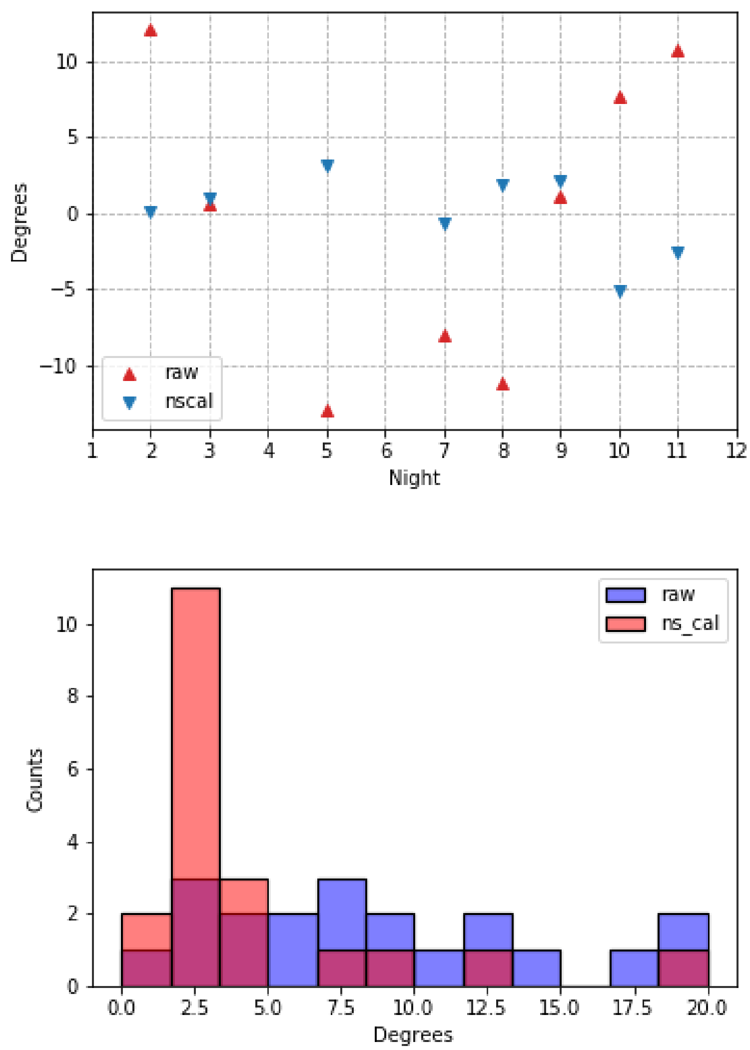

Figure 5.7: Top: Phase variation measured before and after phase calibration with the CNS using `nscal`. The phase variation is calculated as the deviation from the mean of 11 days at the peak of the Cas A transit. Bottom: Histogram of the phase variation for 20 typical baselines, before and after running `nscal`. Application of `nscal` shifted the phase deviations toward 0 and improved the phase deviation significantly.

(about 1 mK), we need to beat the noise which is given by the radiometer equation $T_{\rm rms} = T_{\rm sys}/\sqrt{\tau\Delta\nu}$. For the Tianlai array with $T_{\rm sys} \sim 80$K and $\Delta\nu = 0.25$ MHz, we would need $\tau = 2.56 \times 10^4$ seconds to get down to the HI signal. Since this plot shows we can only integrate for ~ 300 seconds per day, it would take about 85 days of observation to achieve this target.

5.2.4 Tianlai data analysis pipeline

The digital outputs from the cross-correlator are stored on hard drive arrays in HDF5 format for offline processing. The offline data is then passed to a custom data processing pipeline, named `tlpipe`, for additional scientific analysis and map making. `tlpipe` is written in Python and designed to be modular, with each module serving a different function. A schematic of `tlpipe` is shown in Figure 5.9, with independent processing tasks in purple rectangular boxes.

Users of `tlpipe` can write a customized task in the pipeline in a `*.pipe` file. The general procedures consists of the following tasks:

Input Data: The digital output from the correlator. For the dish array, the integration time is 1 second, and there are 32(32+1)/2 visibilities from 16 dual-polarization feeds (including auto-polarization). The data is saved at 60 minute intervals into a new HDF5 file. `tlpipe` can read one or multiples of those files.

Radio Frequency Interference (RFI) Flagging: RFI signals much bigger than typical white noise or astronomical noise are flagged as anomalous and masked from the visibility. The Boolean mask is stored separately from the visibility. `tlpipe` has multiple flagging algorithms, but mainly uses two: the sum threshold method [102] and the scale-invariant rank (SIR) operator method [103]. The RFI masked from 1 hour of nighttime data using the sum threshold and SIR operator method is shown in

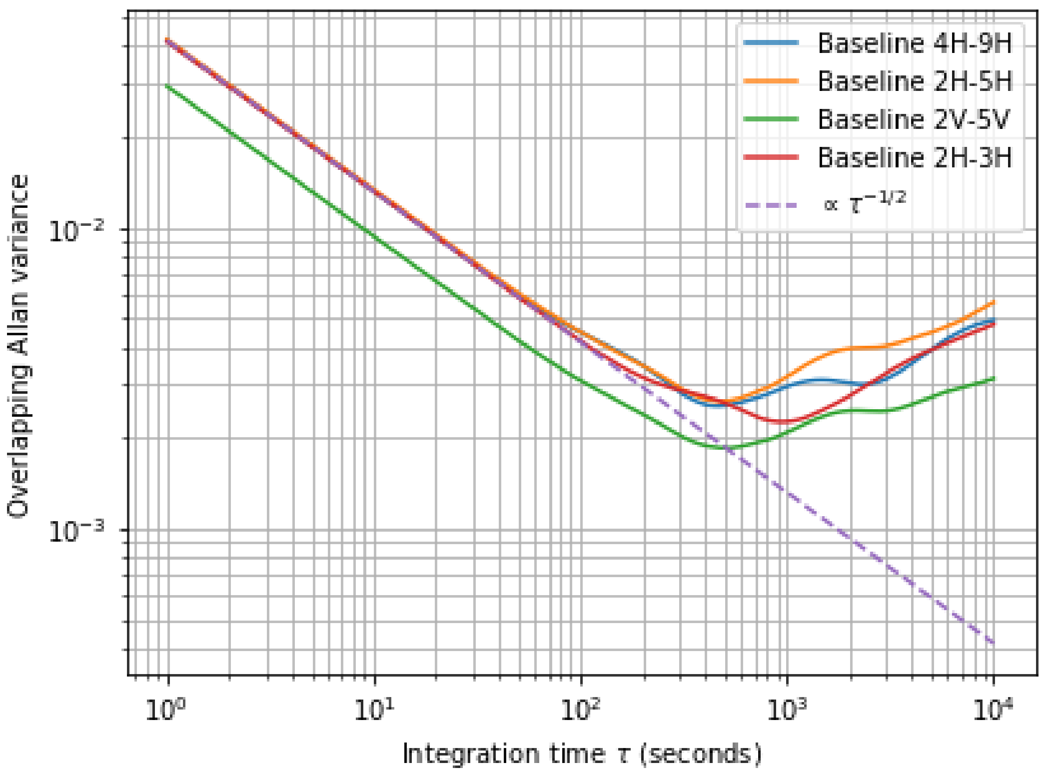

Figure 5.8: Overlapping Allan variance [114] versus integration time, τ, for four typical baselines centered at 747.5 MHz with bandwidth 0.244 MHz during the nighttime only, for the real part of the visibility. The visibility is uncalibrated and the vertical axis is in arbitrary units. The intercept of the dashed purple line $\propto \tau^{-1/2}$ is adjusted so that it matches the variance trend. The plot shows that the noise integrates down as $1/\sqrt{\tau}$, as expected, for about 300 seconds. The imaginary part of the visibility shows similar behavior.

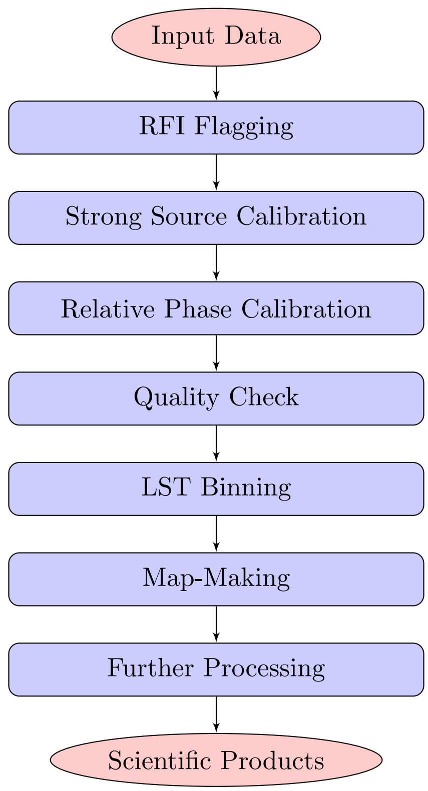

Figure 5.9: Flowchart for the data processing pipeline. Figure from [142].

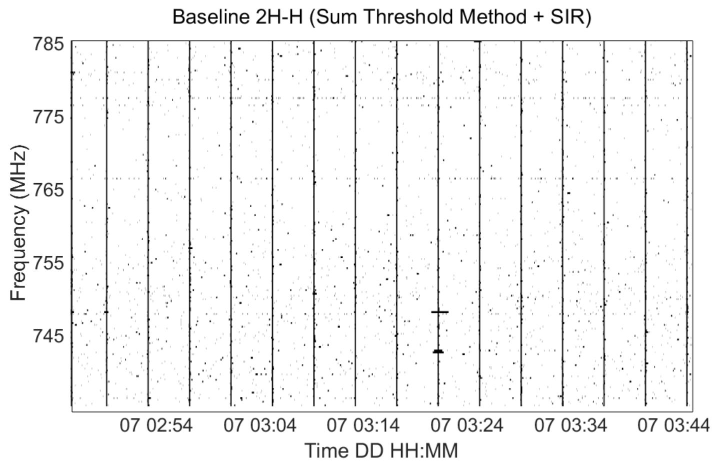

Figure 5.10: Masking of the CNS and RFI after applying both the sum threshold method and the SIR operator method. The vertical lines show times and frequencies masked when the signals from the CNS are used for calibration. Besides the CNS, there is a small amount of data at discrete frequencies and times masked as RFI. The masked RFI (dots and lines) are magnified for better readability. Some faint horizontal lines (at about 777 MHz and 767 MHz) are from intermittent RFI.

Figure 5.10. The periodic vertical stripes show the mask when the CNS is turned on, and the dots and small horizontal lines are the RFI. Since the telescope is located in a radio-quiet site, only about 0.6% of the data is lost due to RFI.

All the metadata, including calibration sources, CNS on time, temperature of the analog electronics room, site temperature, dew point, humidity, precipitation level, wind direction, wind speed, barometric pressure, etc. are stored along with the visibilities in the `HDF5` files. These metadata can be later use to check the electronic gain variation of the system versus different weather variables.

Relative phase calibration: A strong and regularly-broadcast calibration noise source (CNS) is used for phase calibration. It is currently used to remove the phase variations over time but may be used for amplitude calibration in the future. The noise source can be viewed as a near-field source with visibility

$$V_{ij}^{\mathrm{ns}} = C \cdot e^{ik(r_i - r_j)}, \tag{5.16}$$

where C is a real constant, and the subscript ij corresponds to baseline i, j. For each baseline, the task defines the visibility during the on and off cycles of the CNS to be

$$\begin{aligned} V_{ij}^{\mathrm{on}} &= G_{ij} \left(V_{ij}^{\mathrm{sky}} + V_{ij}^{\mathrm{ns}} + n_{ij} \right) \\ V_{ij}^{\mathrm{off}} &= G_{ij} \left(V_{ij}^{\mathrm{sky}} + n_{ij} \right), \end{aligned} \tag{5.17}$$

where G_{ij} is the complex gain of baseline i, j and n is the noise, and V_{ij}^{sky} is the observed visibility from the sky.

The difference between V_{ij}^{on} and V_{ij}^{off} is then given by

$$\begin{aligned} V_{ij}^{\mathrm{on}} - V_{ij}^{\mathrm{off}} &= G_{ij} V_{ij}^{\mathrm{ns}} \\ &= \left| G_{ij} \right| e^{ik\Delta L} C \cdot e^{ik(r_i - r_j)} \\ &= C \left| G_{ij} \right| e^{ik(\Delta L + (r_i - r_j))}, \end{aligned} \tag{5.18}$$

where ΔL is the difference in cable length. The phase introduced by the CNS is then

$$\phi_{ij} = \mathrm{Arg} \left(V_{ij}^{\mathrm{on}} - V_{ij}^{\mathrm{off}} \right) = k \left(\Delta L + (r_i - r_j) \right) = k \Delta L + \text{ const} \tag{5.19}$$

The corrected sky visibility, after CNS calibration, is

$$V_{ij}^{\text{rel-cal}} = e^{-i\phi_{ij}} V_{ij} = e^{-i \, \mathrm{Arg}\left(V_{ij}^{\mathrm{ea}} - V_{ij}^{\mathrm{off}} \right)} V_{ij}. \tag{5.20}$$

Point source calibration: The next step after relative phase calibration with the CNS is absolute gain calibration. The telescope uses transits of strong astronomical radio sources to calibrate the amplitudes and phases of the complex gains for each feed. The pipeline fits the transit signal for each baseline and frequency independently. The complex gains for each feed (both horizontal and vertical polarizations) are then obtained and saved. The computed complex gains are applied to the entire dataset until the next radio sources transits.

Map-making: The map making code uses m-mode analysis [122, 124]. The results are 3-dimensional maps, with 2 angular dimensions and 1 redshift dimension.

Other tasks: There are other tasks in the pipeline for plotting, time or frequency averaging, bad channel identifications, or calculating the 21 cm power spectrum from maps.

5.3 Goals and Future Plans

Figure 5.12 shows a simulation of the brightness temperature of the HI signal and foreground contamination as a function of observed frequency and source redshift. The blue line demonstrates the example HI signal along a typical line-of-sight that traces the over-density by small objects at low redshift. The red line shows what we expect to see from the foregrounds: several orders of magnitude brighter but spectrally smooth. The most challenging requirements for future telescopes is to tackle this bright astrophysical foreground.

Foreground contamination sets the design requirements for all stages of the instrument design, simulation, and data analysis. To remove the bright foregrounds, more accurate calibration is needed: the beam, or the instrument response must be understood to 0.1%, and the gain, or the time-dependent response of the instrument, must be calibrated to 1% [124]. Knowledge of the beam needs higher precision to reduce mode mixing from the foreground into the HI signal. Currently, we rely on sky signals for

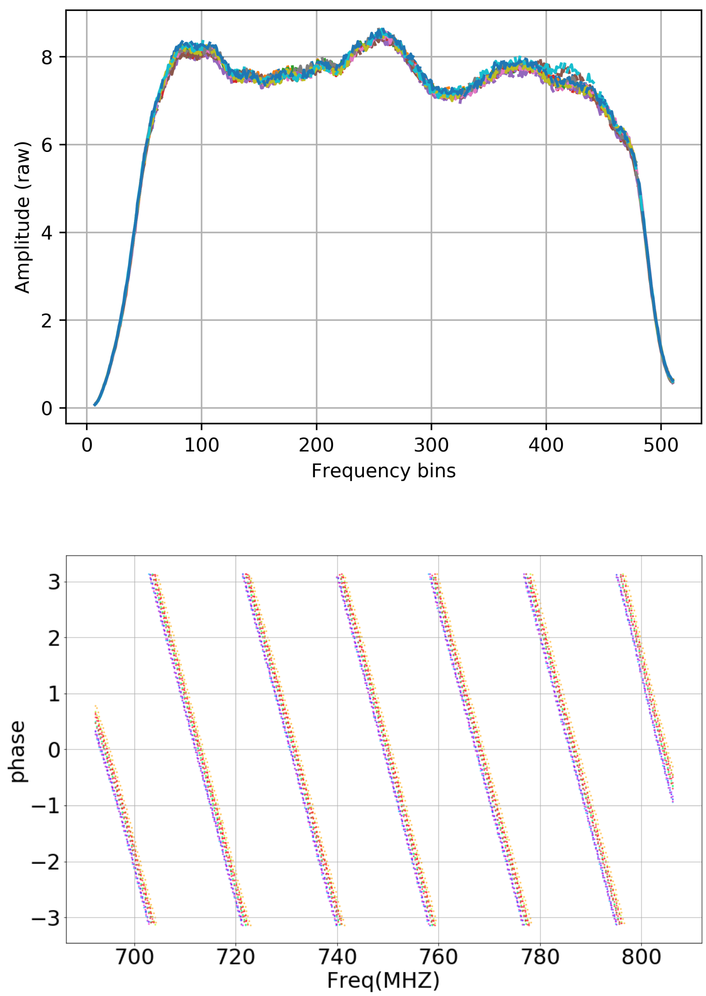

Figure 5.11: Uncalibrated gain amplitude (top) and phase (bottom) versus frequency during transits of Cas A over 11 nights for baseline 4H-9H. Each colored curve represents the peak response during the transit for each night.

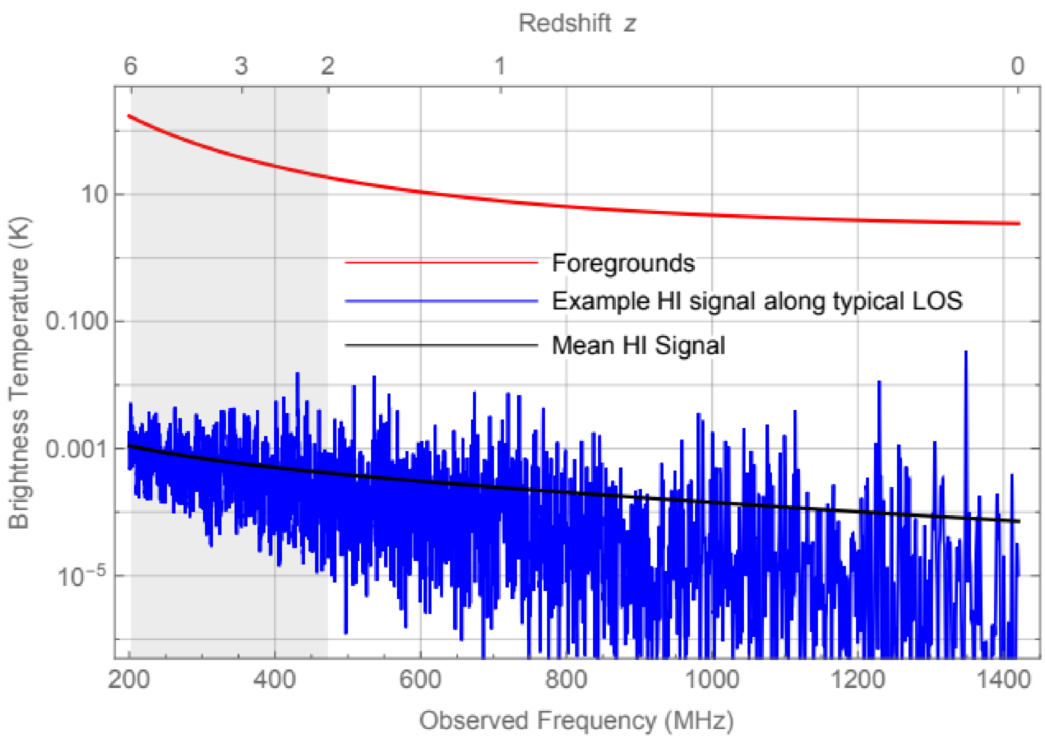

Figure 5.12: Simulation of the brightness temperature of the HI signal and foreground contamination as a function of observed frequency and source redshift. The blue line demonstrates the example HI signal along a typical line-of-sight, with the black line showing the mean HI signal. The red line shows shows the brightness temperature of the foreground. Figure from [14].

both beam and gain calibration, and this has not achieved the required accuracy for adequately removing the foregrounds. Better drone measurements and electromagnetic simulation, including environmental factors such as scattering from the Earth also needs to be considered. The future low-z Tianlai bands overlap with most geo-positioning satellites, so RFI removal will need to be improved. In addition, the ionosphere plasma couples with the Earth's magnetic field to rotate the polarization of the incoming light. The rotation is proportional to λ^2 (longer wavelengths are more affected) and time-dependent with the number of free electrons present in the ionosphere. This affects EoR experiments but is a negligible effect at Tianlai observing frequencies.

One of the largest sources of noise comes from the system noise temperature, which is dominated by the low noise amplifier. The system noise temperatures for both the Tianlai dish and cylinder arrays are $80-85$K, depending on frequency. Deploying more antennas can achieve a better signal-to-noise ratio. Synchrotron is still subdominant to the receiver noise for the current Tianlai bands. Data processing also needs to be optimized for the huge amount of data coming in. Both the dish and cylinder arrays generate about 20GB of data each day without compression. This poses a challenge in data transfer, storage, distribution, and analysis.

5.3.1 Cross-correlation with galaxy surveys

Intensity mapping is a relatively recent field, enabled by the improved sensitivity in the observing instruments and data processing. Optical galaxy surveys, on the other hand, are a more well-developed observational tool. Galaxy catalogs have grown tremendously to millions of galaxies, both at low redshift resolution (photometric) and high resolution (spectroscopic). Galaxy surveys therefore can augment the data obtained with 21 cm intensity mapping, which aggregates the emission of galaxy clusters and provide a means for testing the intensity mapping technique. By cross-correlating with galaxy surveys, this allows us to accurately map the full 3D structure in the Universe. In the near future, the Tianlai Pathfinder arrays will be tuned to operate in the frequency range 1330-1430 MHz, corresponding to redshift $0.07 \geq z \geq -0.01$, to facilitate cross-correlation with low redshift HI and galaxy surveys.

In 21 cm intensity mapping, the dominant sources of noise come from sky noise and thermal noise. We measure the aggregate intensity from all objects, including small and faint signals, so the Poisson noise (or shot noise) is currently not a concern. The shot noise comes from the fact the we are sampling a continuous field using a finite number of objects. To mitigate shot noise, we need to take spectra of more objects up to a given flux. In the future, intensity mapping experiments will start to be sensitive to shot noise. Galaxy surveys allow us to probe individual sub-samples with certain properties to reduce the shot noise in future intensity mapping experiments.

In addition, even though galaxy surveys excel at lower redshifts, the become more challenging at higher redshifts. Since we are looking at the younger Universe, the observations must be done in the infrared, where detectors are less efficient and more expensive than traditional optical sensors. The infrared range also contains more sky contamination and more variables. Intensity mapping may one day complement the weakness of galaxy surveys at higher redshifts.

Chapter 6

Sun Removal with AlgoSCR

Several ongoing and upcoming intensity mapping radio interferometers, such as Tianlai, CHIME, HERA, HIRAX, etc., are designed to map large swaths of the sky by drift scanning over periods of many months. One of the challenges of the observations is that the daytime data are contaminated by strong radio signals from the Sun. In the case of Tianlai, this results in almost half of the measured data being unusable. In this chapter, we try to address this issue by developing an algorithm for solar contamination removal (AlgoSCR) from the radio data. The algorithm is based on an eigenvalue analysis of the visibility matrix, and hence is applicable only to interferometers. We apply AlgoSCR to simulated visibilities, as well as real daytime data from the Tianlai dish array. The algorithm can reduce strong solar contamination by about 95% without seriously affecting other weaker sky signals and thus makes the data usable for certain applications. The content of this chapter is also given in the paper "AlgoSCR: an algorithm for solar contamination removal from radio interferometric data" published on MNRAS in 2022. The full version of the paper is available here https://academic.oup.com/mnras/article-abstract/512/3/3520/6544640.

6.1 Introduction

Cosmologists study the Universe on the largest observable distance scales in order to understand its origin and evolution. In the past few decades, cosmic microwave background (CMB) instruments have mapped almost the entire sky with high sensitivity and fine angular resolution. These maps measure the intensity and polarization fluctuations at the last scattering surface and remain a primary tool for studying the Universe. However, for understanding the nature of dark matter and dark energy, it is essential to study the evolution of structure as a function of time. Galaxy redshift surveys have been extremely successful in mapping the large scale structure of the Universe by cataloging the distribution of luminous galaxies in redshift space. These maps can be used, for example, to observe the characteristic baryon-acoustic oscillation (BAO) signal, which can be used as a standard ruler to extract cosmological parameters. However, as we map larger and more distant volumes of the Universe, the method faces multiple challenges. For example, the galaxies become fainter and spectral lines are redshifted to wavelengths that are difficult to detect from the ground.

Hydrogen intensity mapping, a radically different technique, creates 3D maps using the 21 cm emission of neutral hydrogen (HI) without resolving individual galaxies. This line is unique in cosmology as, for $\lambda > 21\,\mathrm{cm}$, it is the dominant astronomical line emission for all redshifts. Hence, to a good approximation the wavelength of a spectral feature can be converted to a redshift without having to first identify the atomic transition. In principle, HI intensity-mapping could be used to make 3D maps of matter at all redshifts up into the "dark ages" ($z \approx 100$), even before galaxies have formed.

The first HI intensity mapping observations began over a decade ago [2, 108, 95, 30, 90] and interest has continued to grow [15, 83, 125]. A number of dedicated projects have been launched to detect the signal and turn the technique into a useful cosmological tool. These are mainly interferometers, such as CHIME [19, 99], Tianlai [31, 137, 38, 79, 135], MWA [130], LWA [52], HERA [44], HIRAX [100], and PUMA [125], but also include single dishes with multiple feed antennas, such as BINGO [21, 46, 136] and FAST [64]. Intensity mapping instruments can address questions at a variety of redshift ranges. At $z \sim 10$ they probe the Epoch of Reionization (EoR), star formation, and galaxy assembly, while at lower redshifts they trace large scale structure for studies of dark energy, etc. [108, 27, 23, 2, 30, 90, 95].

So far, the HI signal has not been detected using intensity mapping by itself. Intensity mapping observations, in the post-recombination epoch, have detected HI when cross-correlated with galaxy redshift surveys [92, 91, 13]. A number of challenging systematic effects must be overcome to allow autocorrelation detections. The foremost of these is separating the HI signal from Galactic and extra-Galactic astronomical foregrounds, which are $\sim 4 - 5$ orders of magnitude brighter [83]. The Sun represents an astronomical foreground that is even brighter and of a different character.

The daytime data from radio interferometer arrays in general, and the Tianlai dish array in particular, are contaminated by the solar signal, making the data unusable for most astronomical analyses. The lost data have a significant impact on observing efficiency; reaching the required survey sensitivity means observing the sky for almost twice the number of days. This penalty is particularly problematic for HI intensity mapping, where long integration times (months or years) are necessary to detect the

HI signal. Furthermore, this data loss prevents obtaining continuous, 24 hr data sets, which allow dense coverage of the uv plane and facilitate detection of periodic signals. Furthermore, not having 24 hours of continuous usable observations prevents the application of m-mode map making techniques [123]. The (u, v) coverage can still be quite good with nighttime data, and one can recover full 24 hours RA coverage by combining nighttime data from observations about 6 months apart. In this paper we try to remove the solar contamination from the daytime data from a radio interferometer. We have used the data from the Tianlai dish array as our test sample. However, the problem is not unique to Tianlai; the same algorithm may be used for other radio interferometric observations. While daytime observations with single dish radio telescopes are also plagued by the Sun's signal, this algorithm is only applicable to interferometer arrays.

The Tianlai Project is led by the National Astronomical Observatory of China (NAOC). It consists of two pathfinder radio interferometers: an array of cylinder antennas and an array of dishes, at a radio-quiet site in Xinjiang, China [32, 79, 135]. The objective is to obtain high fidelity 3D images of the northern sky using HI intensity mapping. The analysis described in this paper concentrates on the dish array data. The Tianlai dish array consists of 16 steerable, 6 m diameter dishes; a schematic is shown in Fig. 5.2, which also shows the dish numbering scheme. We use these dish numbers for referring to different baselines in the paper. The dish array currently operates between 685 MHz and 810 MHz, corresponding to redshift $0.75 < z < 1.07$, divided into 512 equally spaced frequency bins of width 244 kHz ($\delta z = 0.0002$). The 16 dual-polarization feeds yield 32 autocorrelation visibilities and $32 \times (32-1)/2 = 496$ cross-correlation visibilities, which are currently sampled every second. The system noise temperatures for the dish antennas are $80 - 85$ K [79, 140, 38, 135].

Different methods have been proposed to remove broadband RFI and the solar contamination [104, 26] based on SVD decomposition and other methods in the time and frequency domain. The solar contamination is extremely strong in the Tianlai data and methods using SVD decomposition in the time and frequency domain remove the background signal along with the solar contamination. The objective of this paper is to describe an eigenvalue-based approach that operates in the baseline space for removing solar contamination from radio interferometric data without affecting the background signal. We propose an algorithm, AlgoSCR, that can remove most of the solar contamination, provided the Sun is the strongest source in the sky along with other weaker sources. The paper is organized as follows. In the second section, we discuss the solar contamination problem in the Tianlai dish array in detail. The third and the fourth sections give the detailed algorithm for removing the solar contamination. We also show the results of our analysis on the real Tianlai data. To test what fraction of the Sun signal can be removed by our algorithm, and how much signal from other cosmic sources is removed by it, in section five we perform two tests. First, we apply it to Sun-contaminated Tianlai data and compare the cleaned visibilities to the same sidereal times observed during nighttime. Second, we apply it to simulated data where the amount of solar contamination and foreground point sources are known. In the discussion section we assess the efficacy of the method and the issues that we face when applying it. We also describe some future directions to pursue with this approach.

6.2 The solar contamination problem

The Tianlai data show strong contamination from the solar signal during the daytime. In Fig. 6.1 we plot the sum of the absolute visibility from 10 frequency

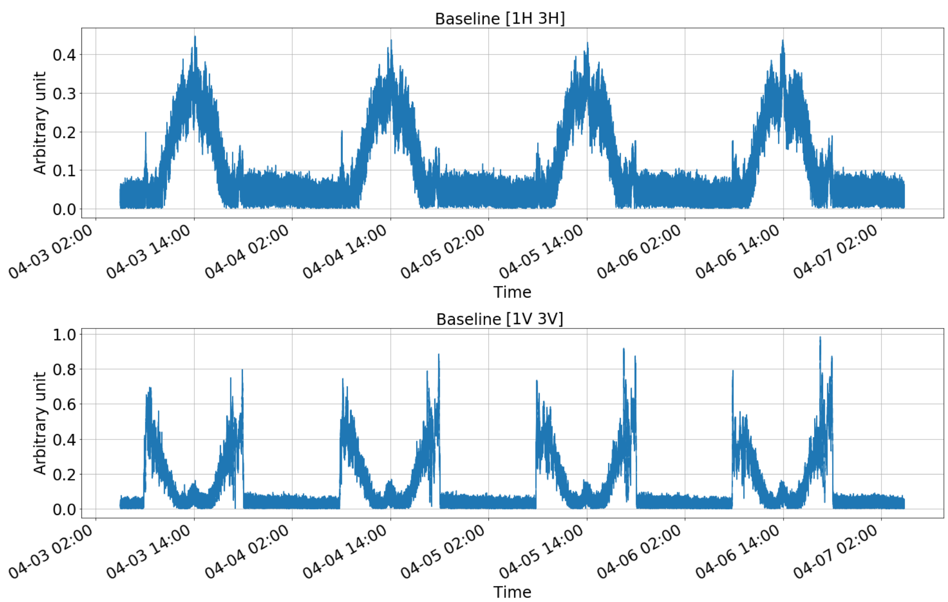

Figure 6.1: Value of the (uncalibrated) cross-correlation visibility amplitudes averaged over the 10 central frequency bands during 4 days of observations of the NCP in April, 2019. Integration time is one second. Each plot corresponds to a different baseline, as indicated. The baseline numbering scheme appears in Fig. 5.2. Time is given in local time.

channels (out of 512 channels) at the center of the band, for 4 consecutive days (total 96 hours) for two different baselines. The horizontal axis shows the time in hours, starting at the beginning of the observations. The 2 different plots are for 2 representative visibilities. We can see a roughly smooth visibility amplitude for about 10 hours every day and then a sudden increase in the absolute visibility and a noisy pattern for about the next 14 hours.

The plots clearly show that the daytime signal is several times stronger than the night. The shape of the contamination pattern also varies with baseline. Some of the baselines show a bumpy feature with the strongest visibility occurring near noon, whereas for other baselines the signal is strongest during Sunrise and Sunset and shows a 'dip' feature during the daytime. The top plot is the auto-polarization visibility corresponding to two horizontal feeds, whereas the bottom plot shows an auto-polarization visibility from two vertical feeds. The auto-polarization signals from similar feeds on other baselines show roughly similar types of patterns, except for a couple of baselines. The data are taken during observations of the North Celestial Pole (NCP) in April, 2019. During this observing period the path of the Sun is located at an angle of approximately 85° from the direction of the main beam. The plot gives an overview of the magnitude of the solar contamination problem in the Tianlai dish array.

An obvious conclusion of this strong daytime visibility is that the telescopes are responding to the Sun's illumination of their far sidelobes. For the baselines measuring correlations of the H polarization, the antenna sidelobes are aligned with the direction of Sun near noon, providing a strong visibility at mid-day, while for the V polarization the Sun falls between two side lobes at noon, producing stronger signal during Sunrise

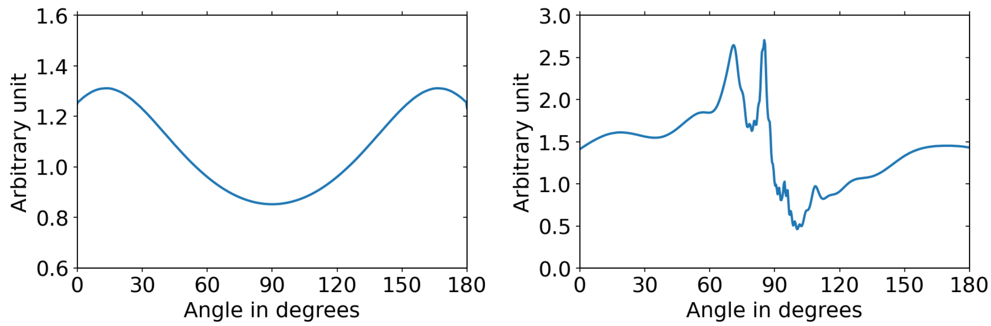

Figure 6.2: Simulated antenna directivity along the Sun's track across the beam during daytime (180° corresponding to 12 hours) when the dish array is observing the NCP in April. The scale is in arbitrary linear units (not in dB). The plot shows that for one of the polarizations, the signal from the Sun has a"dip" in power during the daytime, whereas for the other polarization there is high power during certain parts of the daytime. Even though the plots do not show the exact features that we observe during daytime (Fig. 6.1), we must remember that the sidelobes from the electromagnetic simulations are not exactly the same as those of the real antenna.

and the Sunset. These effects are consistent with the expected responses of the feed antennas, which are essentially orthogonally oriented crossed dipoles.

In Fig. 6.2, we show a cut through the simulated beam pattern for a single dish measured using an electromagnetic simulation (EM) package (CST [1]), corresponding to the Sun's track during the daytime. We can see that for one of the polarizations (left plot) we are getting a low amplitude during the midday whereas for the other polarization (right plot) the amplitude is comparatively high at noon. The simulated patterns shown in Fig. 6.2 don't exactly replicate the observed pattern of Fig 6.3, because the sidelobes from these EM simulations don't exactly match those of the real beam. The sidelobes at this particular angle are also highly cluttered. A couple of degrees change in the path gives rise to a very different shape in the sidelobes, making it difficult to reconstruct the exact pattern through such an EM simulation.

[1] https://www.3ds.com/products-services/simulia/products/cst-studio-suite/

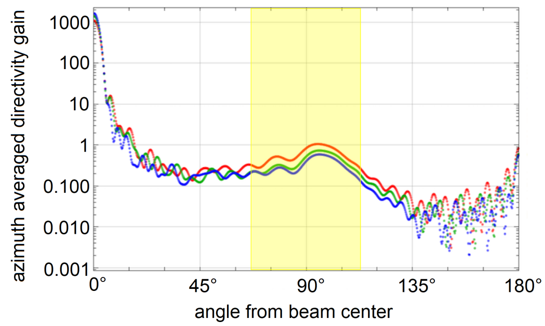

Figure 6.3: Simulated beam directivity as a function of beam angle θ from the beam center of the antennas for 3 different frequencies, 700 (red), 750 (green) and 800 MHz (blue). Each plot shows the absolute co-polar directive gain averaged over the azimuthal angle. The angle is the polar angle calculated from the center of the beam. The yellow shaded region shows the range of polar angles for which the Sun appears in the sidelobes of the beam, ranging from 66.55° at the Summer Solstice, to 113.45° at the Winter Solstice, when the dish array observes the NCP. The gain is relatively flat over this range of angles and causes the Sun signal to vary by only a factor of about 6 over the year.

This daily response to the Sun signal is relatively constant over a period of a year. Fig. 6.3 shows the directive gain of the dish antennas as computed by an EM simulation. The Sun enters the sidelobes of the antennas over a range of polar angles for which the beam patterns are relatively flat. The simulation is consistent with measurements of the daytime visibilities at different times of the year. Using the eigenvalue analysis described below, Fig. 6.4 shows the contribution by the Sun to the visibility for a typical baseline during January, 2018 and then again in April, 2019. As the paths of the Sun through the sidelobes of the antennas are different at different times of the year, the visibilities are also slightly different, but the overall patterns of the signals are similar. We can see that the amplitudes are within $\sim 30\%$ of each other. Part of this amplitude variation was induced by the variation of system gain, which is caused by the different air temperature in Jan. and Apr. The fast oscillating fringes from the Sun are present in both plots.

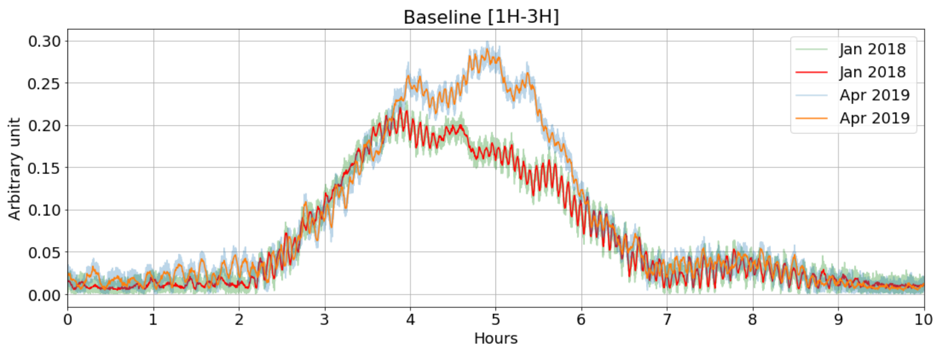

Figure 6.4: The amplitude of the daytime visibilities in January, 2018 and April, 2019. Due to the difference in the time of Sunrise in January and April, the 0 hr of each curve is adjusted so that the Sun signals from both data sets peak at about the same time. The green and blue curves are the amplitude of the visibility obtained from the telescope for January, 2018 and April, 2019, respectively. The 'fast oscillation' fringes are seen in both observations. The red and orange curves are corresponding spline fits to better highlight the fast oscillation fringes.

The complex visibility for a typical baseline is shown as a 'waterfall plot' in Fig. 6.5 for a 24 hour period.We can see that the daytime data is dominated by bright fringes caused by the Sun. On the other hand, the pattern in the nighttime data comes from the much dimmer radio sky and has a very different character. The dominant fringes in the nighttime data come from a combination of weak sources near the NCP and bright sources far from the NCP, particularly Cassiopeia A (Cas A) and Cygnus A (Cyg A).

6.3 Removing Sun contamination using eigenvalue analysis

We start by defining the notation used in this paper. The visibility matrix is given by

$$\mathbf{V} = [\mathbf{D}^{\mathrm{s}}\mathbf{G}]^{\dagger}\,[\mathbf{D}^{\mathrm{s}}\mathbf{G}] + \left\langle [\mathbf{N}]^{\dagger}[\mathbf{N}] \right\rangle, \tag{6.1}$$

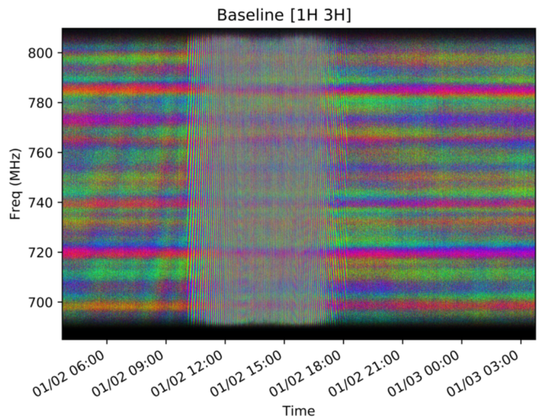

Figure 6.5: The complex visibility for a typical baseline plotted over a 24 hour period in January, 2019. We represent the phase of the complex visibility by hue (color) and the amplitude by value (brightness) in a HSV (hue, saturation, value) display of the color model (see Fig. 6.27 for details). The local time proceeds linearly from left to right, with a sampling interval of 1 s. The frequency increases linearly from bottom (685 MHz) to top (810 MHz) in 512 equally spaced frequency bins. The time interval from about 10:00 to 18:00 is dominated by the Sun.

where $\mathbf{D}^s$ is the voltage signal from the antenna in matrix form, $\mathbf{G}$ is a direction-independent complex gain matrix, $\mathbf{N}$ describes the noise from the receivers, and † represents the conjugate transpose. The individual components are given by

$$\mathbf{V}_{(i,j)} = \langle E_i^* E_j \rangle \, . \tag{6.2}$$

E_i represents the complex voltage from receiver i, with E_i^* being its complex conjugate. E_i is given by

$$E_i = \left(\sum_s D_i \left(\vec{\omega}_s \right) e^{i\mathbf{k}\cdot\mathbf{r}_i} F_s \right) G_i + N_i, \tag{6.3}$$

where F_s is the electric field of the radio wave coming from a source on the celestial sphere, $D_i(\vec{\omega}_s)$ is the primary beam of antenna i, and this is a function of the direction vector $\vec{\omega}_s$. $\mathbf{k}$ is the 3-dimensional wavenumber, the Fourier dual to the position vector $\mathbf{r}_i$ of feed i.

The intensity of the source at any frequency ν, is given by

$$I_s(\nu) = |F_s(\nu)|^2 = F_s^*(\nu)F_s(\nu)\,. \tag{6.4}$$

For extended sources we need to integrate over different directions for calculating E_i:

$$E_i \;=\; \left(\int D_i\left(\vec{\omega}_s\right) e^{i\mathbf{k}\cdot\vec{r}_i} F_s d\omega_s\right) G_i + N_i\,. \tag{6.5}$$

The visibility is an ensemble average of the $E_i^* E_j$, i.e.

$$\begin{aligned}\mathbf{V}_{(i,j)} &= \langle\, E_i^*\, E_j\,\rangle_{\tau_{\text{int}}} \\ &= \Big[\,\frac{1}{\tau_{\text{int}}}\int_0^{\tau_{\text{int}}} E_i^*\, E_j\; \mathrm{d}t\,\Big]\end{aligned} \tag{6.6}$$

where τ_{int} is the integration time, which is constant for any time and frequency bin (t,ν). For the current Tianlai setup, the integration time is 1 s. The asterisk ($*$) represents the complex conjugate and the bracket $\langle\;\;\rangle$ represents the ensemble average.

Here we should note that the visibilities from different astrophysical sources are additive. Provided there is only one point source on the sky, the visibility matrix, i.e. $\mathbf{V}_{(i,j)}$, at any time can be written as an outer product, of the electric field from the source measured at different feed antennas. Therefore, if the visibility matrix is decomposed into its corresponding eigenvalues and eigenvectors, there should be only one nonzero eigenvalue. In the presence of other weaker sources, the largest eigenvalue should correspond to the Sun signal and the eigenvector corresponding to the largest eigenvalue will roughly point toward the direction of that source in the eigenspace. The contributions from additional, weaker sources and noise may alter the direction slightly.

Comparing the visibility amplitudes between the daytime and the nighttime data in Fig. 6.1, we can infer that the largest contribution to the daytime signal is from the Sun, entering through the antenna sidelobes. Therefore, in the eigen-decomposition of the visibility matrix the largest eigenvalue should represent the solar contamination.

6.3.1 Issues with the autocorrelation signal

The voltage from the feeds contain a contribution from the receiver noise. Therefore, the measured signal or voltage E_i for a given feed i is the sum of the sky signal, $E_{\mathrm{Sky}\,i}$ and the instrument noise, N_i, i.e. $E_i = E_{\mathrm{Sky}\,i} + N_i$.

Under the assumption that the noise terms from separate feeds are uncorrelated, we can say that the ensemble average of the noise from feed i and feed j is zero, i.e. $\langle N_i^* N_j \rangle \approx 0$. Therefore, the visibility for cross-correlated feed i and j, where $i \neq j$, is $\mathbf{V}_{(i,j)} \approx \langle E^*_{\mathrm{Sky}\,i} E_{\mathrm{Sky}\,j} \rangle$.

However, for the autocorrelations, the visibilities, $V_{(i,i)}$ are dominated by the positive noise term $\langle N_i^* N_i \rangle$. The amplitudes of the autocorrelation signals are much higher than those of the cross-correlation signals. Therefore, in an eigen-decomposition of the visibility matrix, the eigenvectors are dominated by the noise signals from the autocorrelation, as the sky signals are typically much smaller than the noise.

It is not possible to ignore these auto-correlation signals or simply set them to 0 during the eigenvalue decomposition. To overcome this difficulty, we replace the

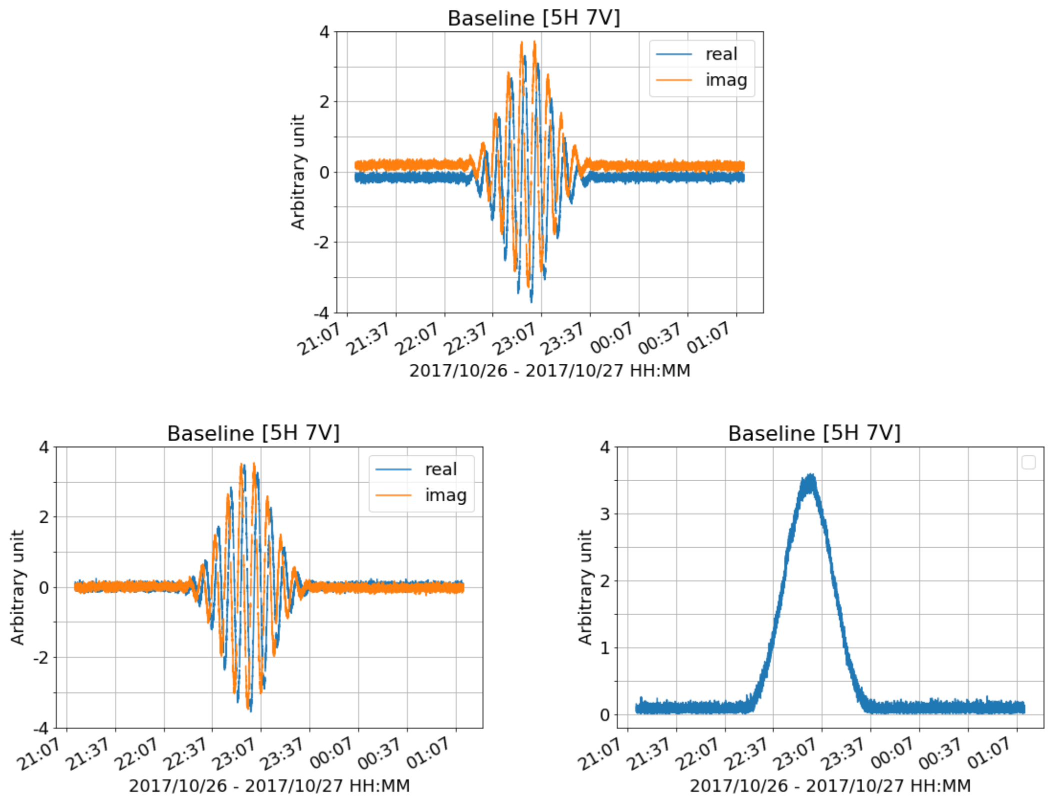

Figure 6.6: Top: The real and imaginary components of the raw visibility of baseline [5H 7V] during transit of Cas A in October 2017. We can see that there is a small DC offset in both the real and imaginary components. Bottom left: The real and imaginary components of the raw visibility after removing the offset from each of the components. Bottom right: Amplitude of the visibility of baseline [5H 7V] after removing the mean. We can see a perfectly Gaussian transit peak.

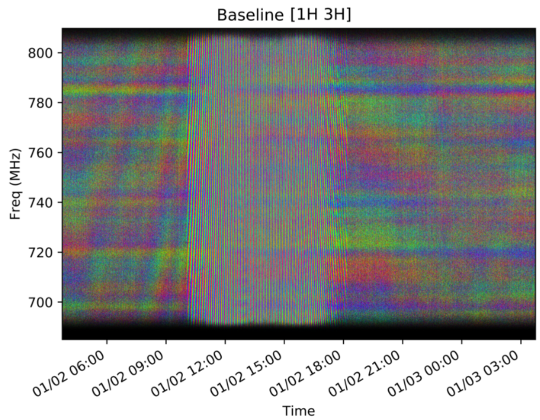

Figure 6.7: The waterfall plot of the complex visibility (same as Fig. 6.5) after the nightly mean subtraction. We can see that most of the horizontal stripes, which probably are caused by crosstalk, are now gone from the waterfall plot. The structures from the sky are more prominent.

corresponding terms in the visibility matrix by the following quantity as a proxy for the autocorrelation visibilities:

$$\mathbf{V}_{(i,i)} = \frac{1}{n}\sum_{k,j} \mathrm{abs}\left[\frac{\mathbf{V}_{(i,k)}\mathbf{V}_{(j,i)}}{\mathbf{V}_{(j,k)}}\right] \ , \quad \forall i \neq j \neq k. \tag{6.7}$$

The receiver noise component in the correlation matrix is bypassed by using (7) to replace the autocorrelations. But its long term effects remain uninvestigated. Here, n is the number of values over which we are doing the sum, i.e. the number of (j,k) pairs. This brings the level of the amplitude of the autocorrelation to the order of the cross-correlation amplitude and we can do a meaningful eigenvalue decomposition.

6.3.2 DC offset in the visibility

If there is no strong source in the sky then the real and the imaginary parts of the visibility are expected to randomly fluctuate around 0. However, often in radio interferometers, there are some DC offsets in the real and imaginary components of the visibility. The offsets may originate from a variety of systematic effects, and cross-coupling of signals between the antennas is one of them. In the Tianlai data, we see it in multiple baselines as colored horizontal stripes in the waterfall plots of the complex visibility (see Fig. 6.5).

In the top plot of Fig. 6.6, we show the real and the imaginary parts of the visibility from a transit of Cas A observed by baseline [5H 7V]. The amplitude of the visibility during the transit is expected to form a Gaussian profile. However, as there is some DC offset, we can expect the plot to show some wavy feature modulating the Gaussian. To prevent this we need to remove the DC offset. In the middle panel of Fig. 6.6, we show the real and the imaginary components of the visibility, after subtracting the mean of the nighttime data from both the real and imaginary components of the visibility. We remove the night-time mean from each frequency channel and each baseline. The amplitude of the visibility, after DC offset removal, shows a Gaussian peak during the transit of Cyg A, as expected and is shown in the bottom panel of the same figure.

We are investigating the source of the DC offset. However, the nighttime mean subtraction substantially reduces the night-to-night variation in absolute terms and as a fraction of the remaining signal, as discussed in [135]. This nightly mean subtraction removes much of the correlated noise as well as a significant fraction of the signal (gain times sky). Because the sky signal should be the same at the same local sidereal time, it does not contribute to the nightly variation that can be caused by variations in gain or correlated noise. If the variations were due only to gain fluctuations, we would not see a decrease in fractional variation. Thus, much of the subtracted signal is correlated noise.

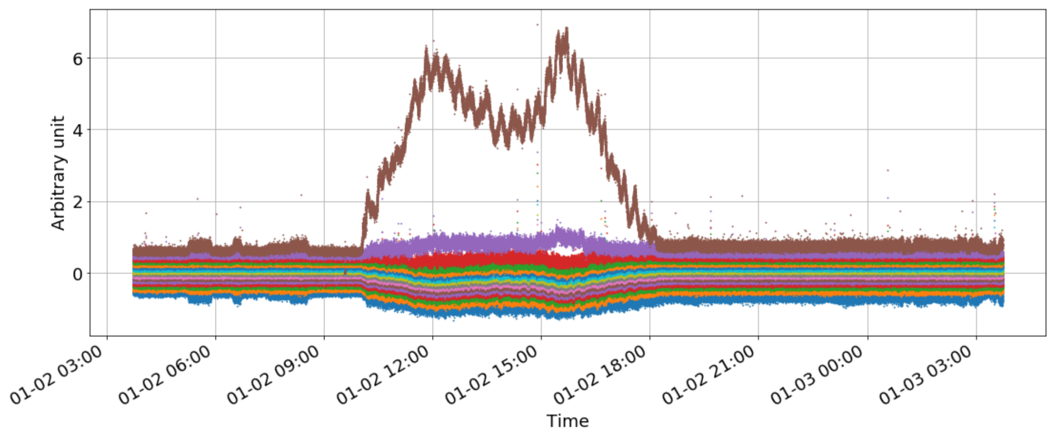

Figure 6.8: Plot of all the 16 eigenvalues in the eigen-decomposition of the horizontal polarization visibility $\mathbf{V}^{(H)}$. We can see that one of the eigenvalues is much larger than the other eigenvalues during daytime. This particular eigenvalue is coming from the solar contamination of the daytime data. We can see that the other eigenvalues are also affected during daytime. This happens due to the change in the eigenvectors, one of which (the eigenvector corresponding to the largest eigenvalue) is oriented towards the Sun during the daytime. The plot here is shown at the central frequency (747.5 MHz) of the observed Tianlai Dish Array band. All other one-dimensional plots also use this frequency.

The presence of this DC offset may also introduce an error in the eigen-decomposition and it must be removed before running the Sun removal algorithm described below. We subtract the mean value of the real and imaginary parts of the visibility for each night of data. We do not include the daytime data when computing the mean, because it is contaminated by the Sun. However, the DC offset is very stable over each night and from night to night. So we remove the nightly mean from the entire 24 hours of data, including the daytime data.

In Fig 6.7, we show a waterfall plot of the complex visibility from one baseline after the nighttime mean removal. We can see the nighttime structures more prominently after the mean subtraction.

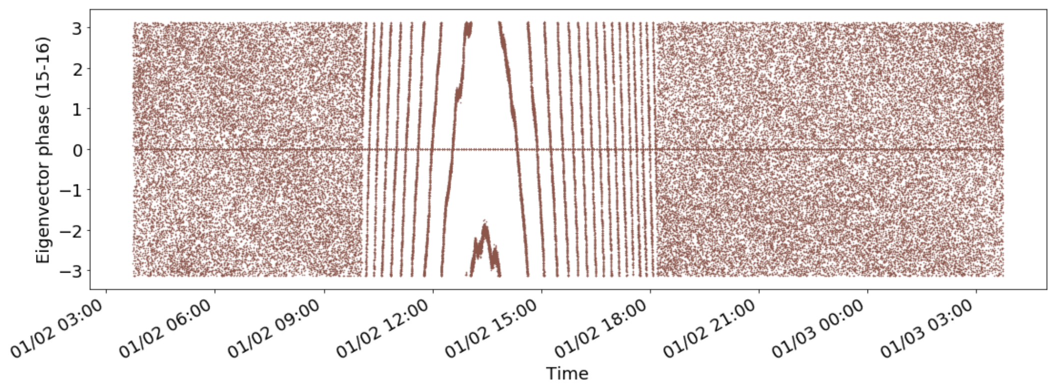

Figure 6.9: Phase plot of the 15th component of the eigenvector corresponding to the largest eigenvalue of the horizontal polarization visibility, $\mathbf{V}^{(H)}$. We can see the strong fringes during daytime, which confirms that the eigenvalue is coming from a single strong source, the Sun. During night, as there is no single strong source, the phase is varying randomly. The horizontal line in the center is caused by the calibration noise source, which is turned on and off periodically.

Note that, for simplicity, we have only considered the auto-polarization signals. If we use both the auto-polarization and cross-polarization signals, we expect to get two large eigenvalues, each corresponding to one of the polarizations. However, at present, we are in the process of understanding different systematic effects involved in measuring the cross polarization signals in the Tianlai data. Different systematic effects, e.g. mutual coupling between two feeds in the same dish, which are in close proximity to each other, are more complicated for cross-polarization data than the same polarization and require detail investigation both in data analysis as well as the instrumentation level. These are beyond the scope of this paper. Therefore, in this paper, we set the cross polarization signal to 0, making the visibility matrix look like

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}^{(H)} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^{(V)} \end{bmatrix}, \tag{6.8}$$

where the superscripts H and V refer to the horizontal or vertical polarization, respectively. $\mathbf{V}$ is the visibility matrix at each time and frequency (t, ν).

6.3.3 Understanding the eigenvalue decomposition

For removing the solar contamination from the real data, we first remove the DC offset from all the cross-correlation channels. We write the visibility matrix for each frequency and at every time bin in the form shown in Eq.(6.8). We replace the autocorrelation signals using the formula given in Eq.(6.7) and decompose the matrix into eigenvalues and eigenvectors as $\mathbf{V} = \boldsymbol{\mathcal{E}}\boldsymbol{\Lambda}\boldsymbol{\mathcal{E}}^{-1}$. At this point it should be noted that the eigenvalue decomposition is invariant under a U(1) transformation, i.e. if we multiply the full eigenvector matrix, $\boldsymbol{\mathcal{E}}$, by a factor of $e^{i\psi}$ for any real ψ, then the corresponding eigenvalue matrix $\boldsymbol{\Lambda}$ will remain invariant. Therefore, without loss of generality, we choose the first component of the eigenvector for each time and frequency component to be real and positive.

Also, in our case the visibility, $\mathbf{V}^{(X)}$ (where $X = \{H, V\}$) is a block diagonal matrix. Therefore, the eigenvalues and the eigenvectors of the matrix will be the eigenvalues and eigenvectors from each of the blocks, i.e. $\mathbf{V}^{(X)} = \boldsymbol{\mathcal{E}}^{(X)}\boldsymbol{\Lambda}^{(X)}\boldsymbol{\mathcal{E}}^{(X)\,-1}$. For each t and ν, $\boldsymbol{\mathcal{E}}$ is a $n \times n$ matrix whose i-th column is the complex normalized eigenvector, $\mathcal{E}_i^{(X)}$ of $\mathbf{V}$, and $\boldsymbol{\Lambda}$ is the diagonal matrix whose diagonal elements, $\boldsymbol{\Lambda}_{ii} = \lambda_i$, are the corresponding eigenvalues.

Fig. 6.8 shows 16 eigenvalues calculated from the horizontal polarization matrix ($\mathbf{V}^{(H)}$) as a function of time. The plot clearly shows that one eigenvalue is much higher than the other values during daytime. We can undoubtedly infer that the major contribution to the power in that particular eigenvalue comes from the solar contamination, as the Sun is by far the strongest source in the sky during day time.

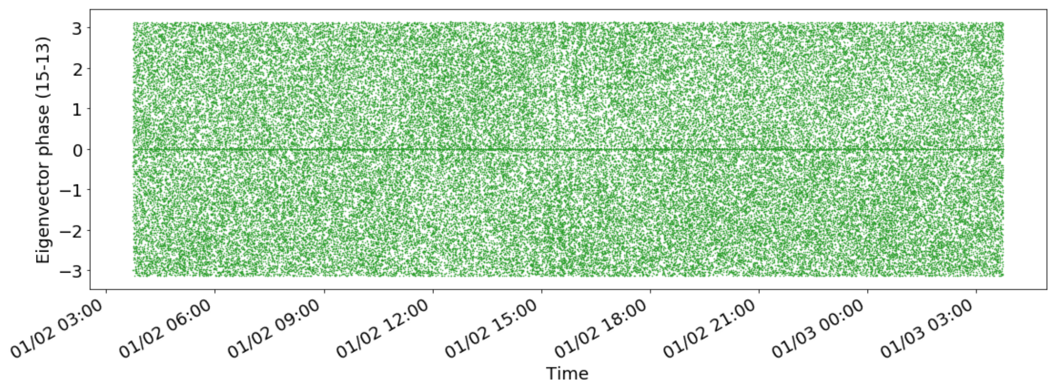

Figure 6.10: Phase plot of the 15th component of the eigenvector corresponding to the 4th largest eigenvalue in the horizontal polarization visibility $\mathbf{V}^{(H)}$. Here we don't see any fringes, showing that no individual strong source is contributing to this particular eigenvalue.

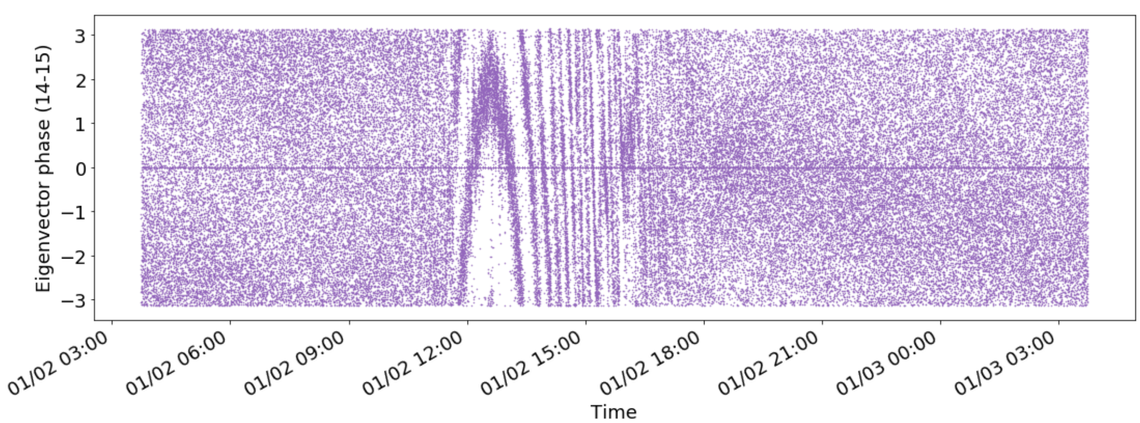

Figure 6.11: The phase of one component of the eigenvector corresponding to the 2nd largest eigenvalue. We can see weak fringes, indicating that some of the solar contribution is present in the second largest eigenvalue.

Fig. 6.9 shows the phase of one of the components of the eigenvector that corresponds to the largest eigenvalue: $\boldsymbol{\mathcal{E}}^{(X)}_{(15,16)}$. Eigenvalues are sorted according to the daytime amplitude. The 16th eigenvalue is the largest and we have shown the 15th component of the corresponding eigenvector.

Clear fringes are visible in the daytime data, showing that the daytime signal in the eigenvector is coming from a single strong source. In the nighttime data we can see the phase varies completely randomly, proving the absence of any single strong source at the nighttime data.

In Fig. 6.10 we show the phase from one of the components of the 4th largest eigenvector, $\boldsymbol{\mathcal{E}}^{(X)}_{(15,13)}$. Unlike Fig. 6.9, no fringes are visible, indicating that there is no single strong source being detected by that particular baseline and the signal is coming from the background sky. The same thing is true for any of the other smaller eigenvalues. In Fig. 6.11 we have plotted the phase from one component of the eigenvector corresponding to the second largest eigenvalue. We can find weak fringes during the daytime, indicating that some of the Sun signal has 'leaked' into this eigenvector. Ideally, the second eigenvalue represents the second strongest sources in the sky, and this leakage may be due to the presence of other sources and the background noise, which includes diffuse sources from the sky and thermal noise. In addition, the re-normalization of the autocorrelation signal using Eq. 6.7 is another possible cause of this leakage. Finally, it may also be that some of the solar radiation is being reflected from the ground and illuminating the feeds from a different direction from the main Sun signal.

If we plot the phase from any component of the eigenvector corresponding to the third largest eigenvalue, we can still see some fringe pattern in the daytime data.

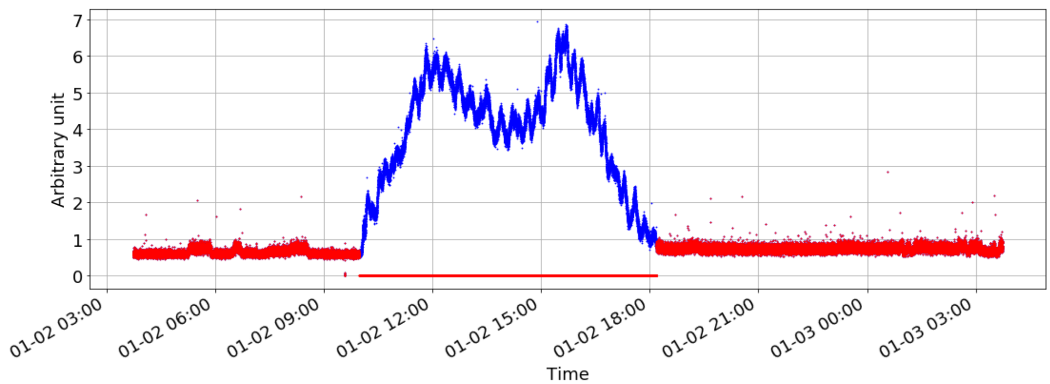

Figure 6.12: Blue: The original largest eigenvalue from the eigen-decomposition, during daytime. Red: The same largest eigenvalue but with zero value during the daytime. This step simulates removing the Sun signal in Eq. 6.11, since the largest eigenvector during the daytime points in the direction of the Sun in the eigenspace.

However, these fringes are much weaker in comparison to $\boldsymbol{\mathcal{E}}^{(X)}_{(15,x)}$ showing that the leakage of solar power is mostly restricted to the second largest eigenvalue.

6.3.4 A first attempt to subtract the solar contamination signal

Because the solar signal is contributing mostly to the largest eigenvalue, as a first step in removing the Sun signal we can set the largest eigenvalue during the daytime data to 0, and then reconstruct the visibility. In Fig. 6.12 we show the largest eigenvalue as a function of time (in blue during the daytime). The red curve shows the value after setting the largest eigenvalue during daytime to be 0. All the other eigenvalues are kept fixed. In Fig. 6.13, we show the waterfall plot of the complex visibility of one baseline that we recover after this step. The plot shows that most of the contamination is removed. However, some solar contamination signal is still discernible in the visibility plot. We can see clear, faint fringes for the baseline plotted in Fig. 6.13.

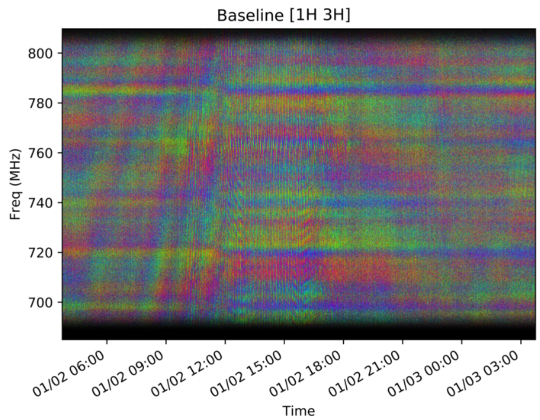

Figure 6.13: The waterfall plot of the complex visibility after zeroing the largest eigenvalue during daytime for a typical baseline. Nightly mean subtraction has been applied. There is still some residual solar contamination signal, which is causing the weak fringes during the daytime.

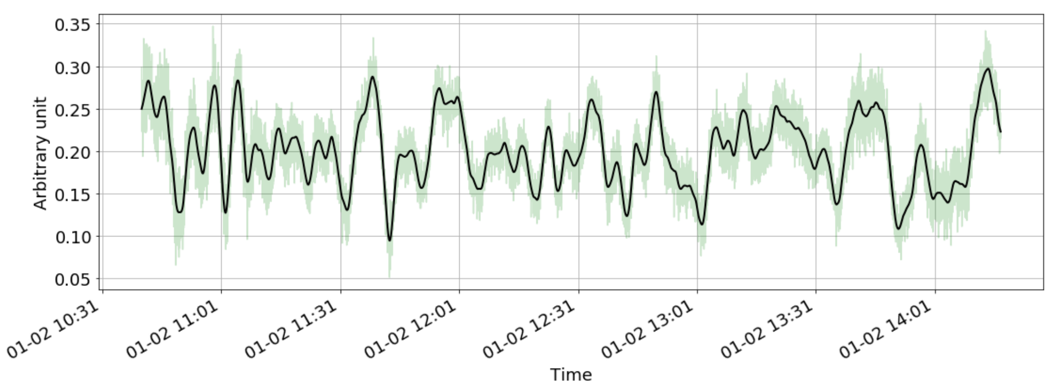

Figure 6.14: A 4-hour segment of the amplitude of one component of the eigenvector corresponding to the largest eigenvalue is shown in the light green curve. The sampling interval is 1 s. The random fluctuations in the data come from the noise. The black curve shows this component after smoothing the data, as described in section 6.4.1.

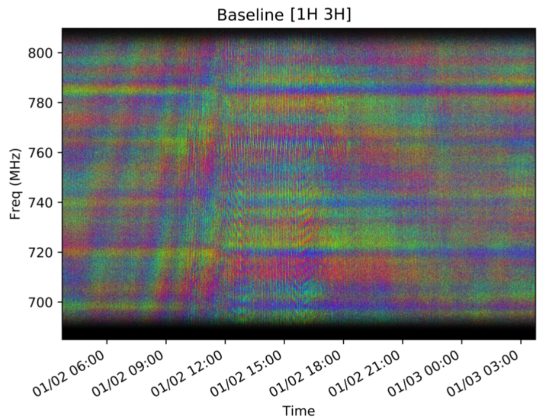

Figure 6.15: The complex visibility after smoothing the components of the largest eigenvector in spherical coordinates (described in Sec. 6.4.1, before scaling by a gain factor, described in Sec. 6.4.2) for a typical baseline. Nightly mean subtraction has been applied. There is still some signal from the Sun that has not been removed.

6.4 Improving the Sun subtraction

As we can see from Fig. 6.13, some of the solar signal still remains in the visibility matrix, mainly the signal that leaked to the second largest or even to the third largest eigenvalue. We attempt to remove this residual signal through the following 2 steps.

6.4.1 Smoothing the eigenvalues and eigenvectors

The problem with the direct eigenvalue removal method described above is that it is based on the assumption that there is a single source on the sky. This is not true in this case, as the visibility matrix contains signals from other sources as well as instrument noise. In Fig. 6.14, one component of the eigenvector corresponding to the largest eigenvalue is shown in light green over a period of 4 hours. The data, sampled every second, are noisy. However, as the Sun moves smoothly and the beam is not expected to be structured on small scales, we expect the eigenvector to vary smoothly with time. These fluctuations in the eigenvalue probably come from noise. The long term (minute level and longer) fluctuations originate in the structure of the sidelobes of the telescopes.

As the noise in the visibility matrix may cause the Sun signal to leak from the largest eigenvalue to other eigenvalues, in this section we try to reduce the effect of noise. For doing that, we fit a smooth curve (black line) through the eigenvectors corresponding to the Sun signal. This smoothed signal from the largest eigenvector is then subtracted from the original visibility to construct the Sun-removed visibility.

The cleaning routine can be summarized as follows. The visibility matrix $\mathbf{V}^{(X)}$ is first decomposed into the eigenvalue and the eigenvectors for each time and each frequency bin,

$$\mathbf{V}^{(X)} = \boldsymbol{\mathcal{E}}^{(X)}\boldsymbol{\Lambda}^{(X)}\boldsymbol{\mathcal{E}}^{(X)\,-1}, \qquad X = \{H, V\}. \tag{6.9}$$

Suppose $\mathcal{E}_S$ is the eigenvector corresponding to the largest eigenvalue, λ_S. As shown in Fig. 6.14, the direction of $\mathcal{E}_S$ will vary in every second. The n-dimensional complex eigenvectors have only $2n - 1$ degrees of freedom as we have already set the first component to be real and positive. As the eigenvectors are unit vectors, the total number of independent components becomes $2n - 2$. If we fit a smooth line through each of the $2n-1$ components, then we will overfit and the amplitude of the eigenvectors will not be 1. To keep the eigenvector normalized while doing the fitting, we express each (complex) component of the eigenvector in n-dimensional spherical coordinates and then fit a smooth line through the tangents of the angles in spherical coordinates and convert back to Cartesian space. This gives the black line, shown in Fig. 6.14. Smoothing in spherical coordinates ensures that the normalization of the eigenvector is preserved during the smoothing procedure. (See Appendix. **??** for details.)

Let the smoothed components of the largest eigenvectors be $\tilde{\mathcal{E}}_S$. If $\tilde{\lambda}_S$ is the contribution to the visibility from the direction of the eigenvector $\tilde{\mathcal{E}}_S$, then we can write, $\tilde{\lambda}_S = \tilde{\mathcal{E}}_S^T \mathbf{V} \tilde{\mathcal{E}}_S$. If we assume that this smoothed component comes from the Sun signal, then the contribution to the visibility from the Sun is given by

$$\tilde{\mathbf{V}}_S = \tilde{\lambda}_S \left[\tilde{E}_S \otimes \tilde{E}_S \right] . \tag{6.10}$$

After subtracting the Sun signal, the contribution to the visibility from the rest of the radio sky and noise is given by

$$\mathbf{V}_{\text{Sky}} = \mathbf{V} - \tilde{\mathbf{V}}_S \, . \tag{6.11}$$

In Fig. 6.15 we show the complex visibility after removing the Sun signal using this particular algorithm. In comparison to the simplest algorithm, of just removing the largest eigenvalue, this new algorithm works better. However, we can see that some of the Sun signal is still present in the visibility.

6.4.2 Scaling the Sun signal from eigenvalue analysis

The above eigenvalue analysis is based on the assumption that the signal coming from Sun is contained in the largest eigenvalue. As discussed before, this assumption is not correct because of the leakage of power into other eigenvalues.

To overcome this issue, we consider that during the daytime the signals from the sky are much smaller than the solar signal. Therefore, the Sun signal, $\mathbf{V}_S(t, \nu)$, calculated from our analysis should roughly match with the visibility $\mathbf{V}(t, \nu)$ during the daytime as the other signals are negligible in comparison to the $\mathbf{V}_S(t, \nu)$, provided that there

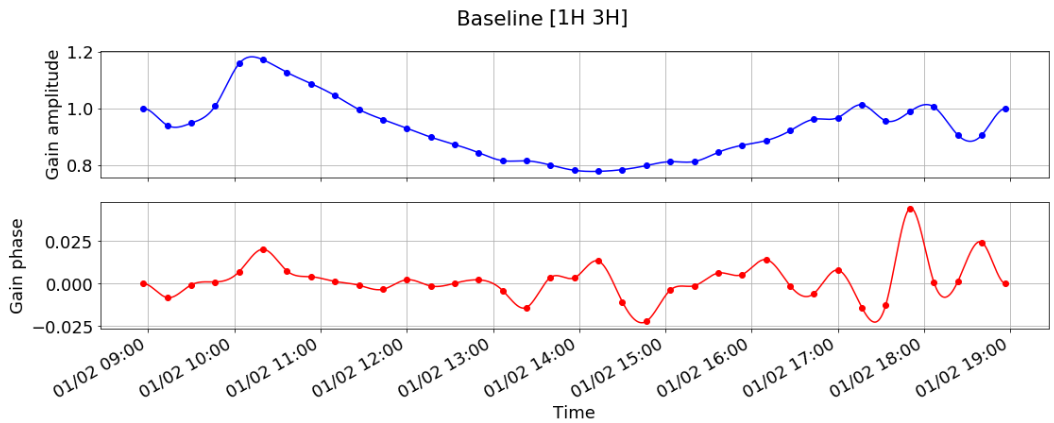

Figure 6.16: Plot of the gain $g = Ae^{i\phi}$ after χ^2 minimization for 10 hours during the daytime. Blue: Plot of the gain amplitude A. Red: Plot of the gain phase ϕ in radians. The dots represent the points of χ^2 minimization that occur every 1000 seconds. These gain values are extrapolated to the intervening points for a total of 36,000 seconds.

are no other strong sources during the day. To do that we introduce a scaling (gain) factor, $g = Ae^{i\phi}$, for each 1000 seconds (about 15 min) of daytime data and minimize

$$\begin{aligned} \chi^2 &= \sum_{t,\nu} \left[\Re(\mathbf{V} - g\mathbf{V}_S(t,\nu))\right]^2 \\ &+ \sum_{t,\nu} \left[\Im(\mathbf{V} - g\mathbf{V}_S(t,\nu))\right]^2 . \end{aligned} \tag{6.12}$$

Here $\Re(\,)$ and $\Im(\,)$ are the real and imaginary parts of the quantity inside the bracket. We get 36 gain factors (g), calculated from 10 hours ($36,000$ seconds) of daytime data. In Fig. 6.16, we show the plot of g over 10 hours of daytime, with circular dots. The smooth lines show the interpolated data. We can see that g varies smoothly throughout the day. The expectation is that the $|g|$ should be very close to 1 and very smooth, and the phase variation should be very small. This is because the Sun and the sky move smoothly through the beams over the day. As long as the Sun signal is strong enough in comparison to the background sky we can expect that the power leakage

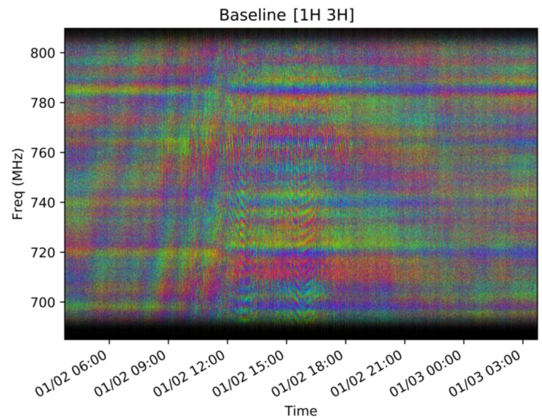

Figure 6.17: The waterfall plot of the complex visibility after χ^2-optimization scaling the Sun signal by a gain factor (Section 6.4.2) for a typical baseline.

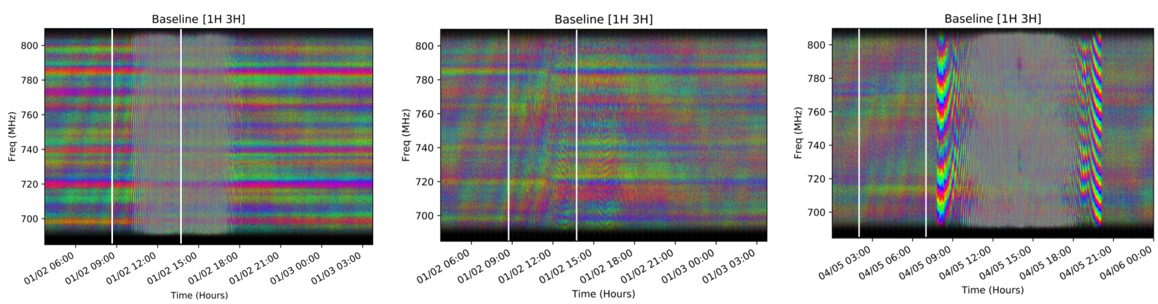

Figure 6.18: Left: Waterfall plot of original (Sun-contaminated) January 2018 complex visibility over 24 hours (same as Fig. 6.5). Middle: waterfall plot after applying removing solar contamination with AlgoSCR (same as Fig. 6.17). Right: waterfall plot of visibility data taken in April 2019. The areas between the white lines show periods of sidereal time when observations occur during daytime in January, and nighttime in April.

will vary smoothly and the gain variation should also be smooth. As the signal in the largest eigenvector and leaked power both are coming from Sun we can expect the phase variation to be minimun. Fig. 6.16 shows that the assumption is a good one in this case. However, near sunrise and sunset the amplitude and the phase change rapidly, possibly because the Sun signal is weaker at those times.

The interpolated g is used as a multiplication factor to determine the solar contribution $g \times \tilde{\mathbf{V}}_S(t,\nu)$, which is finally subtracted from $\mathbf{V}(t,\nu)$. This gives our final Sun-removed signal from the daytime data, i.e.

$$\mathbf{V}_{\mathrm{Sky}}(t,\nu) = \mathbf{V} - g_{\mathrm{int}}(t,\nu)\tilde{\mathbf{V}}_S(t,\nu)\,. \tag{6.13}$$

In Fig. 6.17 we show the complex visibility after the solar contamination removal using Eq. 6.13. We can see by visual inspection that most of the contamination signal is removed and the fringes from the weaker sources in the background sky are visible. This is the best that we get from AlgoSCR. However, on closer inspection we can see that a small amount of the Sun signal is still present in the data in the form of weak fringes. In the next section, we will make a first estimate of the performance of our solar signal subtraction and its effect on the signals from the fainter sources.

6.5 Testing the efficiency of the algorithm

6.5.1 Comparison with uncontaminated data

Due to the orbital motion of Earth, the solar signal contaminates observations made at different sidereal times, or sky orientations. Therefore, to quantify the fraction of the solar contamination removed by AlgoSCR, in this section we compare the Sun-removed visibility to the uncontaminated visibility observed during the same sidereal time at an interval of 4 sidereal months. In the left plot of Fig. 6.18, we show the raw complex visibility from January 2018 (same as Fig. 6.5). The middle plot is the visibility after the solar contamination removal using AlgoSCR (same as Fig. 6.17). Finally, the plot on the right shows the complex visibility of the sky observed in April 2019. All plots show the complex visibilities over one sidereal day. About 5 hours of the April 2019 visibility plot is not contaminated by the solar signal, and we mark that period with two white lines in all the plots.

The plots show that visibilities of the same sky orientations are similar. The large fringes in the Sun-cleaned plot coincide with those from nighttime of April 2019. Here one should note that, in the plot in the middle, we can also see some additional weak, rapidly oscillating fringes, which are not present in the April 2019 data. However, upon careful inspection, we can see these 'fast fringes' are originally present in the left plot (before Sun removal). Therefore, the algorithm does not introduce any obvious additional signal.

For understanding these results quantitatively, we take 1-minute time averages and 384-bin frequency averages, and then calculate the ratio of the residual Sun signal to the background signal over the same 5-hour range of sidereal time:

$$\frac{\sum_{t,\nu}|\mathbf{V}_{\mathrm{Sky}}(t,\nu) - \mathbf{V}_{\text{org-2019}}(t,\nu)|}{\sum_{t,\nu}|\mathbf{V}_{\text{org-2019}}(t,\nu)|} = 0.37. \tag{6.14}$$

Here, $\mathbf{V}_{\text{org-2019}}$ is the original (i.e. uncleaned) visibility from April 2019. The ratio of the original daytime signal to the background signal over the same sidereal time is:

$$\frac{\sum_{t,\nu}|\mathbf{V}_{\text{org-2018}}(t,\nu) - \mathbf{V}_{\text{org-2019}}(t,\nu)|}{\sum_{t,\nu}|\mathbf{V}_{\text{org-2019}}(t,\nu)|} = 7.48. \tag{6.15}$$

Assuming that the difference measured by Eq. 6.14 is dominated by residual Sun contamination, we see that AlgoSCR reduces the solar contamination by about 95%.

6.5.2 Comparison with simulated data

We test the efficiency of AlgoSCR when applied to simulated data sets. This provides us with another way to check the fraction of the solar contaminant signal that is removed and to determine how much of the sky signal we are erroneously removing by the analysis.

Construction of simulated data

For constructing a simulated visibility signal $\mathbf{V}_{\rm sim}$, we assume that the electric field at each feed antenna contains contributions from Sun, the sky, and noise. We have assumed that the noise variance is the same throughout the analysis.

The receiver noise is modeled as Gaussian noise in the electric field, $E_{{\rm noise}\, i}$, at the feed antenna. We consider the noise contribution to the electric field to be Gaussian in each sample.

In the Tianlai dish array the integration time in the correlator is 1-sec. The correlator takes in the data that are collected every few microseconds and averages them in an interval of 1-sec. To simulate this process, we add Gaussian random noise in the electric field with a sampling interval of 10 ms. We then calculate the noise contribution to the visibility as $V_{{\rm noise}\,(i,j)} \equiv \langle E^*_{{\rm noise}\, i}\, E_{{\rm noise}\, j}\rangle_{\tau_{\tt int}}$, where $\langle\ \rangle_{\tau_{\tt int}}$ represents the ensemble average over integration period, $\tau_{\tt int} = 1$ second. Here we have 100 data points for every second on which the average is carried out. This method also ensures that the autocorrelation visibilities follow a χ^2 distribution and the cross-correlation visibilities follow a product normal distribution. The mean and variance of $E_{{\rm noise}\, i}$

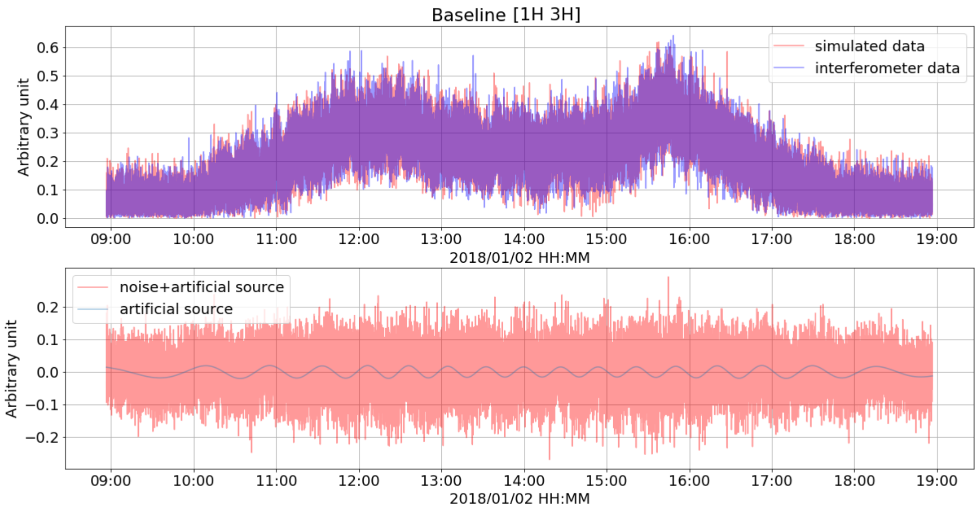

Figure 6.19: Top: The amplitude of the visibility for baseline [1H 3H]. The actual data from the Tianlai dish array are colored blue, while the simulated data are shown in red. The actual data are very similar to the simulation; regions of overlap appear purple. Bottom: In red is the real part of the noise plus the artificial sources. The blue line shows the real part of the signal from the artificial sources that is added to the data. The imaginary part (not shown) is similar to the real part.

are chosen empirically so that the simulated visibility, $\mathbf{V}_{\text{sim}}$, matches the observed visibility. The mathematical details on how to calculate the visibilities from artificial point sources in the sky are shown in Appendix 6.9.

To create the simulated Sun signal, we have taken the largest eigenvalue and correponding eigenvector from Eq. 6.22 from the Tianlai dish array data and treated it as the solar signal. The electric field for the Sun, thus calculated, is added to the simulated noise.

For the simulated artificial sources, we assume the telescope array is pointed at the NCP. The artificial sources are three made-up sources near the NCP. All the artificial sources are visible within the main beam, which is assumed to be Gaussian. Their brightnesses are chosen so that the amplitude of their combined visibility is about 10 times smaller than the noise. (An analysis with different source strengths

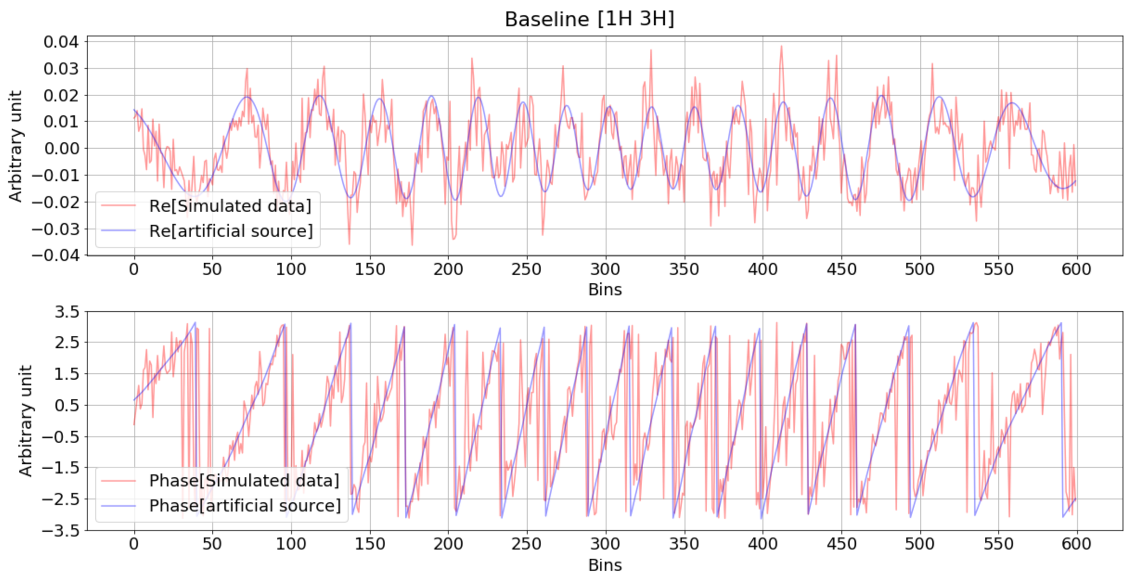

Figure 6.20: The real part (top) and the phase (bottom) of the Sun-removed visibility from simulated data for baseline [1H 3H] after 60-second averaging. The signal from the artificial sources is shown in blue.

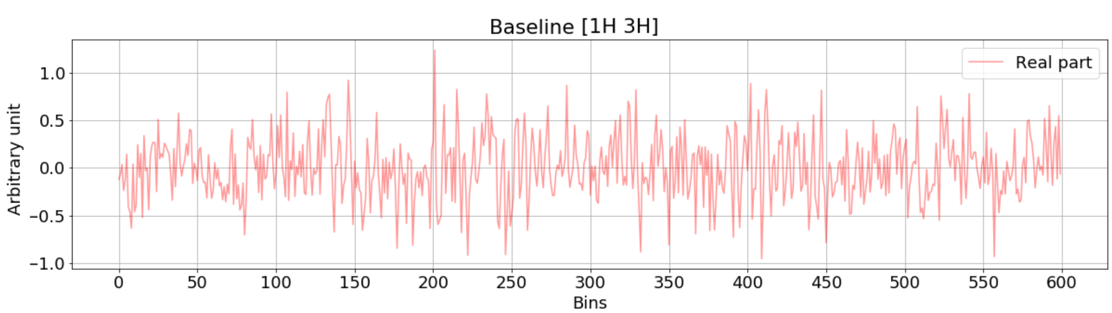

Figure 6.21: Plot of the real part of $(V_{\text{sim}} - V_{\text{org}})/\sigma$, where σ^2 is the variance of the added noise. We can see that the ratios for the real part are roughly within 3. As the noise is Gaussian, we can expect that the noise signal should be with 3σ. we conclude that the Sun removal algorithm does not introduce any significant additional noise in this analysis.

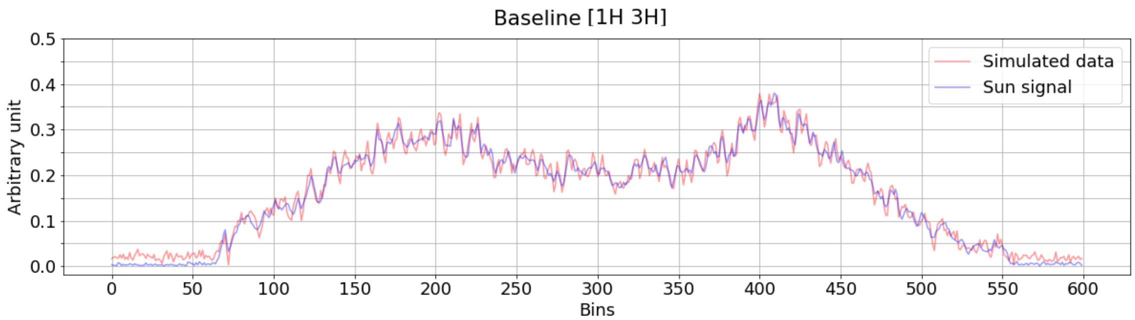

Figure 6.22: The difference between the simulated visibility (shown in Fig. 6.19) and the Sun-removed visibility (shown in Fig. 6.20) is shown in red with 1 second averaging. The signal from the largest eigenvector, which is used as the Sun signal during the daytime, is shown in blue. The data are averaged in 60 second time bins to reduce the noise.

is presented in the next section.) The artificial source visibilities are frequency- and baseline-dependent, just as visibilities from real sources on the sky. We also assume that the visibilities for the Sun and artificial sources are uncorrelated, i.e., there is no cross-term between the Sun and the artificial sources. This makes the visibilities for the Sun and artificial sources additive, as shown in Eq. 6.16. Please check Appendix 6.9 for details.

$$\mathbf{V}_{\text{sim}} = \mathbf{V}_{\text{noise}} + \mathbf{V}_{\text{S}} + \mathbf{V}_{\text{artificial sources}} \tag{6.16}$$

Results from the simulated data

We generated simulated data as shown in Fig. 6.19. The top panel of Fig. 6.19 shows the amplitude of the visibility for baseline [1H 3H] for both simulated and actual Tianlai dish array data: the plot in blue shows the actual complex visibilities from the Tianlai dish array, and the red plot shows the simulated data in our simulation (see Equation 6.16). The bottom panel shows the real part of the signal from the artificial sources

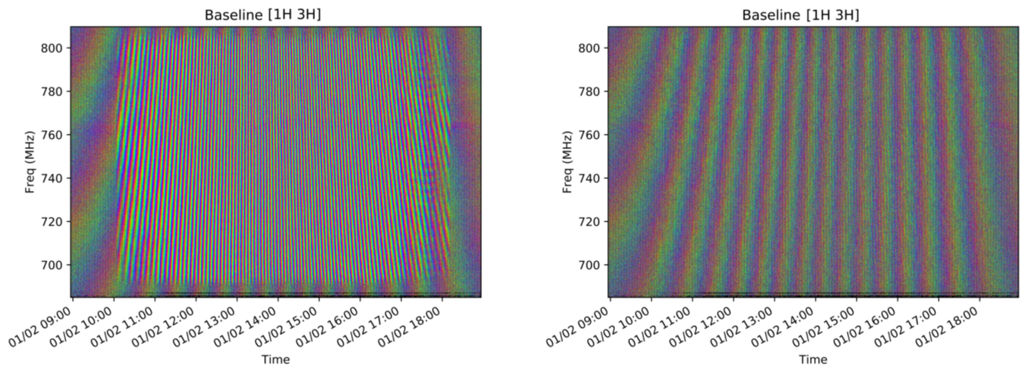

Figure 6.23: Left: Simulated daytime visibilities for a typical baseline. The bright fringes are from the Sun and the weaker fringes are from the artificial sources. Right: Sun-removed visibility using AlgoSCR. The fringes from the artificial sources are clearly visible.

(in blue). The real part of the combined signal (simulated noise and the visibility of the artificial sources that is added to the Sun) is shown in red.

We apply AlgoSCR to the simulated data. Top panel in Fig. 6.20 shows the real part of the visibility (in red) after applying the Sun removal algorithm, along with the real part of the visibility of the artificial sources (in blue) for baseline [1H 3H]. As the nature of the imaginary part will be similar, we have not explicitly shown it in the plot. The phase is shown in the bottom panel of the same figure.

The ratio of the difference between the simulated and the original visibility $V_{\rm sim} - V_{\rm org}$, and the noise standard deviation, σ, is plotted in Fig. 6.21. We can see that the ratio is within 3. As the injected noise is Gaussian, we can expect that most of the visibility should also fall within 3σ. Therefore, Fig. 6.21 ensures that the recovery of the signal using AlgoSCR does not introduce additional noise.

The red plot in Fig. 6.22, shows the difference between the amplitude of the simulated visibility (including the Sun), and the visibility that we are getting after applying AlgoSCR. This gives the contribution from the Sun in our simulated data. The blue curve shows the Sun signal that we introduced for generating the simulated data. We can see that the plots match very well. Top plot is constructed using the data from each second and the bottom plot is after averaging the data over a minute.

In the next set of plots, Fig. 6.23, we show the complex visibilities for one baseline, before and after the solar contamination removal by AlgoSCR. The visibility data show that the artificial sources that we had introduced are clearly visible after the solar signal removal, even though the source strength was much smaller than the Sun signal and the noise. This plot shows qualitatively the potential of the Sun removal algorithm. In the next section we quantify the amount of the signal from artificial sources that is removed.

Comparing efficiency of the method for different external source strengths

Here we compare the efficiency of AlgoSCR in recovering the artificial sources for different source strengths. For this analysis we use the real daytime visibility data taken by the Tianlai dish array as the base visibility. To this data we add the artificial visibility signal with different source strengths. We assume that sources are not correlated with the visibility data and the visibilities are additive.

After running AlgoSCR to remove the Sun, we using a χ^2 statistic to compare the signal with the source visibility that was originally inserted. The plot of the reduced χ^2 for baseline [1H 3H] is shown in Fig. 6.24 against the source strength. Here χ^2 is defined as $\chi^2 = \frac{1}{n_t \sigma^2} \sum_{t,\nu} |\mathbf{V}_{\rm org}(t,\nu) - \mathbf{V}_{\rm sim}(t,\nu)|^2$ during 10 hours of daytime. Here σ^2 is the noise variance and n_t is the number of time-steps, which is the number of degrees of freedom in this case. As we are sampling each second for a total of 10 hours, the number of degrees of freedom $n_t = 36000$.

We can see the χ^2 value is small for the cases in which the artificial source amplitudes are small compared to the Sun signal amplitude. As the artificial source strengths

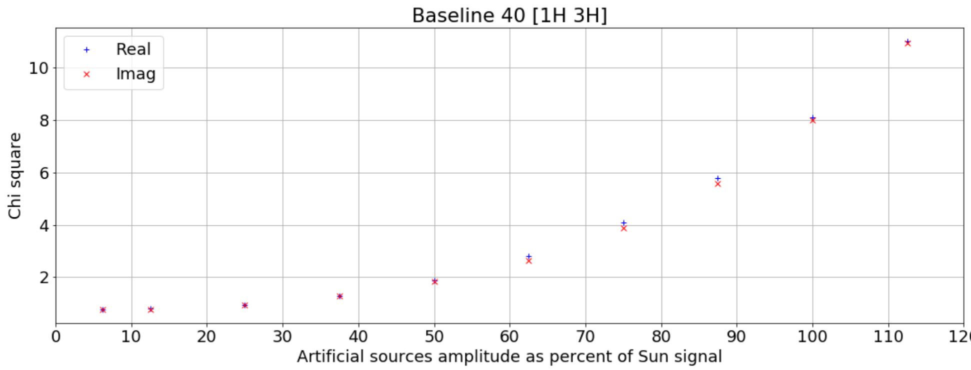

Figure 6.24: Plots of $\chi^2 = \frac{1}{n_t \sigma^2} \sum_{t,\nu} (V_{\text{org}} - V_{\text{sim}})^2$ from the real and imaginary parts of the visibility for different artificial source amplitudes. n_t is the number of sample points in the time direction and σ^2 is the noise variance. The amplitude of the original visibility V_{org} and the simulated visibility V_{sim} are shown in Fig. 6.20. We can see that the χ^2 is increasing as we increase the amplitude of the artificial sources.

increase, the fit gets worse. This is because our analysis is based on the assumption that the solar signal is the only dominant signal. As the strength of the artificial sources increases, the assumption slowly breaks down. In such cases, the largest eigenvalue starts to capture signal from the artificial sources. When the artificial sources are larger than the Sun, the largest eigenvalue provides the contribution from the artificial sources and not the Sun. In such cases we are essentially removing the artificial sources and thus the χ^2 grows quadratically.

In Fig. 6.25 we have plotted the same χ^2, where instead of dividing by σ^2 we have divided by $|V_{\text{org}}|$. Here we can see that the χ^2 is lowest when the strength of the artificial source is about 40% of the Sun signal. When the source strength is small, the χ^2 is dominated by the noise and the χ^2 is high. On the other hand, when the strength of the artificial sources is high compared to the Sun contamination as described before, the recovery gets worse.

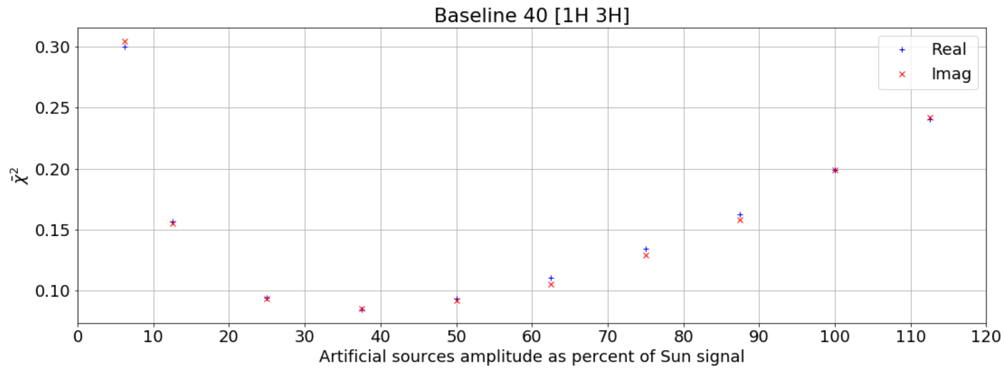

Figure 6.25: Plots of $\bar{\chi}^2 = \frac{1}{n_t|V_{\rm org}|}\sum_{t,\nu}(V_{\rm org} - V_{\rm sim})^2$ from the real and imaginary parts of the visibility for different artificial source amplitudes. Here, the plots are normalized by $|V_{\rm org}|$ instead of σ^2. We can see that the $\bar{\chi}^2$ is lowest at about 40, indicating that the recovery is best when the amplitude of the source is about 40% of the Sun signal.

6.6 Discussion

While developing AlgoSCR we explored multiple techniques and came across various issues. Here we discuss some of the points that are important in the context of optimizing AlgoSCR.

In Sec. 6.4, we address the issue of removing the residual Sun signal after subtraction of the largest eigenvalue. Here we introduce the concept of the multiplication factor g. This procedure raises the question of what will happen if instead of filtering out just the largest eigenvalue, we filter out a smooth component from the two largest eigenvalues. We find that removing the two largest components after smoothing, then the signal from the second largest component, which includes some radio sources, also gets removed, i.e. we will be removing the components from other radio sources and hence the method will not work.

In Sec. 6.4.2, while choosing the gain values, g, we calculate the gains at intervals of 15 min and then interpolate. We find that the results are fairly insensitive to the choice of time interval (say, 10 min or 30 min), as is expected because the gain varies smoothly throughout the day. However, if we choose a long time interval (several hours) for setting the gains, then we expect the results to worsen as the gain may change significantly in that time. However, we have not simulated these cases.

Another important fact that came up during our analysis is that the Sun removal algorithm works better with more baselines, i.e. if we use all 16 dishes from Tianlai instead of, say, 10 dishes, then the effectiveness of the algorithm increases. Analysis with fewer dishes increases the power leakage to other eigenvalues. The exact reason behind this is not known, but it may happen as more baselines reduce the effect of the noise in the eigen-decomposition.

In addition, if we increase the integration time from 1 second to a larger value, the results get worse, which may be due to the fact that the Sun is not a point source. This is different from co-adding the signal from multiple days, which is eventually what the Tianlai array is designed to do. However, we have not tested the algorithm on co-added signals yet.

Our analysis shows that AlgoSCR removes most of the solar contamination during the day. However, it is just a first step. We have not yet tested its effect on map-making and power-spectrum estimation. A critical next step is to make sure that AlgoSCR does not affect the statistics of the maps. This can be checked by comparing the HI power spectra and other statistical quantities from the maps produced using only nighttime data and the maps produced using the full day data after solar contamination removal. Such an analysis requires foreground subtraction and mapmaking and is outside the scope of the present work.

6.7 Conclusion

In this paper, we present a way to separate out the solar contamination from the daytime data observed by an interferometric radio array using eigen-decomposition techniques. The technique is primarily based on the assumption that if the Sun signal is the dominant signal in the sky, along with other weaker sources, and if the signals from the different sources are not correlated, then in the eigen-decomposition of the visibility matrix, the largest eigenvalue is from the strongest source, i.e. the Sun. The eigenvector corresponding to the largest eigenvalue points in the direction of that source in the eigenspace. The technique should filter out this largest eigenvalue while retaining the signals from other sources in the sky.

However, antenna gain fluctuations, noise, sidelobe gain patterns, ground reflection, thermal effects on the instruments and cables, and cross-talk between antennas introduce mixing between the largest eigenvalue and other smaller eigenvalues. For these reasons singling out and removing the Sun signal is not straightforward, and there is some residual contamination from the Sun. Therefore, we apply some novel techniques to remove the leftover Sun signal.

We have tested AlgoSCR in two ways. First, we compared the visibilities obtained by cleaning observations made during sidereal times when the Sun was up with observations made during those same sidereal times at night. We show that AlogSCR can reduce the solar contamination by a factor of 95%. Second, we used simulations to show that our algorithm is able to remove the solar contamination without removing other, weaker sources in the sky. We showed that the efficacy of the algorithm is maximum when the amplitude of the external source is about the 40% of the solar contamination signal.

The fraction of the removed background signal remains significantly small when the external source strength is within a range of 20% - 65% solar contamination.

To the best of our knowledge, this is the first published method for removing solar contamination from radio interferometer data. AlgoSCR can contribute to other ongoing and upcoming radio interferometers for solar contamination removal.

6.8 Summary of AlgoSCR

Here, we review the step-by-step procedure for Sun removal using the algorithm described above.

- For this procedure to work, first we separate the visibility $\mathbf{V}$ into the horizontal and vertical polarizations, $\mathbf{V}^{(H)}$ and $\mathbf{V}^{(V)}$, respectively. If we don't separate the polarizations, the noise and crosstalk in the same dish will give an additional large eigenvalue. The dimension of $\mathbf{V}$ is 32×32, since we have 16 dual-polarization feeds. The dimension of $\mathbf{V}^{(H)}$ and $\mathbf{V}^{(V)}$ will be 16×16.

- Remove the night-time mean from the visibility matrix $\mathbf{V}^{(X)}$: $\mathbf{V}^{(X)} = \mathbf{V}^{(X)} - \langle \mathbf{V}^{(X)} \rangle_{\text{night}}$. Here the average is over the time direction for different frequency channels. This will remove the cross-talk between the antennas.

- Replace the auto-correlations by Eq. 6.7. In practice, if the denominator, $\mathbf{V}^{(X)}_{(i,j)}$, is zero, we replace the term inside the sum by a small number, such as 0.0001.

- Perform an eigen-decomposition of $\mathbf{V}^{(X)}$:

$$\mathbf{V}^{(X)} = \boldsymbol{\mathcal{E}}^{(X)} \boldsymbol{\Lambda}^{(X)} (\boldsymbol{\mathcal{E}}^{(X)})^{-1}. \tag{6.17}$$

- For each second of integration time, let the largest (normalized) eigenvector corresponding to the largest eigenvalue, $\lambda^{(X)}_{S\,(t,\nu)}$, be $\mathcal{E}^{(X)}_{S\,(t,\nu)}$. Now $\mathcal{E}^{(X)}_{S\,(t,\nu)}$ is a vector containing $n = 16$ complex numbers. For fitting the smooth line through these vectors we calculate the tangents, $T^{(X)}_{S\,(t,\nu)}$ as:

$$T^{(X)}_{S\,(t,\nu)}(i) = \frac{\|\mathcal{E}^{(X)}_{S\,(t,\nu)}(i)\|}{\sqrt{\sum\limits_{j=i+1}^{n}\left(\|\mathcal{E}^{(X)}_{S\,(t,\nu)}(j)\|\right)^2}} \quad \forall i \in [1, n-1]. \tag{6.18}$$

- For smoothing the tangents, $T^{(X)}_{S\,(t,\nu)}$, along the time direction we apply a Butterworth low-pass filter to remove the high frequency signal. For our dataset, the filter order is 2 and the -3 dB cut-off frequency is 0.01 Hz. A 0 phase filtering is done by `scipy`'s `filtfilt` function. Let the filtered (smoothed) tangents be $\tilde{T}^{(X)}_{S\,(t,\nu)}$.

- Convert the eigenvectors back to Cartesian coordinates. For each second of integration time,

$$\begin{aligned}\|\tilde{\mathcal{E}}^{(X)}_{S\,(t,\nu)}(i)\| = \sin\left(\tan^{-1}(\tilde{T}^{(X)}_{S\,(t,\nu)}(i))\right) \times \\ \prod_{j=1}^{i}\cos\left(\tan^{-1}(\tilde{T}^{(X)}_{S\,(t,\nu)}(j))\right), \qquad \forall i \in [1, n-1]\end{aligned} \tag{6.19}$$

and

$$\|\tilde{\mathcal{E}}^{(X)}_{S\,(t,\nu)}(n)\| = \prod_{j=1}^{n}\cos\left(\tan^{-1}(\tilde{T}^{(X)}_{S\,(t,\nu)}(j))\right). \tag{6.20}$$

We then calculate the real and the imaginary parts of the of the eigenvectors

$$\tilde{\mathcal{E}}^{(X)}_{S\,(t,\nu)}(i) = \|\tilde{\mathcal{E}}^{(X)}_{S\,(t,\nu)}(i)\|\Big[\cos\left(\theta^{(X)}_{S\,(t,\nu)}(i)\right) + i\sin\left(\theta^{(X)}_{S\,(t,\nu)}(i)\right)\Big] \tag{6.21}$$

where $\theta^{(X)}_{S\,(t,\nu)}(i) = \tan^{-1}\left(\Im(\mathcal{E}^{(X)}_{S\,(t,\nu)})/\Re(\mathcal{E}^{(H)}_{S\,(t,\nu)})\right)$.

- The contribution to the visibility from the Sun is given by

$$\mathbf{V}_{S\,(t,\nu)}^{(X)} = \lambda_{S\,(t,\nu)}^{(X)} \left(\tilde{\mathcal{E}}_{S\,(t,\nu)}^{(X)} \otimes \tilde{\mathcal{E}}_{S\,(t,\nu)}^{(X)} \right) \tag{6.22}$$

where $\otimes$ denotes the outer product between eigenvector $\mathcal{E}_S^{(X)}$ and itself.

- After removing the Sun contribution, the sky contribution to the visibility is

$$\begin{aligned} \mathbf{V}_{Sky}^{(X)} &= \mathbf{V}^{(X)} - \mathbf{V}_S^{(X)} \\ &= \boldsymbol{\mathcal{E}}^{(X)} \boldsymbol{\Lambda}^{(X)} (\boldsymbol{\mathcal{E}}^{(X)})^{-1} - \mathbf{V}_S^{(X)}. \end{aligned} \tag{6.23}$$

However, the above steps still leave some Sun signal contamination, as shown in Fig. 6.13. To better remove this leftover Sun contamination, we multiply the Sun signal in Equation 6.22 by a complex factor of $g = Ae^{i\phi}$:

$$\mathbf{V}_{Sky}^{(X)} = \mathbf{V}^{(X)} - Ae^{i\phi}\mathbf{V}_S^{(X)}. \tag{6.24}$$

- To find A and ϕ, we divide the 10 hours of daytime data into 36 intervals (1000 s each) and minimize the χ^2 for each interval. Here we define the χ^2 as

$$\begin{aligned} \chi^2 = \sum_{t,\nu} \left(\Re \left[\mathbf{V}^{(H)} - Ae^{i\phi} \mathbf{V}_{S\,(t,\nu)}^{(H)} \right] \right)^2 + \\ \sum_{t,\nu} \left(\Im \left[\mathbf{V}^{(H)} - Ae^{i\phi} \mathbf{V}_{S\,(t,\nu)}^{(H)} \right] \right)^2 \end{aligned} \tag{6.25}$$

The sum is done over all seconds in the chosen interval and frequency 700.625 MHz to 794.375 MHz. We don't sum the frequency channels before 700 MHz and after

800 MHz, because we don't want to include the edge of the band-pass filter. At the end of this process, we have 36 $[A, \phi]$ pairs.

- We have 36 $[A, \phi]$ pairs corresponding to thirty-six 1000 sec intervals in 10 hours of daytime data. We use a cubic spline to interpolate a $[A, \phi]$ pair for each second in 10 hours of daytime data.

- Subtract the corrected Sun signal:

$$\mathbf{V}^{(H)}_{\mathrm{Sky}\,(t,\nu)} = \mathbf{V}^{(H)} - A_{\mathrm{int}\,(t,\nu)} e^{i\phi_{\mathrm{int}\,(t,\nu)}} \mathbf{V}^{(H)}_{\mathrm{S}\,(t,\nu)}. \tag{6.26}$$

  This gives us the final solar contamination removed result, and the results are shown in Fig. 6.16.

6.9 Calculating the artificial source visibilities

The visibilities of the artificial sources come from three made up sources near the NCP. The three simulated sources are randomly chosen to be at (RA, DEC) = (75.75,81.25), (79.5, 80.5) and (245.0,79,75) with constant brightness temperatures across all observed frequencies. We calculated the visibilty for each frequency in Tianlai's 512 frequency bins (equally spaced between 685 MHz and 810 MHz).

Each astronomical source exhibits a linearly varying phase with time and frequency, since the visibility is proportional to $e^{-i\varphi}$, where φ is the fringe phase and is defined as

$$\varphi = 2\pi\nu\tau_g(\nu, t) = \frac{2\pi\nu\mathbf{b}\cdot\mathbf{s}}{c}\,. \tag{6.27}$$

$\tau_g(\nu, t)$ is the frequency independent geometric delay and is equal to

$$\begin{aligned} \tau_g(\nu, t) \quad \equiv \quad & \frac{\mathbf{b} \cdot \mathbf{s}}{c} = \frac{b_x}{c} \cos\delta \cos H(t) \\ & -\frac{b_y}{c} \cos\delta \sin H(t) + \frac{b_z}{c} \sin\delta, \end{aligned} \tag{6.28}$$

where δ is the source declination, $H(t)$ is the source hour angle as a function of sidereal time, $\mathbf{b} = (b_x, b_y, b_z)$ are the baseline components with units of length in the radial, eastern and northern polar directions and $\mathbf{s}$ is the source vector. c is the speed of light. We can also calculate the fringe rates as follows:

$$\frac{\partial\varphi}{\partial t} = \frac{2\pi\nu}{c}[-b_x \sin H(t) + b_y \cos H(t)] \cos\delta, \tag{6.29}$$

For each second of integration time and each frequency, the visiblity for dish i and j is calculated as follows

$$\mathbf{V}_{(i,j)} = \sum_{k=1}^{n} \mathcal{F}_k A(\mathbf{s}) e^{-i\varphi} \tag{6.30}$$

where $\mathcal{F}_k$ is the flux of source k. In our simulations, we used three sources with $\mathcal{F} = (546, 170, 128)$ Jansky. $A(\mathbf{s})$ is the gain of the antenna in the direction of the source vector $\mathbf{s}$. For simplicity, $A(\mathbf{s})$, the simulated main beam gain, is taken as a Gaussian distribution with a standard deviation of 3° (FWHM = 7°), and we assume that all three artificial sources fall within the main beam. Therefore we did not model the beam sidelobe gain. $A(s)$ is also assumed to be independent of frequency. In the real Tianlai Dish beam pattern, the mainbeam (excluding the sidelobe) FWHM is about 5° ([135]). The simulated baselines are identical to the real Tianlai dish array, and the procedure for calculating the visibility is repeated for every baseline. The waterfall plots of the simulated visibilities for a few typical baselines are shown in

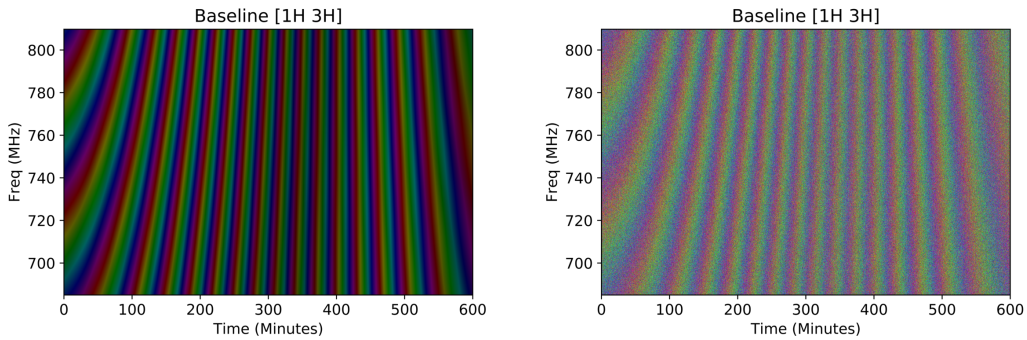

Figure 6.26: Left: Combined visibilities of three artificial sources for baseline [1H 3H], shown here for 10 hours. Right: Mock observed simulated visibility with those three artificial sources.

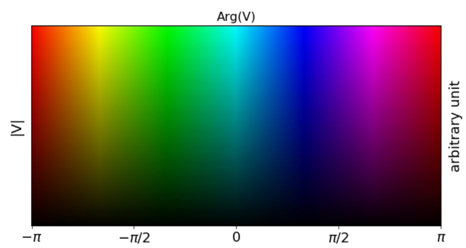

Figure 6.27: The color palette used to represent complex visibilities in this paper is shown in this plot. The phase of the complex visibility is represented by the hue and the magnitude is represented by the brightness. For more details see Appendix A of [135].

Fig. 6.26 using the same representation of waterfall plots that is used throughout this paper and is described in Fig. 6.27. As expected, longer baselines give higher fringe rates, and for a given baseline, we see a faster fringe rate at lower frequency.

Chapter 7

Foreground removal with machine learning

Astrophysical foregrounds have always been a challenge in HI detection. There are several components with different emission characteristics. Foreground removal has traditionally been done with blind techniques such as Principal Component Analysis (PCA). However, advances in computing have made foreground removal with deep learning easier to implement. One example of such progress is the `deep21` program by Makinen et al. (2021), which used deep learning on sky maps that have been partially-cleaned with PCA [87] and on which this chapter is based. In this chapter, we explore this idea further by adding cross-correlation with a mock galaxy survey and quantify the results.

7.1 Astrophysical foregrounds

The primary challenge for intensity mapping is the presence of galactic and extragalactic foregrounds with amplitudes a few orders of magnitude larger than the 21 cm signal.

In order to do cosmological analysis, we need to devise ways to mitigate the presence of foregrounds. Fortunately, the spectra of the foregrounds are mostly smooth compared to the frequency structure of the HI, and in theory it should be possible to subtract them.

Foregrounds signals can be mitigated by observing in an uncontaminated region in k space known as the EoR window. However in practice, due to mode-mixing, radio telescopes can couple the anisotropy in the foregrounds into the spectral structure with an amplitude that far exceeds that of the cosmological signal. This is exacerbated by instrumental systematics such as gain variations and beam imperfection.

Foreground cleaning is therefore required. First, we need a detailed knowledge of the beam and the gain to effectively deconvolve and remove spectral structure introduced by the instrument. Next, we need to subtract the astrophysical foregrounds, which will be discussed in the next section.

So far, HI has not been detected using intensity mapping without cross-correlation with galaxy surveys. Anderson et al. 2018 [12] reported intensity maps acquired from the Parkes radio telescope and cross-correlated with galaxy maps from the 2dF galaxy survey. CHIME performed a similar cross correlation by "stacking" HI maps on the locations of galaxies and quasars from the eBOSS catalog[34]. The intermediate step for intensity mapping requires cross-correlation before future standalone intensity mapping surveys become feasible. In this chapter, we will explore whether machine learning, together with galaxy cross-correlation, is a viable route in helping separate the foreground.

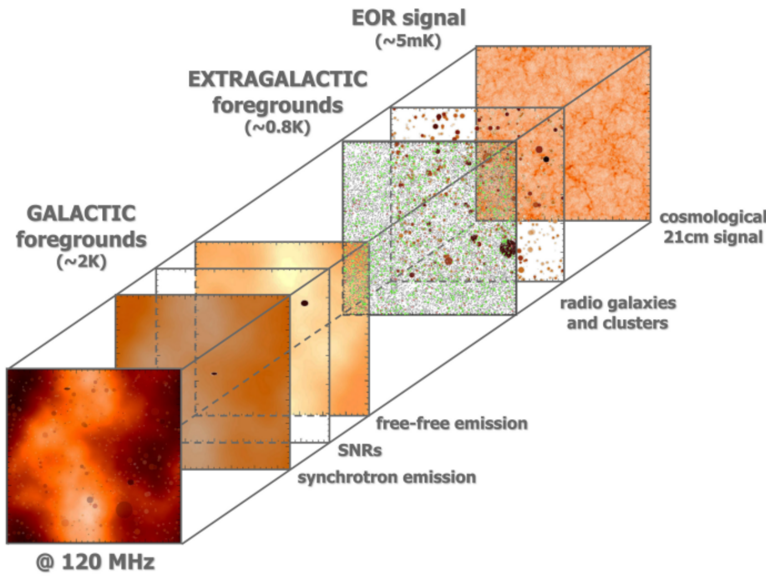

Figure 7.1: Different foreground components as a function of redshift at 120 MHz. Figure from https://ned.ipac.caltech.edu/level5/March14/Zaroubi/Zaroubi5.html.

7.1.1 Types of foregrounds

There are four main types of astrophysical foregrounds that dominate the 21 cm signal [10]:

Galactic synchrotron

Galactic synchrotron radiation is by far the largest contribution to the total radio emission. It contributes up to 75% of the foregrounds. The emission occurs when high-energy electrons are accelerated through a magnetic field. The electrons are from relativistic cosmic rays which are then accelerated by the Galactic magnetic field. Unpolarized synchrotron radiation is spectrally smooth and fairly isotropic outside the bright galactic plane. However, as synchrotron photons travel through the magnetized interstellar medium, their polarization angles change. This effect is known as Faraday rotation, and it is frequency-dependent and not spectrally smooth. It is less polarized in the galactic plane because of random, incoherent superposition. Instrumental effects can mix the polarized emission into the unpolarized part, and this can introduce erroneous cosmological structures.

Galactic and extra-galactic free-free emission

Also known as bremsstrahlung, this is the electromagnetic radiation produced by the scattering of a charged particle by another charged particle. This interaction can produce radio wavelengths similar to those of the redshifted 21 cm line, and the radiation is spectrally smooth. It can occur both inside and outside the Milky Way galaxy.

Extra-galactic point sources

Extra-galactic point sources are objects beyond the Milky Way galaxy, including sources such as active galactic nuclei (AGN). They primarily emit both synchrotron and free-free radiation. These objects follow the same matter distribution as the cosmological signal. Thus, this type of foreground is correlated with the cosmological signal at similar redshifts.

7.1.2 Foregrounds and HI simulations

Foregrounds

We lack a detailed analytical description of the foregrounds, but many numerical solutions exist. To simulate the foregrounds, we used the `CRIME` simulation package [10]. Five different types of previously mentioned foregrounds are implemented. The

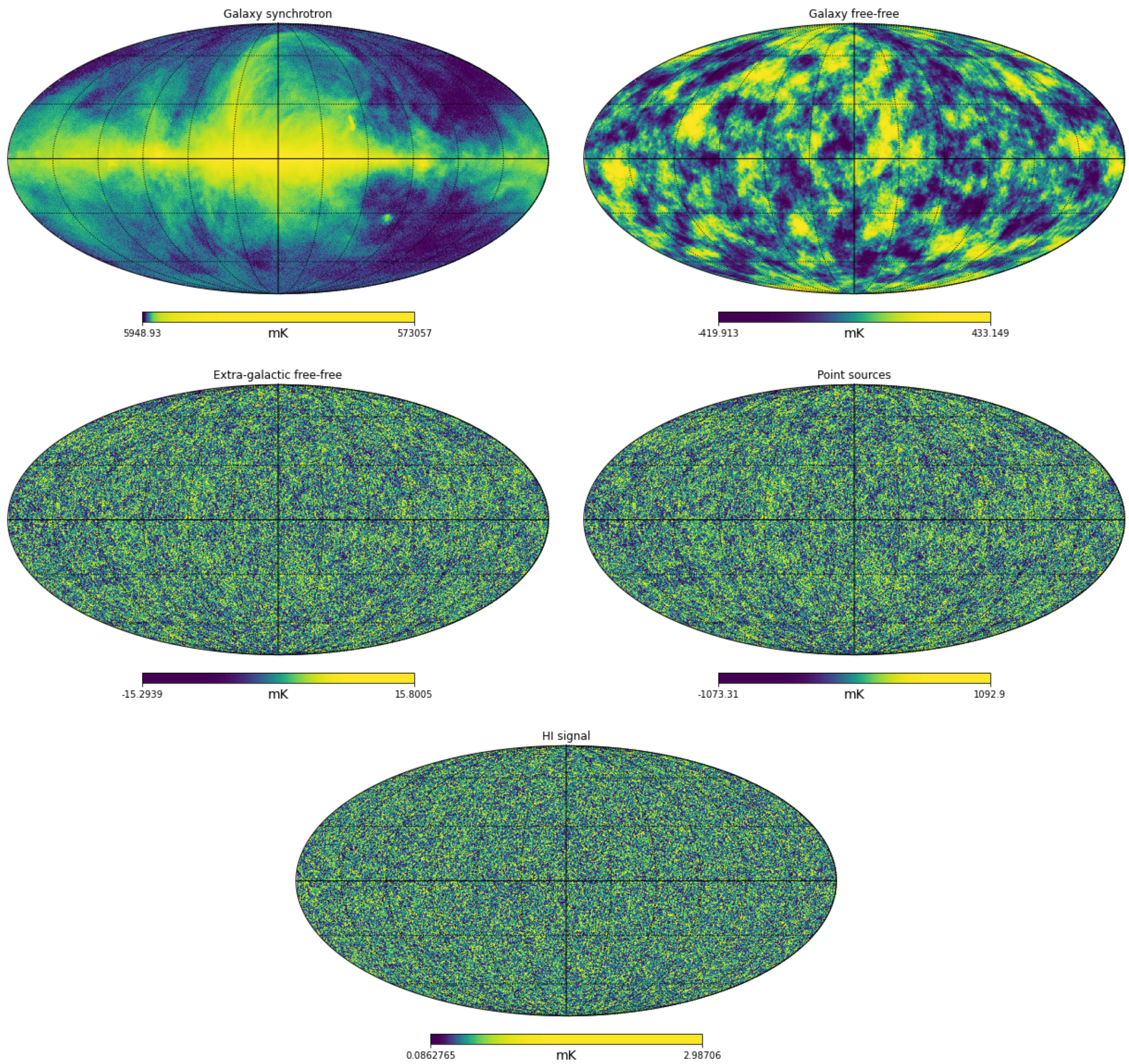

Figure 7.2: Full sky temperature maps (in mK) for different types of foregrounds at 441 MHz as generated by `CRIME`. Top row: galactic synchrotron, galactic free-free; Middle row: extra-galactic free-free, extra-galactic point sources; Bottom row: Cosmological HI signal.

Foreground Component	$A\,[\mathrm{mK}^2]$	β	α	ξ
Galactic Synchrotron	1100	3.3	2.80	4.0
Point Sources	57	1.1	2.07	1.0
Galactic free-free	0.088	3.0	2.15	35
Extragalactic free-free	0.014	1.0	2.10	35

Table 7.1: Foreground simulation parameters used in `CRIME` simulation.

galactic synchrotron radiation is the largest contribution to the total emissions. The unpolarized synchrotron should be spectrally smooth, and far from the galactic plane, can be modeled as an isotropic field. To model this and weaker foregrounds such as point sources and free-free emission, a power-spectrum model is used [118]:

$$C_\ell\left(\nu_1,\nu_2\right) = A\left(\frac{\ell_{\mathrm{ref}}}{\ell}\right)^{\beta}\left(\frac{\nu_{\mathrm{ref}}^2}{\nu_1\nu_2}\right)^{\alpha}\exp\left(\frac{-\log^2\left(\nu_1/\nu_2\right)}{2\xi^2}\right), \tag{7.1}$$

with values given in Table 7.1. To include the shape of the emission from the galactic plane, `CRIME` uses the Haslam map, which contains the full-sky synchrotron emission at $\nu_H = 408$ MHz. It then extrapolate the foregrounds to other frequencies using the Planck Sky Model to generate full sky maps of the synchrotron spectral index $\beta(\hat{\mathbf{n}})$:

$$T_0(\nu,\hat{\mathbf{n}}) = T_{\mathrm{Haslam}}\left(\hat{\mathbf{n}}\right)\left(\frac{\nu_H}{\nu}\right)^{\beta(\hat{\mathbf{n}})}. \tag{7.2}$$

Cosmological HI signal

The `CRIME` simulation generates a dark matter field and then uses a log-normal model to generate mock HI intensity maps. The random Gaussian density and velocity perturbations are generated in a cubic box of comoving size L with N_{grid}^3 cubical cells of size $l_c \equiv L/N_{\mathrm{grid}}$. For the simulations, $N_{\mathrm{grid}} = 3072$ and $L = 8850\ h^{-1}$ Mpc per side. The scales probed are approximately $2\pi/L < k < 2\pi/l_c$. The observer is placed at the center of the grid, and the signal is projected onto the observer's lightcone. The overdensity field and radial velocity in the lightcone is calculated by computing the redshift to each cell through the distance-redshift relation. At the same time the log-normal transformation on the Gaussian overdensity field is performed to generate the non-uniform HI density field. In a cell at $\mathbf{x}$ with redshift $z(\mathbf{x})$, the overdensity and radial velocity are given by [10]:

$$\begin{aligned} 1+\delta_{\mathrm{HI}}(\mathbf{x}) &= \exp\left[G(z)\delta_G(\mathbf{x}, z=0) - G^2(z)\sigma_G^2/2\right] \\ v_r(\mathbf{x}) &= \frac{f(z)H(z)D(z)}{(1+z)f_0H_0}v_r(\mathbf{x}, z=0), \end{aligned} \tag{7.3}$$

where $\sigma_G^2 \equiv \langle\delta_G^2\rangle$ is the variance of the Gaussian overdensity at $z=0$ and the factor $G(z) \equiv D(z)b(z)$ describes the growth of perturbations and possible linear galaxy bias b. The box is divided into spherical shells at different frequencies corresponding to the redshifts from the observer, using the `HEALPIX` pixelization scheme with resolution $N_{\mathrm{side}} = 256$, which corresponds to a per-pixel frequency-independent resolution of $\theta_{\mathrm{pix}} \approx 14'$. To each pixel, the mean brightness temperature associated with the hydrogen mass is calculated as

$$T_{21}(z, \hat{\mathbf{n}}) = (0.19055\ \mathrm{K})\frac{\Omega_{\mathrm{b}}h(1+z)^2x_{\mathrm{HI}}(z)}{\sqrt{\Omega_{\mathrm{M}}(1+z)^3+\Omega_\Lambda}}\left(1+\delta_{\mathrm{HI}}\right). \tag{7.4}$$

The full sky maps of the foreground and cosmological simulations are shown in Figure **??**.

Noise

To model a realistic observation, we need to add observational noise (mostly thermal noise). Observational noise can be modeled as Gaussian noise with mean zero and a variable range. The noise model has the following parameters:

$$\begin{aligned} \alpha_{\text{noise}} &\sim \log \mathcal{U}(0.05, 0.5) \\ \sigma_{\text{noise}} &= \alpha_{\text{noise}} \left\langle T_b(\nu) \right\rangle \\ \epsilon_{b,i} &\sim \mathcal{N}\left(0, \sigma_{\text{noise}}\right) \\ \hat{T}_{b,i} &= T_{b,i} + \epsilon_{b,i}. \end{aligned} \tag{7.5}$$

The variance of the noise is proportional to the average fiducial cosmological temperature at a given frequency $\langle T_b(\nu) \rangle$ to simulate noise with variable range. The observed temperature at pixel i, $\hat{T}_{b,i}$, is the sum of the true temperature, $T_{b,i}$, with some Gaussain noise $\epsilon_{b,i}$. Putting it all together, the observed temperature at a given frequency ν and line-of-sight direction $\hat{\mathbf{n}}$ consists of three components: foreground, cosmological HI signal, and observational noise:

$$T_{\text{obs}}\left(\nu, \hat{\mathbf{n}}\right) = T_{\text{fg}}(\nu, \hat{\mathbf{n}}) + T_{\text{cosmo}}\left(\nu, \hat{\mathbf{n}}\right) + T_{\text{noise}}\left(\nu\right) \tag{7.6}$$

7.2 Foreground Removal

7.2.1 Foreground Removal with PCA

As previously mentioned, the foreground and the HI signal should have different spectral signatures. While the foreground is spectrally smooth, the HI signal, while being many orders of magnitude smaller, has spectral structures. Blind foreground subtraction methods such as PCA assume that the observed brightness temperature can be decomposed along each line of sight $\hat{\mathbf{n}}$ as a function of frequency ν as:

$$T_{\text{obs}}(\nu, \hat{\boldsymbol{n}}) = \sum_{k=1}^{N_{\text{fg}}} f_k(\nu) S_k(\hat{\boldsymbol{n}}) + T_{\text{cosmo}}(\nu, \hat{\boldsymbol{n}}) + T_{\text{noise}}(\nu, \hat{\boldsymbol{n}}), \tag{7.7}$$

where N_{fg} is the number of foreground degrees of freedom to subtract, $f_k(\nu)$ are a set of smooth basis functions of the frequency, and $S_k(\hat{\boldsymbol{n}})$ are the foreground sky maps. For a particular line of sight $\hat{\boldsymbol{n}}$ measured at a discrete set of N_ν frequencies, this model can be written as a linear system:

$$\mathbf{x} = \hat{\mathbf{A}} \cdot \mathbf{s} + \mathcal{C}_{\mathbf{0}}, \tag{7.8}$$

where $x_i = T_{\text{obs}}(\nu_i, \hat{\mathbf{n}})$, $A_{ik} = f_k(\nu_i)$, and $s_k = S_k(\hat{\mathbf{n}})$. The residual, $\mathcal{C}_0$, is the sum $T_{\text{cosmo}}(\nu, \hat{\mathbf{n}}) + T_{\text{noise}}(\nu)$. Foreground removal is performed by reconstructing the residual $\mathcal{C}_0$ as accurately as possible.

PCA computes the principle components and uses them to implement a change of basis. The principle components are the directions that successively maximize variance in the data. The largest modes in a PCA decomposition represent the most correlated variables. The foreground is expected to be smooth and has high spectral correlation; removing the few largest eigenvalue is expected to preserve the cosmological signal on large angular scales since it has much lower correlation in frequency space.

We follow the PCA procedure of Alonso et al. [9] and Makinen et al. [87]. The idea is to find both the foreground components s_k and an optimal set of basis functions A_{ik} at the same time.

The combined foreground and HI maps are divided into 64 frequency bins. The frequency-frequency covariance matrix, $\mathbf{C}$, is defined as the average over $N_{\rm pix}$ pixels in the simulation:

$$C_{ij} = \frac{1}{N_{\rm pix}} \sum_{n=1}^{N_{\rm pix}} \frac{T_{\rm obs}\left(\nu_i, \hat{\mathbf{n}}_n\right) T_{\rm obs}\left(\nu_j, \hat{\mathbf{n}}_n\right)}{\sigma_i \sigma_j}, \tag{7.9}$$

where σ_i are root-mean-square fluctuations of $\mathcal{C}_0$ in mK in the ith frequency bin and each σ_i is estimated iteratively from the data. If the frequency-frequency covariance matrix contains components that are highly correlated in frequency, most of the information will be concentrated in a small set of very large eigenvalues while the other ones are negligibly small. We can then subtract the foreground by removing the components which correspond to the eigenvectors of the frequency covariance matrix with the $N_{\rm fg}$ largest associated eigenvalues.

The frequency-frequency covariance matrix $\mathbf{C}$ can be diagonalized via eigenvalue decomposition:

$$\mathbf{\Lambda} = \mathbf{U}\mathbf{C}\mathbf{U}^T = \mathrm{diag}\left(\lambda_1, \ldots, \lambda_{N_\nu}\right), \tag{7.10}$$

where $\lambda_i > \lambda_{i+1}$ for all i are the (ordered) eigenvalues of $\mathbf{C}$, and $\mathbf{U}$ is an orthogonal matrix whose columns are the corresponding eigenvectors.

We then identify and remove the $N_{\rm fg}$ eigenvalues corresponding to the foregrounds because they are much larger than the rest of the eigenvalues. We can then build the matrix $\mathbf{U}_{\rm fg}$ from the columns of $\mathbf{U}$ corresponding to the $N_{\rm fg}$ eigenvalues and find the brightness temperature along the line of sight as:

$$\mathbf{x} = \mathbf{U}_{\rm fg} \cdot \mathbf{s} + \mathcal{C}_{\mathbf{0}}. \tag{7.11}$$

The foreground sky maps $\mathbf{s}$ are found by projecting $\mathbf{x}$ onto the basis eigenvectors of the covariance matrix $\mathbf{C}$:

$$\boldsymbol{s} = \mathbf{U}_{\mathrm{fg}}^{\mathrm{T}} \cdot \boldsymbol{x}. \tag{7.12}$$

Even though PCA is able to identify and separate foreground maps, it does not produce clean HI intensity maps since it removes the mean of the observed signal. In a real experiment, the situation will be even more complicated. Cosmological signal also exhibits smoothly varying structure on large scales. Polarization leakage from galactic synchrotron can introduce erroneous structures into the cosmological signal. Intrumental noise can also be correlated in frequency.

In our analysis, we use PCA-cleaned maps before feeding them into the neural network. We repeat wthe choice of Makinen et al. [87] of removing the first three principal components from the foreground maps. Three principal components removes most of the foregrounds and reduces the amplitude in the observed maps from thousands of Kelvin to a fluctuation of a few milliKelvin around zero. Removing only two principal components only reduces the amplitudes to a few hundred Kelvins. Subtracting more modes will remove too much small-scale cosmological clustering. PCA cleaning will also be helpful for the subsequent step, as it scales the input maps to around the range $[-1, 1]$ where neural network works best.

7.2.2 Foreground cleaning with convolutional neural network

After PCA cleaning, which removes most of the foreground contamination, the residual maps are then fed into the input of a convolutional neural network (CNN). The CNN aims to recover the cosmological HI signal from the PCA residuals by minimizing

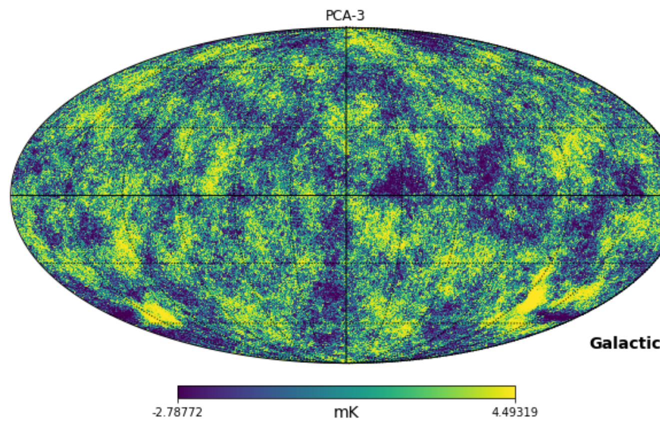

Figure 7.3: Full sky maps at 441 MHz with three largest principal components removed.

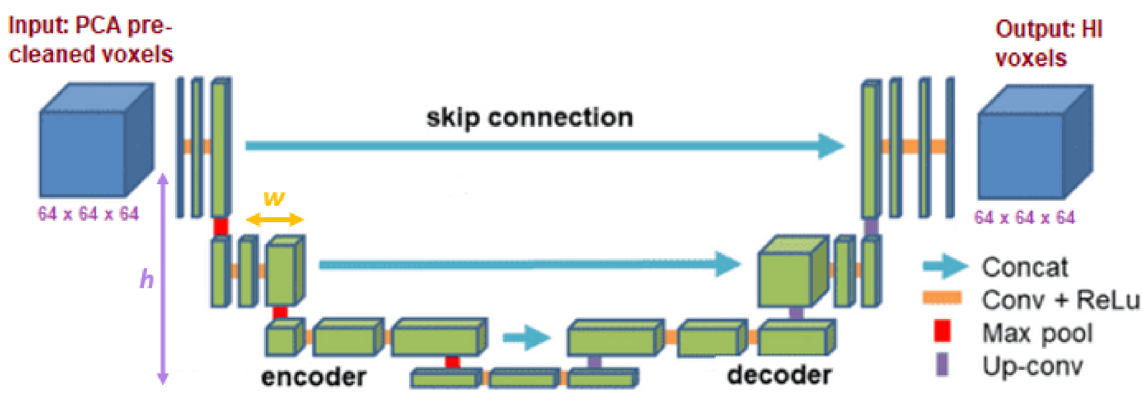

Figure 7.4: UNet architecture showing the encoder (contracting path) and the decoder (expanding path). The PCA pre-cleaned voxels of size $(64, 64, 64)$ are fed into the network. The hyperparameters to be tuned, h and w, are the number of number of down-convolutions and the number of convolutions for each convolution block.

the difference between the CNN's output maps and the simulated cosmological signal. Later, we will further improve the performance by introducing cross-correlation with mock galaxy surveys.

The full dataset consists of 100 full-sky cosmological and foreground simulations generated by `CRIME`. Each sky has an unique random seed. All the simulations assume a standard flat ΛCDM Universe with fiducial parameters $\{\Omega_m, \Omega_b, h, n_s, \sigma_8\} = \{0.315, 0.049, 0.67, 0.96, 0.83\}$. For each random seed, there are 64 frequency bins ranging from 350 MHz to 492.5 MHz with a step size $\Delta\nu \approx 2$ MHz, which corresponds to redshift $1.89 < z < 3.05$. This frequency range is chosen because it is the most contaminated with the foregrounds, especially galaxy synchrotron radiation. Each sky is divided into 192 `HEALPix` cubic voxels, so it has the shape $\left(N_{\text{voxels}}, N_{\theta_x}, N_{\theta_y}, N_\nu\right) = (192, 64, 64, 64)$. In total, we have 192 voxels/sky $\times$ 100 skies = 19200 voxels in the dataset. We do a standard 80:10:10 training:validation:test sets split. The validation set is used to tune and optimize the hyper-parameters of the model, where the test set is used to give an unbiased estimate of the performance of the final optimized model.

The CNN architecture used is the U-Net architecture, which works well on image data to achieve separation of different structures within the image. Similar to an autoencoder, it consists of a contracting (encoder) path and an expanding (decoding) path (see Figure 7.4). For each contracting step, the side of each voxel is halved while the number of filters is doubled. The filters use a stride step of 2. The hyperparameters h and w characterize the number of down-convolutions and the number of convolutions for each convolution block, and they are the parameters to be tuned. 3D convolutional kernels are used because they better capture the full angular information θ_x, θ_y as well as the spectral information ν. The data then go through the symmetric expanding path. Skip connections (or shortcut connections) concatenate features between the encoder and the decoder parts. They feed some information from encoder layer to decoder layer by skipping a few layers in between. Without the skip connections, some information would be lost during the downsampling stage. They also allow the network to extract the information on all different scales. Batch normalization is carried out at each layer to handle internal covariate shift. The final outputs from the U-Net are the foreground-cleaned voxels.

Loss function

The first objective of the foreground cleaning procedure is to minimize the difference between the predicted, p, and true value, t, of each i^{th} voxel. This difference is quantified by the loss function $\mathcal{L}(p,t)$. There are many ways to define the loss function: Mean Absolute Error (MAE): $\mathcal{L}(p,t) = \frac{1}{n}\sum_i |p_i - t_i|$, where the absolute value operation is carried out pixel-wise; Mean Square Error (MSE): $\mathcal{L}(p,t) = \frac{1}{n}\sum_i (p_i - t_i)^2$; and $\log\cosh$ loss:

$$\mathcal{L}(p,t) = \sum_i \log\cosh\left(p_i - t_i\right). \tag{7.13}$$

MSE converges faster than MAE but is more prone to over-smoothing. MAE is robust to outliers in the data, since MSE squares the error and hence will see much larger error than MAE. MAE is hence more likely to remain stable and has only small changes when exposed to noise. $\log(\cosh(x))$ is approximately equal to $\frac{1}{2}x^2$ for small x and to $|x|$ for large x and is twice differentiable everywhere. It works similarly to the mean square error, but is less affected by occasional wildly incorrect predictions. As a result, $\log\cosh$ loss is chosen for its robustness and stability.

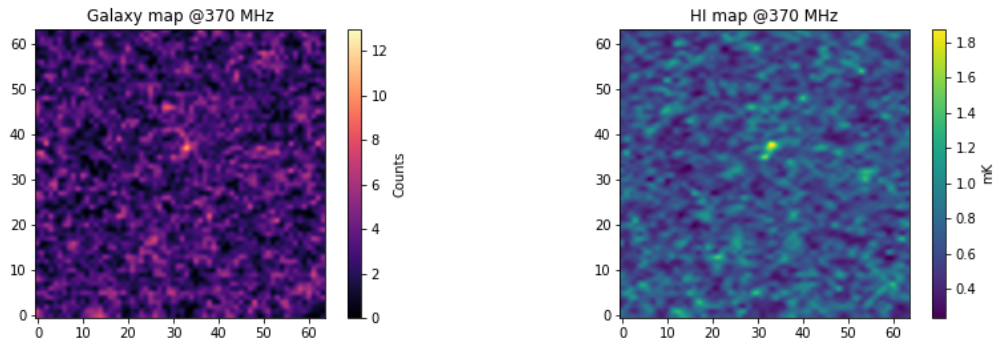

Figure 7.5: On the left, a 2D slice of the galaxy map at 370 MHz. On the right, we have an equivalent HI slice that the galaxy map was derived from.

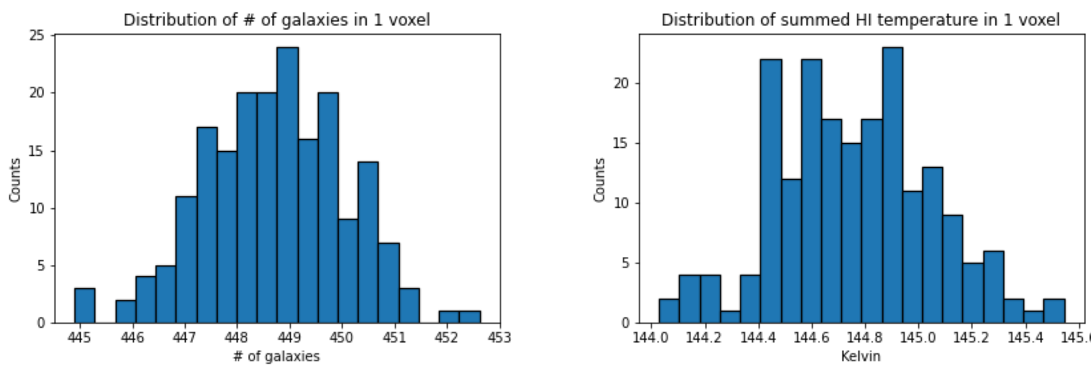

Figure 7.6: Left: Histogram for the distribution of the number of galaxies in one voxel of size $\left(N_{\theta_x}, N_{\theta_y}, N_\nu\right) = (64, 64, 64)$ over one sky consisting of 192 voxels. Right: Histogram for the summed HI temperature over the same sky. Note that the distributions look quite similar.

7.2.3 Mock galaxy surveys

Mock galaxy maps were created from simulated cosmological HI sky maps. For each voxel in the HI simulation, we draw a random sample from the Poisson distribution with the probability mass function given by $f(k;\lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$. k is the number of galaxies in a given voxel, and the expected value in each voxel, λ, is proportional to the HI temperature. This assumes that the number of galaxies found within a given volume is proportional to the amount of neutral hydrogen present. A more physical model would include a galaxy bias factor, since galaxy count is only a biased tracer of the underlying dark matter. However, we assume that the scale is large enough (each side of a voxel is ~ 7 Mpc) so that the bias factor is linear. On smaller scales, non-linear effects become more important. Figure 7.5 shows the mock galaxy maps obtained from the simulated HI maps. Figure 7.6 shows the similarity between the distribution of the number of galaxies in each voxel over one sky (192 voxels) and the distribution of the summed HI temperature over the same sky.

The second objective of this foreground cleaning procedure is to maximize the correlation between the output HI maps and the mock galaxy maps. Let the mock galaxy catalog be g. The correlation is defined as

$$\mathrm{Corr}(g, p) = \frac{\sum_i (g_i - \bar{g})(p_i - \bar{p})}{\sqrt{\sum_i (g_i - \bar{g})^2} \sqrt{\sum_i (p_i - \bar{p})^2}}. \tag{7.14}$$

In the ideal case, the correlation should approach that between the mock galaxy maps and the originally simulated HI maps (about 0.65). The loss function to maximize the agreement between the predicted value and galaxy catalog value can be defined as:

$$\mathcal{L}(g, t) = 1 - \mathrm{Corr}(g, p). \tag{7.15}$$

The correlation leaves too much freedom for the prediction, since it is invariant to additive and multiplicative shifts by a constant value. We could try the concordance correlation coefficient which addresses this issue. Here we will combine the $\log\cosh$ loss function with the correlation loss:

$$\mathcal{L}(p, g, t) = \alpha \left(\sum_i \log\cosh (p_i - t_i) \right) + \beta \left(1 - \mathrm{Corr}(g, p) \right), \tag{7.16}$$

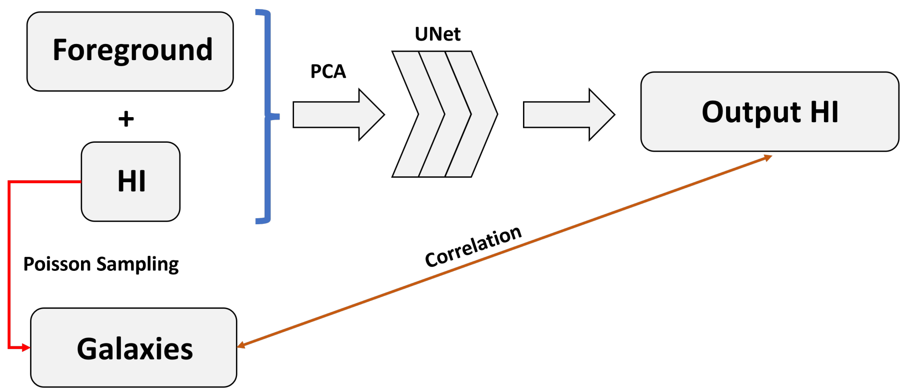

Figure 7.7: Diagram showing the overall foreground cleaning procedure

where α and β are the weighting factors for the log cosh and correlation loss functions, respectively. We refer to this as the *custom loss*. Those weighting factors are to be determined.

We tested tuning the hyperparameters with a smaller dataset (10 skies with 80:10:10 split) on a Google Colab Pro notebook with NVIDIA P100 GPU with 16GB integrated RAM. The full dataset with 100 skies is run on a Google Cloud Compute machine with 2x NVIDIA A100 GPUs with 40GB integrated RAM each. We add the foreground, HI cosmological signal, and simulated observational noise with $\alpha_{\rm noise} = 0.25$ to form mock observed maps. These serve as the input for PCA pre-cleaning (3 largest principal components removed). The PCA pre-cleaned maps are then fed into the UNet network, which outputs HI maps that we hope are free of foreground. The overall process is summarized in Figure 7.7.

For the test run, we chose $w = 3$ and $h = 3$ with a batch size of 16 (meaning the hyperparameters are updated every time the network runs through 16 training voxels). The number of training epochs is 10, so the network runs through the entire training dataset 10 times. This is enough to ensure that the loss function plateaus. We use an Adaptive Moment (Adam) Optimizer with a fixed learning rate with weight decay

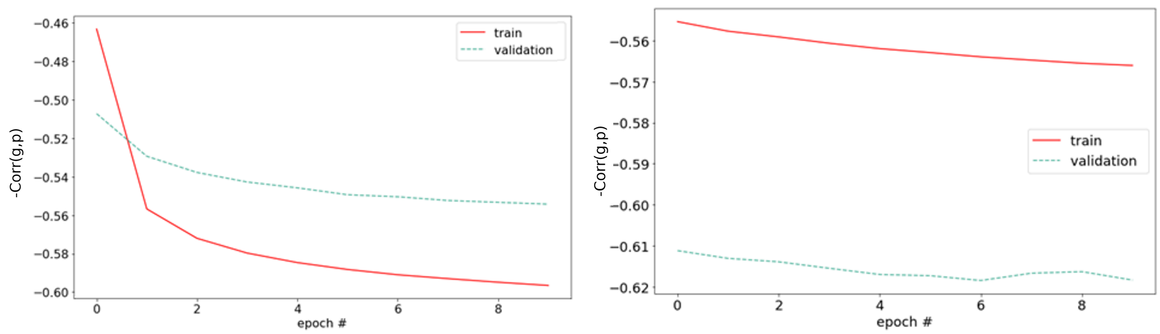

Figure 7.8: $-\text{Corr}(g, p)$ as a function of training epochs for the pure $\log\cosh$ loss function (left) and the custom loss function (right).

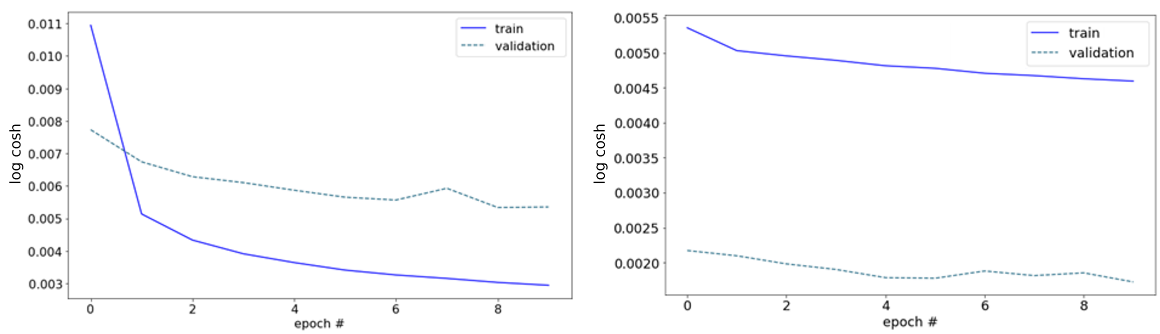

Figure 7.9: $\log\cosh$ loss value as a function of training epoch for pure $\log\cosh$ loss (left) and custom loss (right).

to minimize the loss function. Batch normalization is employed in each convolutional step. The exact value of the hyperparameter set depends on the size of the training set to achieve a balance between minimizing the loss function and avoiding overfitting the training set. The test run contains 11.68 million trainable parameters.

7.3 Results

We compare the difference between the results from using the custom loss function and the $\log\cosh$ loss function that is used in the original paper by Makinen et al. [87]. Figure 7.8 shows the negative of the correlation ($-\text{Corr}(g, p)$) as a function of training epochs for pure log cosh loss (Eq. 7.13) on the left and for the custom loss (Eq. 7.16) on the right. We see that log cosh loss function attains a higher correlation on the training set as the number of training epochs goes up, but it does not generalize as well to the validation set, implying that the model may have slightly overfitted the training data. On the other hand, the correlation for the custom loss function only goes up slightly as the number of training epochs increases, but it generalizes better to the validation data. The correlation on the validation set is roughly on par with the correlation on the training set, meaning that the model with the custom loss generalizes better to unseen data.

Figure 7.9 shows the value of the log cosh loss function for pure log cosh loss on the left and the custom loss on the right. As expected, the log cosh value is lower when we minimize with respect to the log cosh loss function. However, we see a similar behavior with the correlation case: the custom loss tends to not overfit the training data and and it generalizes better to validation data. This is due to the extra information contained in the galaxy cross-correlation.

Signal recovery

Figure 7.10 shows the average temperature in mK at each frequency for the pure log cosh loss on the left and the custom loss on the right. The average temperature for the PCA pre-cleaned map is zero at all frequencies, as expected, since the PCA procedure in the frequency direction removes the mean value at each frequency. For both types of loss function, the UNet predicted average temperature at each frequency are very close to the original HI signal despite the fact the PCA pre-cleaning step removes the mean of the observations.

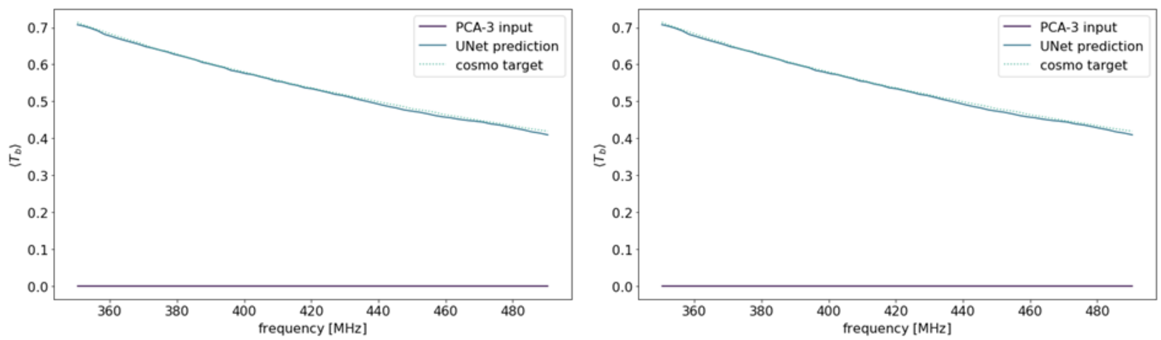

Figure 7.10: Average temperature $\langle T_b \rangle$ in mK versus frequency for the pure log cosh loss (left) and the custom loss (right).

Figures 7.11 and 7.12 show the cleaning performance for the pure log cosh loss and the custom loss, respectively, on a random `HEALPix` map at three different frequencies. In each panel, the original cosmological HI signals are shown in the left column, the PCA pre-cleaned maps in the middle column, and the UNet reconstruction in the right column. With both loss functions, the reconstructed maps are very similar to the original cosmological signal, and both do better at higher frequency where foreground contamination is less severe.

Power spectrum

As discussed in previous chapters, the power spectrum can quantify important information about the correlation between the matter over-density at all different scales. Since the matter distribution is three-dimensional, we need to consider both the angular and radial power spectra in order to capture the clustering information perpendicular and parallel to the line of sight, but here we only discuss the angular power spectrum. The angular power spectrum of the brightness temperature is found by decomposing it in terms of the spherical harmonic basis function $Y_{lm}(\hat{\mathbf{n}})$:

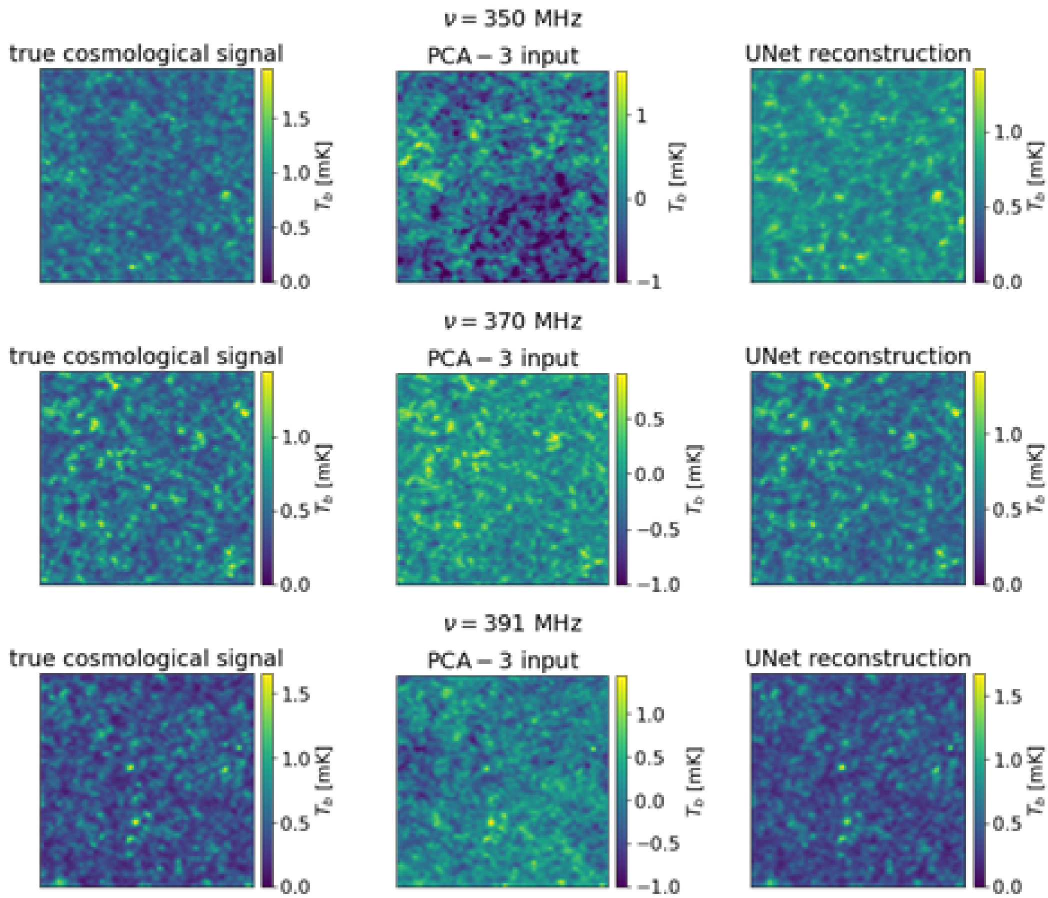

Figure 7.11: 2D slices at three different frequencies $\nu = 350, 370, 391$ MHz comparing the true cosmological signal (left), PCA pre-cleaned map (middle), and UNet prediction (right) for the pure $\log\cosh$ loss function.

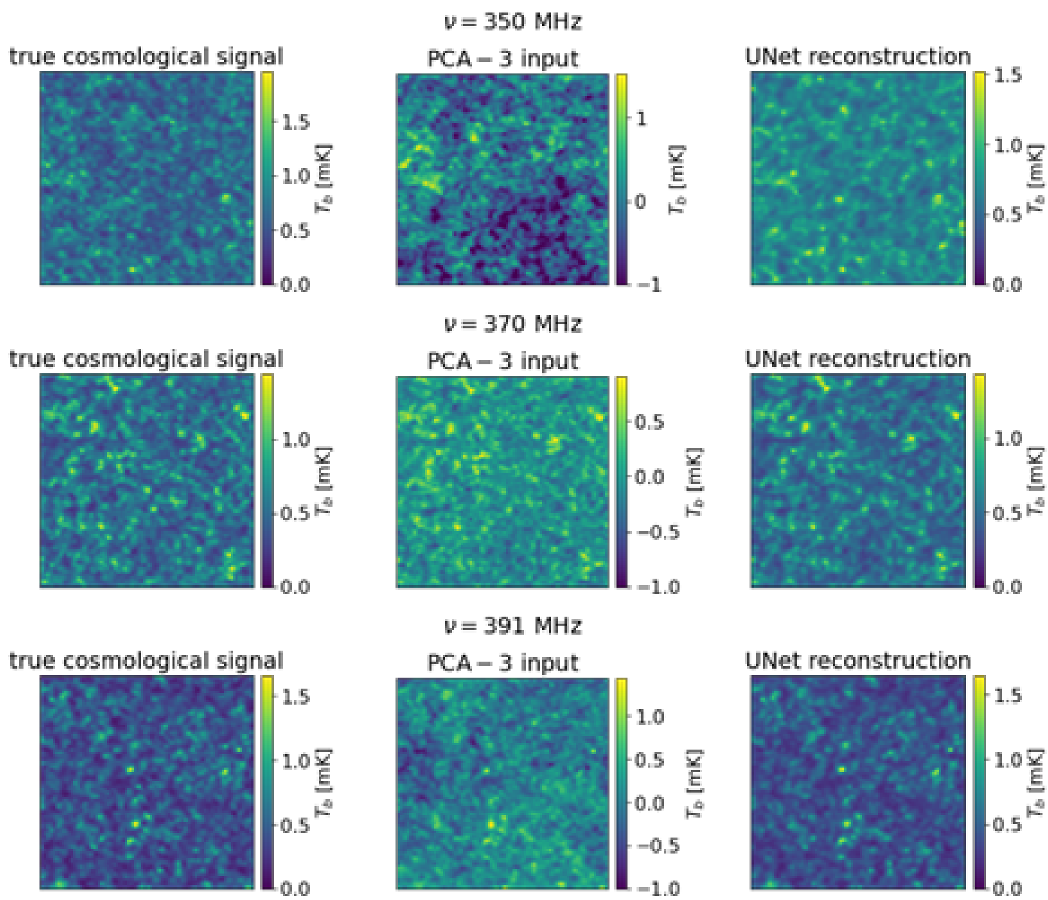

Figure 7.12: 2D slices at three different frequencies $\nu = 350, 370, 391$ MHz comparing the true cosmological signal (left), PCA pre-cleaned map (middle), and UNet prediction (right) for the custom loss function.

$$a_{\ell m}(\nu) = \int T_b(\nu, \hat{\mathbf{n}}) Y^*_{\ell m}(\hat{\mathbf{n}})\, \mathrm{d}\hat{\mathbf{n}}^2. \tag{7.17}$$

The angular power spectrum at a particular ℓ is found by adding all the squared coefficients of each m mode and dividing by the number of such modes:

$$C_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2 . \tag{7.18}$$

To characterize the cleaned map, we define the residual map as $p-t$, where p and t are the predicted and true (simulated) maps, respectively. The power spectrum of the residual map as a fraction of the true cosmological HI power spectrum is defined as

$$\rho_{\mathrm{res}}(\ell) = \frac{C_\ell(p-t)}{C_\ell(t)}. \tag{7.19}$$

We can also define the bias introduced by the cleaning method as a fraction of the true power spectrum as

$$\varepsilon(\ell) = \frac{C_\ell(p) - C_\ell(t)}{C_\ell(t)}. \tag{7.20}$$

A good foreground removal will result in $\rho_{\mathrm{res}}(\ell)$ and $\varepsilon(\ell)$ being as close to zero as possible. Figure 7.13 shows ρ_{res} as a function of ℓ for both log cosh loss on the left column and custom loss on the right column. The results are quite similar for both loss functions, with the UNet outperforming PCA on all angular scales and at all frequencies. Figure 7.14 shows the bias $\varepsilon(\ell$ as a function of ℓ for the two loss functions. The PCA method seems to have lower bias than the UNet network on the small scales at the middle frequency. However, for the custom loss function, the bias is on par with the PCA method and better than the pure log cosh loss on the intermediate angular scales. This is where galaxy cross-correlation may have an added benefit. In all cases, the UNet network is mostly scale-independent, in contrast to PCA, where the effectiveness of foreground removal is highly dependent on the scale.

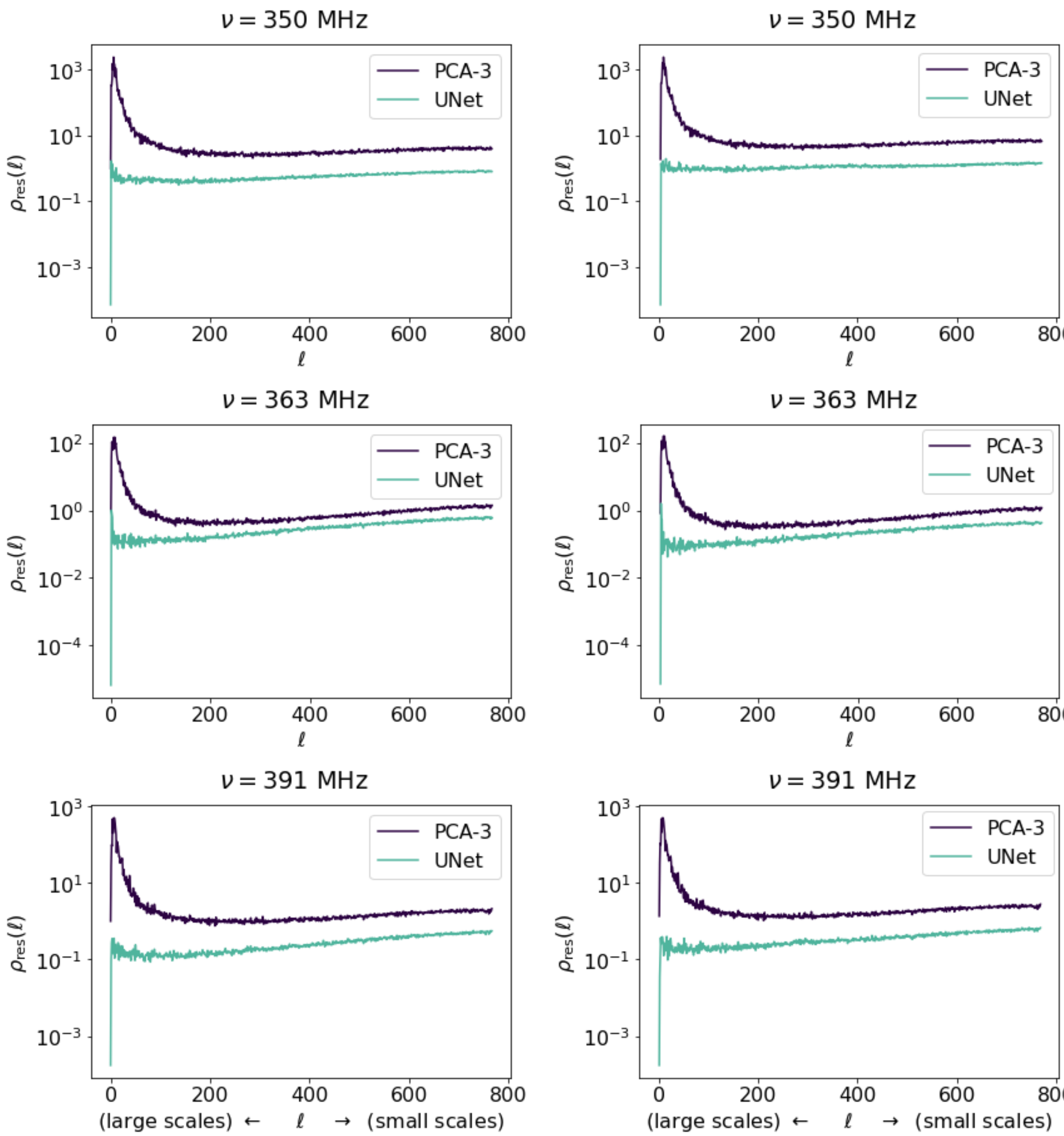

Figure 7.13: Ratios of the power spectrum of the residual maps $p - t$ over the power spectrum of the true map t (Eq 7.19) for pure $\log\cosh$ loss (left) and custom loss (right) at three different frequencies. The ratios for PCA pre-cleaned maps are shown in purple, and in teal for UNet reconstructed maps.

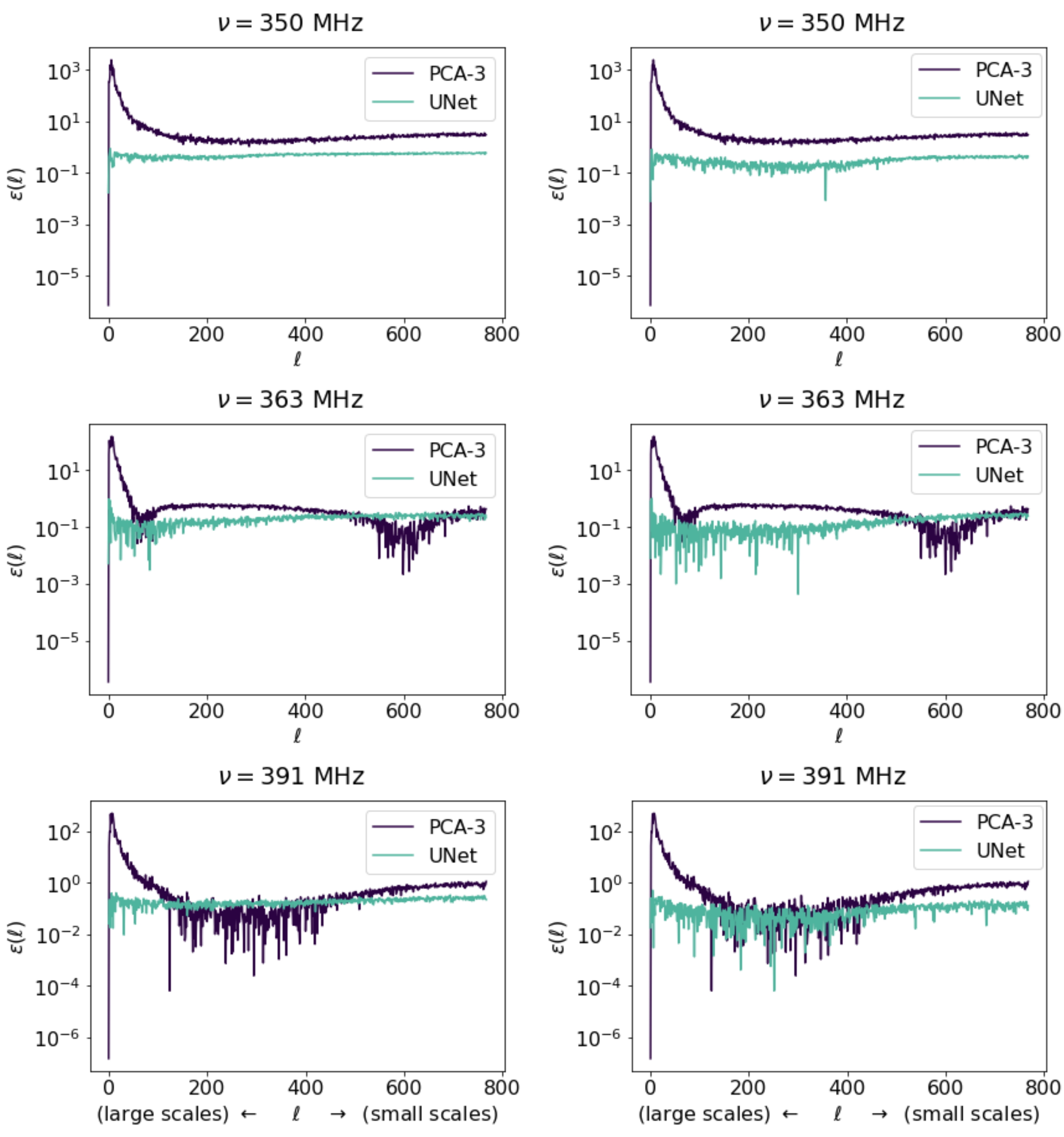

Figure 7.14: Bias introduced by PCA and UNet (Eq 7.19) for pure log cosh loss (left) and custom loss (right) at three different frequencies. The ratios for PCA pre-cleaned maps are shown in purple, and in teal for UNet reconstructed maps.

7.4 Future work

Since intensity mapping is in its infancy stage, there has been no undisputed detected HI signal from any standalone observation. In the immediate future, we need to look at cross-correlation with galaxy redshift surveys to separate the cosmological HI signal from the continuum foreground emission.

This chapter presented preliminary results on adding galaxy survey cross-correlation information to an existing method using a UNet convolutional neural network. We see that galaxy cross-correlation offers minor improvement on the bias of the power spectrum at certain frequencies. However, more work needs to be done to confirm the benefit of galaxy cross-correlation. Given more time, I would run the model on the full data set with 100 skies or more. In addition, I would also need to build confidence intervals for the network's prediction by training an ensemble of UNets, i.e. training a number of different models and building a statistic on the predicted outputs.

There are a few challenges in doing so: The `CRIME` simulations are very resource intensive and take a lot of time to complete, even on the computing cluster at the Center for High Throughput Computing (CHTC) at UW-Madison. The full UNet model with 100 skies requires two NVIDIA A100 GPUs with 80GB of dedicated graphics memory, which is half the capacity of CHTC. Running the training and hyperparameter tuning on the Google Cloud Compute Engine is expensive.

The `CRIME` simulation only generates synthetic maps observed by a single dish telescope. In the future, we will extend this to simulations of interferometers. This is more complicated due to the frequency-dependent beam effects, which will generate false structures in the frequency direction. Li et al. [80] have experimented with separating the EoR signal with a denoising autoencoder on a simulated SKA dataset.

In summary, the biggest advantage of using neural networks in cleaning the foreground signal is that we do not need an exact analytical description of the various physical foreground components. However, there are a few disadvantages that come with it. The intermediate steps inside the neural network are not generally interpretable. It requires a lot of simulated data and is computationally costly. It might also fail to clean skies where the cosmological parameters or the foreground physics are different from those in the training data. However, given the rapid development and application of deep learning techniques, we are positive that these challenges can be overcome and machine learning will be a standard tool in the astrophysicists' toolbox in the future.

Chapter 8

Conclusion

8.1 Summary

The goal of observational cosmology is to better understand the Universe, its origin, history, content, geometry, and its evolution. There are many approaches to accomplish that goal. We can get a lot of information about the early Universe by studying the CMB anisotropy. It offers a window through which we can test different inflationary models, the content of the Universe, and the density perturbations that grow to become galaxies millions of years later. The BAO peaks can provide information on different cosmological parameters in the ΛCDM model such as the topology of the Universe, the Hubble parameter, and the dark matter and dark energy fractions. We talked about how Boltzmann codes such as `CMBAns` allow us to infer the values of different cosmological parameters in the ΛCDM model from observed power spectra.

However, there are components in ΛCDM that we do not yet understand. One notable example is the nature of dark matter and its distribution. There are many current and planned experiments to study dark matter that take advantage of the entire electromagnetic radiation spectrum. Using galaxies as tracers of the underlying dark matter intensity field, galaxy surveys, both photometric and spectroscopic, have provided us with a wealth of information at low redshifts in recent years. However, there are other tracers that could be even more powerful than galaxy surveys. One such tracer is neutral hydrogen and the associated 21 cm emission intensity mapping. In principle, intensity mapping can complement CMB observation by providing gravitational potential maps that could be used to de-lense CMB B-mode in the hope of searching for inflationary B-mode signal.

With 21 cm intensity mapping, we could in principle map much larger volumes in a shorter time compared to other approaches. However, intensity mapping is still a relatively young field and there are many technical issues to be addressed before we can unlock its full power. The main challenge is that the foregrounds are several orders of magnitude brighter than the 21 cm signal. Fortunately, the foregrounds are smooth in frequency space, while the 21 cm signal contains spectral structures. This allows the foregrounds to be, in theory, removed so that we can analyze the 21 cm signal. While waiting for observations to catch up, we also need to improve the theoretical models for HI, as well as foreground removal techniques to extract the most out of the HI signal.

In this thesis, we have:

- Introduced a modular Boltzmann code to predict the CMB power spectrum from cosmological parameters.
- Evaluated the performance of the Tianlai transit radio interferometer by doing calibration, characterizing the beam, and quantifying gain stability.
- Developed a new method to remove solar contamination from radio interferometric data.

- Studied foreground cleaning for future intensity mapping with machine learning and cross-correlation with galaxy surveys.

Chapter 2 offered a general introduction into the basic concepts in cosmology, which served to give a better understanding of the succeeding chapters. We briefly described the homogeneous Universe, its composition and the evolution of its components. We described how small imhomogeneities started by quantum fluctuations in the early Universe later grew to form anisotropies in the CMB.

Chapter 3 introduced a cosmological Boltzmann code named `CMBAns`. It is capable of calculating the CMB angular power spectra for different cosmological parameters up to high multipoles. We focus on the perturbations in a flat-Universe. It is written in a modular format that users can use to write their codes.

Chapter 4 is a brief introduction into the physics behind intensity mapping. We described how perturbation in the matter density lead to structure formation. We introduced the correlation function between the over-densities and how it is used to calculate the matter power spectrum. We showed how the expected 21 cm signal depends on the total amount of HI in the Universe and the HI bias. Finally, we can use the HI halo model to make prediction on the HI bias and the shot noise.

In Chapter 5, we described in detail the fundamentals of radio interferometry. Interferometric arrays are increasingly popular for intensity mapping, since they eliminate the need for a single, very large aperture telescope. The Tianlai Dish Array is one such array. It is optimized to perform cosmological studies with 21 cm intensity mapping in the redshift range $2.55 \geq z \geq -0.01$. Since it is still in the early stage, we described work on calibrating the instrument, ensuring the stability of the absolute gain and electronic phase drift. We also introduced the software pipeline that processes and analyzes the data.

Chapter 6 introduced a novel technique for solar contamination removal in any interferometric array, given that the Sun signal is the dominant source in the sky. We showed that the algorithm leaves most other sky signals intact. This enables us to use the daytime data for cosmological analysis, which would otherwise be unusable.

In Chapter 7, we described the different types of foregrounds encountered with intensity mapping observations. From the `CRIME` simulations of foregrounds and HI, we built simple mock galaxy maps. We showed that we can use machine learning techniques to clean the foreground by cross-correlating with galaxy surveys.

8.2 Future work

There are many open questions in cosmology that still need to be addressed, and the CMB and intensity mapping are powerful, complementary tools to study the evolution of the Universe from the surface of last scattering until the present time.

Current CMB experiments such as Planck, ACT, and SPT have measured the CMB data up to very high multipoles. All these cosmological data can be very well fitted with a flat ΛCDM model. However, several recent papers also claim slight deviations from a flat Universe. Therefore, it is desirable for the cosmological Boltzmann codes to be able to calculate the power spectrum for non-flat models, so that even any small deviation from the special flatness can be measured. Currently, the best fit value of Ω_k from Planck TT, TE, EE+lowE+lensing is $\Omega_k = -0.011^{+0.013}_{-0.012}$. When we add BAO measurements from galaxy redshift surveys to this the best fit value comes down to $\Omega_k = 0.0007^{+0.0037}_{-0.0037}$ [6]. The curvature mainly changes the CMB power spectrum by moving the power-spectrum horizontally. This is due to the change in the distance of the last scattering surface. In the near future, we will extend `CMBAns` for the case of a non-flat background metric. In addition, we will address cases in which fundamental constants vary as a function of redshift. Fundamental constants have played an important role in physics. They are either directly connected to the strength of different interactions that happens in nature, such as G, and there are theories in which the fundamental constants may actually be spacetime variables. So far only `CMBAns` is capable of calculating the CMB power spectrum for models with varying fundamental constants. Most of the code for this case is already written, and we will document this process in a future paper.

Even though the information from the CMB is important in its own right, observations during the dark ages and reionization era are crucial for filling the knowledge gap on the formation of structures in the Universe. The neutral hydrogen 21 cm transition is the only signal that can be used to probe this era, since the Universe then contains no bright sources. The measurement of 21 cm temperature maps as a function of redshift will allow us to get information on the evolution of the Universe during the dark ages and EoR. Several dedicated instruments have been built over the last two decades, but no standalone observation of the HI signal has been made. The Tianlai array is one such instrument, designed to cover a broad frequency range (400-1430 MHz). A few thousand hours of observation, mostly toward the NCP, have been performed. Progress in instrument design, electronics, and data processing are being made to reach a higher sensitivity. More performance assessments and better calibration are needed. It will be necessary to determine the beam response of individual antennas through a combination of electromagnetic simulations and on-site beam measurements. Frequency structures in the receiver, such as standing waves, should be minimized and carefully filtered out. The level of cross coupling between antennas needs to be reduced. The contribution of bright sources such as the sun through the antenna far sidelobes, which is frequency dependent, needs to be better understood and modelled. The Tianlai project also serves as a stepping stone for other intensity mapping experiments, such as HIRAX and SKA.

In all intensity mapping experiments, the 21 cm is totally overwhelmed by foreground emissions, which are generally a few thousand times brighter than the cosmological HI signal. Separating the 21 cm signal from the foregrounds is an ongoing tasks. In the past decade, progress in machine learning have been significant. We have shown that it is possible to clean foregrounds on a simulated data set with galaxy cross-correlation. This work is still inconclusive, and we would like to run it on a larger data set and with more computing power. We would also like to extend the model to include polarized foregrounds or varying cosmological parameters. In the future, we can use pre-trained models to perform transfer learning to real data and cross-check with observed power spectra.

Appendix A

The fluid equation

In this section, we will derive the relationship between pressure and density for cosmological fluids such as matter and radiation. First, consider an expanding volume of comoving radius. The volume has radius a, so the rate of change in volume is

$$\frac{\mathrm{d}V}{\mathrm{d}t} = 4\pi a^2 \frac{\mathrm{d}a}{\mathrm{d}t} \tag{A.1}$$

The total energy confined within the volume is

$$E = \left(\frac{4}{3}\pi a^3 \rho\right) c^2 \tag{A.2}$$

The rate of change in the energy with respect to time is

$$\frac{\mathrm{d}E}{\mathrm{d}t} = 4\pi a^2 \rho c^2 \frac{\mathrm{d}a}{\mathrm{d}t} + \frac{4}{3}\pi a^3 c^2 \frac{\mathrm{d}\rho}{\mathrm{d}t} \tag{A.3}$$

Apply the first law of thermodynamics to the above equations

$$\mathrm{d}E + p\mathrm{d}V = T\mathrm{d}S \tag{A.4}$$

and assuming reversible adiabatic expansion $\mathrm{d}S = 0$, we get

$$\boxed{\dot{\rho} + 3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) = 0} \tag{A.5}$$

This is the **fluid equation**. In unit in which $c = 1$, the fluid equation becomes

$$\dot{\rho} + 3\frac{\dot{a}}{a}\left(\rho + p\right) = 0 \tag{A.6}$$

We can assume that there is a unique pressure associated with the corresponding density, so that $p = wp$ where w is a constant as described in (2.28).

- **Matter**: We can solve the fluid equation for matter, which has $p_m = 0$:

$$\dot{\rho}_m + 3\frac{\dot{a}}{a}\rho_m = 0 \quad \Rightarrow \quad \frac{1}{a^3}\frac{\mathrm{d}}{\mathrm{d}t}(\rho_m a^3) = 0 \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}(\rho_m a^3) = 0 \quad \Rightarrow \quad \rho_m \propto a^{-3} \tag{A.7}$$

- **Radiation**: For isotropic radiation, $p_r = \rho_r c^2/3$ ($= \rho_r/3$ in unit in which $c = 1$). The fluid equation becomes

$$\dot{\rho}_r + 4\frac{\dot{a}}{a}\rho_r = 0 \quad \Rightarrow \quad \frac{1}{a^4}\frac{\mathrm{d}}{\mathrm{d}t}(\rho_r a^4) = 0 \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}(\rho_r a^4) = 0 \quad \Rightarrow \quad \rho_r \propto a^{-4} \tag{A.8}$$

- **Vacuum**: For vacuum, the energy density ρ_Λ is constant. It is not dependent on the expansion of the universe, so $\rho_\Lambda \propto a^0$

Appendix B

Cosmic variance

The cosmic variance is define as

$$\frac{\Delta C_\ell}{C_\ell} \equiv \frac{\sqrt{\langle (C_\ell - \hat{C}_\ell)^2 \rangle}}{C_\ell} \tag{B.1}$$

in which

$$\langle (C_\ell - \hat{C}_\ell)^2 \rangle = C_\ell^2 - 2C_\ell \langle \hat{C}_\ell \rangle + \langle \hat{C}_\ell^2 \rangle = -C_\ell^2 + \langle \hat{C}_\ell^2 \rangle \tag{B.2}$$

and

$$\begin{aligned}
\langle \hat{C}_\ell^2 \rangle &= \frac{1}{(2\ell+1)^2} \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a_{\ell(-m)} a_{\ell m'} a_{\ell(-m')} \rangle \\
&= \frac{1}{(2\ell+1)^2} \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a_{\ell(-m)} a_{\ell m'} a_{\ell(-m')} \rangle \\
&= \frac{1}{(2\ell+1)^2} \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \Big(\langle a_{\ell m} a_{\ell(-m)} \rangle \langle a_{\ell m'} a_{\ell(-m')} \rangle \\
&+ \langle a_{\ell m} a_{\ell m'} \rangle \langle a_{\ell(-m)} a_{\ell(-m')} \rangle + \langle a_{\ell m} a_{\ell(-m')} \rangle \langle a_{\ell(-m)} a_{\ell -m'} \rangle \Big)
\end{aligned} \tag{B.3}$$

where we have used the Isserlis' theorem in statistics to get to the last line. In addition,

$$\sum_{m=-\ell}^{\ell}\sum_{m'=-\ell}^{\ell}\langle a_{\ell m}a_{\ell(-m)}\rangle\langle a_{\ell m'}a_{\ell(-m')}\rangle = (2\ell+1)^2 C_\ell^2 \tag{B.4}$$

$$\begin{aligned}\sum_{m=-\ell}^{\ell}\sum_{m'=-\ell}^{\ell}\langle a_{\ell m}a_{\ell m'}\rangle\langle a_{\ell(-m)}a_{\ell(-m')}\rangle &= \sum_{m=-\ell}^{\ell}\langle a_{\ell m}a_{\ell(-m)}\rangle\langle a_{\ell(-m)}a_{\ell m}\rangle \\ &= (2\ell+1)C_\ell^2 \end{aligned}\tag{B.5}$$

Similarly,

$$\sum_{m=-\ell}^{\ell}\sum_{m'=-\ell}^{\ell}\langle a_{\ell m}a_{\ell(-m')}\rangle\langle a_{\ell(-m)}a_{\ell -m'}\rangle = (2\ell+1)C_\ell^2 \tag{B.6}$$

So that

$$\langle \hat{C}_\ell^2\rangle = \frac{C_\ell^2[(2\ell+1)^2+2(2\ell+1)]}{(2\ell+1)^2} = C_\ell^2\left(1+\frac{2}{2\ell+1}\right) \tag{B.7}$$

and

$$\langle (C_\ell-\hat{C}_\ell)^2\rangle = -C_\ell^2 + C_\ell^2\left(1+\frac{2}{2\ell+1}\right) = \frac{2C_\ell^2}{2\ell+1} \tag{B.8}$$

The cosmic variance is therefore,

$$\frac{\Delta C_\ell}{C_\ell} \equiv \frac{\sqrt{\langle (C_\ell-\hat{C}_\ell)^2\rangle}}{C_\ell} = \sqrt{\frac{2}{2\ell+1}} \tag{B.9}$$

Appendix C

`CMBAns` Modularity Example

This appendix shows the modularity functionality of `CMBAns`. Users can simply call the `CMBAns` function by calling `ModuleName.function`. A module in this context is a class in `C`.

C.0.1 calconftime

This function resides inside module `Others`, which contains different standard cosmological functions that are frequently used in different calculations, such as the conformal time at different redshift, distance to the sound horizon, etc. `calconftime` will calculate the conformal time difference between any scale factor to the present era by integrating $\mathrm{d}\tau/\mathrm{d}a$ from scale factor a to the present time. We use the function `rombint` for integrating the function.

$$\tau = \int_a^1 \frac{\mathrm{d}\tau}{\mathrm{d}a}\,\mathrm{d}a \tag{C.1}$$

Format

`double calcconftime(double a)`

Arguments

Name	Kind	In/Out	Description
`a`	`double`	IN	the scale factor

Examples

```
#include<stdio.h>
#include<math.h>
#include "variables.h"
#include "others.h"
#include "neutrino.h"

int main() {
    CMB cmb;
    others Others;
    neutrino Neutrino;

    cmb.OmegaB = 0.05;
    cmb.OmegaC = 0.25;
    cmb.OmegaDE = 0.7;
    cmb.OmegaNmassive = 0.0;
    cmb.H0 = 67.9;
    cmb.Tcmb = 2.7254;
    cmb.nNeutrinoMassive = 0.0;

    cmb.setparam();

    printf("%e",Others.calcconftime(0.5));
    }
```

C.0.2 ionize

Located inside `recombination` module, this function calculates the ionization fraction of hydrogen at a conformal time $\tau + d\tau$ using Peebles equation. The hydrogen ionization fraction is defined as $x_H = n_e/n_H$, where n_e is the number of free electron density and n_H is the total hydrogen number density. For calculating the hydrogen ionization fraction at any given τ, we can set $x_H = 1$ as initial condition at a very high redshift (at $\tau = \tau_0$) and then calculate x_H at any given $\tau = \tau_0 + n\mathrm{d}\tau$ by repeatedly calling this function.

Format

```
double ionize(double tempb,double a,double adot,double dtau,double xe)}
```

Arguments

Name	Kind	In/Out	Description
`tempb`	`double`	IN	the baryon temperature
`a`	`double`	IN	the scale factor
`adot`	`double`	IN	conformal time derivative of the scale factor
`dtau`	`double`	IN	time increment $\mathrm{d}\tau$
`xe`	`double`	IN	hydrogen ionization fraction at τ
		OUT	hydrogen ionization fraction at $\tau + \mathrm{d}\tau$

Examples

```
#include<stdio.h>
#include<math.h>
#include "recombination.h"
#include "variables.h"
```

```
#include "others.h"

int main() {
CMB cmb;
others Others;
recombination Recombination;
cmb.setparam();

// print the reionization fraction at redshift z = 1370 (a = 1/(z+1))
double adot = 1/Others.dtauda(.00072939);
printf("reionization fraction x_e = %f",
                 Recombination.ionize(3740, .00072939, adot, 1, 0.5));
}
```

C.0.3 `void nu1`

This function is located inside `Neutrino` module. For massive neutrinos, its not possible to calculate the pressure and density by direct analytical integration. Therefore, we need to calculate them numerically. The `Neutrino` module contains functions related to neutrino pressure and density calculations. Specifically, the function `void nu1` can be used to calculate massive neutrino density and pressure at a given scale factor a. It calculates these quantities by interpolating the arrays created by `initnul()`. Therefore, it requires `initnul()` to be called before calling `nu1()`for the first time. It calculates the neutrino density and pressure in the dimensionless form.

Format

`void nu1(double a,double rrnu[])`

Arguments

Name	Kind	In/Out	Description
`a`	`double`	IN	The scale factor
`rrnu[]`	`double`	OUT	Output containing the density (first component) and pressure (second component) of the massive neutrinos

Example

```
#include<stdio.h>
#include<math.h>
#include "variables.h"
#include "neutrino.h"

int main() {
neutrino Neutrino;
CMB cmb;
```

```
    double rhonu[2];
double rrnu[2];

cmb.OmegaB = 0.05;
cmb.OmegaC = 0.25;
cmb.OmegaDE = 0.7-0.0054;
// mass of non relativistic neutrinos
cmb.OmegaNmassive = 0.0054;
cmb.H0 = 67.9;
cmb.Tcmb = 2.7254;
cmb.nNeutrinoMassive = 3.0;

    cmb.setparam();

    Neutrino.initnul();

    // interpolate the neutrino density and pressure
Neutrino.nu1(0.005, rrnu);
    // calculating the exact neutrino density and pressure
Neutrino.ninul(0.005, rhonu);

    // print the result of the interpolation
printf("%lf %lf\n", rrnu[0], rrnu[1]);
// print the exact result
printf("%lf %lf\n", rhonu[0], rhonu[1]); }
```

C.0.4 rombint

Module `Numericx` contains the functions for different numerical techniques used in `CMBAns` such as numerical differentiation, integration, spline interpolation, set of linear differential equation solver, etc. `rombint` is a function inside the module `Numericx`. It takes in a function $f(x)$ as input and numerically integrates it from l to u using Romberg's method:

$$\int_l^u f(x)\,\mathrm{d}x.$$

Format

`double rombint(double (*func)(double), double l, double u, double tol)`

Arguments

Name	Kind	In/Out	Description
`*func`	`double (*func)(double)`	IN	function to be integrated
`u`	`double`	IN	upperbound of the integral
`l`	`double`	IN	lowerbound of the integral
`tol`	`double`	IN	numerical tolerance

Example

To integrate $f(x) = \sin(x)$ from 0 to 1, subjecting to a numerical tolerance of $\texttt{tol} = 10^{-8}$, we can use the following code

```
#include<stdio.h>
#include<math.h>
#include "numericx.h"

double f(double x) {
      return sin(x);
}

int main() {
      numericx Numericx;
      double tol = 1e-8;
      double ans = Numericx.rombint(f, 0, 3.14159265359, tol);
      printf("%e",ans);
}
```

Output : `2.000000000e+000`

References

[1] Kevork N Abazajian, Peter Adshead, Zeeshan Ahmed, Steven W Allen, David Alonso, Kam S Arnold, Carlo Baccigalupi, James G Bartlett, Nicholas Battaglia, Bradford A Benson, et al. Cmb-s4 science book. *arXiv preprint arXiv:1610.02743*, 2016.

[2] F. B. Abdalla and S. Rawlings. Probing dark energy with baryonic oscillations and future radio surveys of neutral hydrogen. *MNRAS*, 360:27–40, June 2005.

[3] PAR Ade, N Aghanim, C Armitage-Caplan, M Arnaud, M Ashdown, F Atrio-Barandela, J Aumont, C Baccigalupi, Anthony J Banday, RB Barreiro, et al. Planck 2013 results. xvi. cosmological parameters. *Astronomy & Astrophysics*, 571:A16, 2014.

[4] Peter Ade, James Aguirre, Zeeshan Ahmed, Simone Aiola, Aamir Ali, David Alonso, Marcelo A Alvarez, Kam Arnold, Peter Ashton, Jason Austermann, et al. The simons observatory: science goals and forecasts. *Journal of Cosmology and Astroparticle Physics*, 2019(02):056, 2019.

[5] Peter AR Ade, N Aghanim, M Arnaud, Mark Ashdown, J Aumont, C Baccigalupi, AJ Banday, RB Barreiro, JG Bartlett, N Bartolo, et al. Planck

2015 results-xiii. cosmological parameters. *Astronomy & Astrophysics*, 594:A13, 2016.

[6] Nabila Aghanim, Yashar Akrami, Frederico Arroja, Mark Ashdown, J Aumont, Carlo Baccigalupi, M Ballardini, Anthony J Banday, RB Barreiro, Nicola Bartolo, et al. Planck 2018 results-i. overview and the cosmological legacy of planck. *Astronomy & Astrophysics*, 641:A1, 2020.

[7] Yacine Ali-Haïmoud and Christopher M. Hirata. Hyrec: A fast and highly accurate primordial hydrogen and helium recombination code. 2010.

[8] Yacine Ali-Haïmoud and Christopher M. Hirata. Ultrafast effective multi-level atom method for primordial hydrogen recombination. 2010.

[9] David Alonso, Philip Bull, Pedro G Ferreira, and Mário G Santos. Blind foreground subtraction for intensity mapping experiments. *Monthly Notices of the Royal Astronomical Society*, 447(1):400–416, 2015.

[10] David Alonso, Pedro G Ferreira, and Mario G Santos. Fast simulations for intensity mapping experiments. *Monthly Notices of the Royal Astronomical Society*, 444(4):3183–3197, 2014.

[11] Ralph A Alpher, Hans Bethe, and George Gamow. The origin of chemical elements. *Physical Review*, 73(7):803, 1948.

[12] C. J. Anderson, 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. Low-amplitude clustering in low-redshift 21-cm intensity maps cross-correlated with 2dF galaxy densities. , 476(3):3382–3392, May 2018.

[13] CJ Anderson, NJ Luciw, Y-C Li, CY Kuo, J Yadav, KW Masui, TC Chang, X Chen, N Oppermann, YW Liao, et al. 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, 2018.

[14] Réza Ansari. Cosmic visions dark energy: Inflation and early dark energy with a stage ii, hydrogen intensity mapping experiment. *arXiv preprint arXiv:1810.09572*, 2018.

[15] Réza Ansari, Evan J Arena, Kevin Bandura, Philip Bull, Emanuele Castorina, Tzu-Ching Chang, Shi-Fan Chen, Liam Connor, Simon Foreman, Josef Frisch, et al. Inflation and early dark energy with a stage ii hydrogen intensity mapping experiment. *arXiv preprint arXiv:1810.09572*, 2018.

[16] C. Armendariz-Picon, V. Mukhanov, and Paul J. Steinhardt. Dynamical solution to the problem of a small cosmological constant and late-time cosmic acceleration. *Phys. Rev. Lett.*, 85:4438–4441, Nov 2000.

[17] C. Armendariz-Picon, V. Mukhanov, and Paul J. Steinhardt. Essentials of k-essence. *Phys. Rev. D*, 63:103510, Apr 2001.

[18] J.S. Bagla, H. K. Jassal, and T. Padmanabhan. Cosmology with tachyon field as dark energy. *Phys.Rev.*, D67:063504, 2003.

[19] Kevin Bandura, Graeme E Addison, Mandana Amiri, J Richard Bond, Duncan Campbell-Wilson, Liam Connor, Jean-François Cliche, Greg Davis, Meiling Deng, Nolan Denman, et al. Canadian hydrogen intensity mapping experiment (chime) pathfinder. In *Ground-based and Airborne Telescopes V*, volume 9145, page 914522. International Society for Optics and Photonics, 2014.

[20] James M. Bardeen. Gauge-invariant cosmological perturbations. *Phys. Rev. D*, 22:1882–1905, Oct 1980.

[21] RA Battye, ML Brown, IWA Browne, RJ Davis, P Dewdney, C Dickinson, G Heron, B Maffei, A Pourtsidou, and PN Wilkinson. Bingo: a single dish approach to 21cm intensity mapping. *arXiv preprint arXiv:1209.1041*, 2012.

[22] RA Battye, IWA Browne, C Dickinson, G Heron, B Maffei, and A Pourtsidou. H i intensity mapping: a single dish approach. *Monthly Notices of the Royal Astronomical Society*, 434(2):1239–1256, 2013.

[23] Richard A Battye, Rod D Davies, and Jochen Weller. Neutral hydrogen surveys for high-redshift galaxy clusters and protoclusters. *Monthly Notices of the Royal Astronomical Society*, 355(4):1339–1347, 2004.

[24] Diego Blas, Julien Lesgourgues, and Thomas Tram. The cosmic linear anisotropy solving system (class). part ii: approximation schemes. *Journal of Cosmology and Astroparticle Physics*, 2011(07):034, 2011.

[25] C. G. Boehmer and T. Harko. Dark energy as a massive vector field. *Eur. Phys. J.*, C50:423–429, 2007.

[26] F. H. Briggs, J. F. Bell, and M. J. Kesteven. Removing radio interference from contaminated astronomical spectra using an independent reference signal and closure relations. *The Astronomical Journal*, 120(6):3351–3361, dec 2000.

[27] Philip Bull, Pedro G Ferreira, Prina Patel, and Mario G Santos. Late-time cosmology with 21 cm intensity mapping experiments. *The Astrophysical Journal*, 803(1):21, 2015.

[28] Emanuele Castorina and Francisco Villaescusa-Navarro. 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, 2017.

[29] Tzu-Ching Chang, Ue-Li Pen, Kevin Bandura, and Jeffrey B Peterson. An intensity map of hydrogen 21-cm emission at redshift z [thinsp][ap][thinsp] 0.8. *Nature*, 466(7305):463–465, 2010.

[30] Tzu-Ching Chang, Ue-Li Pen, Jeffery B. Peterson, and Patrick McDonald. Baryon acoustic oscillation intensity mapping of dark energy. *Phys. Rev. Lett.*, 100(091303), 2008.

[31] Xuelei Chen. Radio detection of dark energy—the tianlai project. *Scientia Sinica Physica, Mechanica & Astronomica*, 41(12):1358–1366, 2011.

[32] Xuelei Chen. The tianlai project: a 21cm cosmology experiment. In *International Journal of Modern Physics: Conference Series*, volume 12, pages 256–263. World Scientific, 2012.

[33] Takeshi Chiba, Takahiro Okabe, and Masahide Yamaguchi. Kinetically driven quintessence. *Phys. Rev. D*, 62:023511, Jun 2000.

[34] CHIME Collaboration, Mandana Amiri, 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. Detection of Cosmological 21 cm Emission with the Canadian Hydrogen Intensity Mapping Experiment. *arXiv e-prints*, page arXiv:2202.01242, February 2022.

[35] J. Chluba and R. M. Thomas. Towards a complete treatment of the cosmological recombination problem. 2010.

[36] J. Chluba, G. M. Vasil, and L. J. Dursi. Recombinations to the rydberg states of hydrogen and their effect during the cosmological recombination epoch. 2010.

[37] Santanu Das. Iucaa graduate school report, 2010.

[38] Santanu Das, Christopher J Anderson, Reza Ansari, Jean-Eric Campagne, Daniel Charlet, Xuelei Chen, Zhiping Chen, Aleksander J Cianciara, Pierre Colom, Yanping Cong, et al. Progress in the construction and testing of the tianlai radio interferometers. In *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy IX*, volume 10708, page 1070836. International Society for Optics and Photonics, 2018.

[39] Santanu Das and Anh Phan. Cosmic microwave background anisotropy numerical solution (cmbans). part i. an introduction to cl calculation. *Journal of Cosmology and Astroparticle Physics*, 2020(05):006, 2020.

[40] Santanu Das, Arman Shafieloo, and Tarun Souradeep. ISW effect as probe of features in the expansion history of the Universe. *JCAP*, 1310:016, 2013.

[41] Santanu Das and Tarun Souradeep. SCoPE: An efficient method of Cosmological Parameter Estimation. *JCAP*, 1407:018, 2014.

[42] Santanu Das and Tarun Souradeep. Suppressing CMB low multipoles with ISW effect. *JCAP*, 1402:002, 2014.

[43] Marc Davis, George Efstathiou, Carlos S Frenk, and Simon DM White. The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal*, 292:371–394, 1985.

[44] David R DeBoer, 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, et al. Hydrogen epoch of reionization array (hera). *Publications of the Astronomical Society of the Pacific*, 129(974):045001, 2017.

[45] Robert H Dicke, P James E Peebles, Peter G Roll, and David T Wilkinson. Cosmic black-body radiation. *The Astrophysical Journal*, 142:414–419, 1965.

[46] Clive Dickinson. Bingo-a novel method to detect baos using a total-power radio telescope. *arXiv preprint arXiv:1405.7936*, 2014.

[47] Scott Dodelson. *Modern cosmology*. Academic press, 2003.

[48] A. D. Dolgov, S. H. Hansen, and D. V. Semikoz. Nonequilibrium corrections to the spectra of massless neutrinos in the early universe. 1997.

[49] A. D. Dolgov, S. H. Hansen, and D. V. Semikoz. Nonequilibrium corrections to the spectra of massless neutrinos in the early universe - addendum. 1998.

[50] Michael Doran. CMBEASY: an object oriented code for the cosmic microwave background. *JCAP*, 0510:011, 2005.

[51] Michael Doran and Christian M. Mueller. Analyze This! A Cosmological constraint package for CMBEASY. *JCAP*, 0409:003, 2004.

[52] Michael W Eastwood, Marin M Anderson, Ryan M Monroe, Gregg Hallinan, Benjamin R Barsdell, Stephen A Bourke, MA Clark, Steven W Ellingson, Jayce Dowell, Hugh Garsden, et al. The radio sky at meter wavelengths: m-mode analysis imaging with the ovro-lwa. *The Astronomical Journal*, 156(1):32, 2018.

[53] Daniel J Eisenstein, David H Weinberg, Eric Agol, Hiroaki Aihara, Carlos Allende Prieto, Scott F Anderson, James A Arns, Éric Aubourg, Stephen Bailey, Eduardo Balbinot, et al. Sdss-iii: Massive spectroscopic surveys of the distant universe, the milky way, and extra-solar planetary systems. *The Astronomical Journal*, 142(3):72, 2011.

[54] Miguel Escudero. Neutrino decoupling beyond the standard model: Cmb constraints on the dark matter mass with a fast and precise n_{eff} evaluation. 2018.

[55] DJ Fixsen, ES Cheng, JM Gales, John C Mather, RA Shafer, and EL Wright. The cosmic microwave background spectrum from the full cobe firas data set. *The Astrophysical Journal*, 473(2):576, 1996.

[56] Riccardo Giovanelli. Alfalfa: an exploration of the z= 0 hi universe. *arXiv preprint arXiv:0802.2100*, 2008.

[57] Daniel Grin and Christopher M. Hirata. Cosmological hydrogen recombination: The effect of extremely high-n states. *Phys. Rev. D*, 81:083005, Apr 2010.

[58] E. Grohs, G. M. Fuller, C. T. Kishimoto, M. W. Paris, and A. Vlasenko. Neutrino energy transport in weak decoupling and big bang nucleosynthesis. 2015.

[59] AJS Hamilton. Linear redshift distortions: A review. *The evolving universe*, pages 185–275, 1998.

[60] Wayne Hu. Covariant linear perturbation formalism. 2004.

[61] Wayne Hu and Scott Dodelson. Cosmic microwave background anisotropies. *Annual Review of Astronomy and Astrophysics*, 40(1):171–216, 2002.

[62] Wayne Hu and Takemi Okamoto. Mass reconstruction with cosmic microwave background polarization. *The Astrophysical Journal*, 574(2):566, 2002.

[63] Wayne Hu and Martin White. A cmb polarization primer. *arXiv preprint astro-ph/9706147*, 1997.

[64] Wenkai Hu, Xin Wang, Fengquan Wu, Yougang Wang, Pengjie Zhang, and Xuelei Chen. Forecast for fast: from galaxies survey to intensity mapping. *Monthly Notices of the Royal Astronomical Society*, 493(4):5854–5870, Mar 2020.

[65] Edwin Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, 1929.

[66] Nick Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227(1):1–21, 1987.

[67] Alexander Yu. Kamenshchik, Ugo Moschella, and Vincent Pasquier. An Alternative to quintessence. *Phys. Lett.*, B511:265–268, 2001.

[68] Marc Kamionkowski, Arthur Kosowsky, and Albert Stebbins. Statistics of cosmic microwave background polarization. *Physical Review D*, 55(12):7368, 1997.

[69] Hideo Kodama and Misao Sasaki. Cosmological Perturbation Theory. *Prog.Theor.Phys.Suppl.*, 78:1–166, 1984.

[70] Tomi Koivisto and David F. Mota. Vector Field Models of Inflation and Dark Energy. *JCAP*, 0808:021, 2008.

[71] Julien Lesgourgues. The cosmic linear anisotropy solving system (class) i: Overview. *arXiv preprint arXiv:1104.2932*, 2011.

[72] Julien Lesgourgues. The cosmic linear anisotropy solving system (class) iii: Comparision with camb for lambdacdm. *arXiv preprint arXiv:1104.2934*, 2011.

[73] Julien Lesgourgues and Thomas Tram. The cosmic linear anisotropy solving system (class) iv: efficient implementation of non-cold relics. *Journal of Cosmology and Astroparticle Physics*, 2011(09):032, 2011.

[74] Antony Lewis. Camb. http://camb.info/readme.html.

[75] Antony Lewis. Linear effects of perturbed recombination. *Physical Review D*, 76(6):063001, 2007.

[76] Antony Lewis. Efficient sampling of fast and slow cosmological parameters. *Phys.Rev.*, D87(10):103529, 2013.

[77] Antony Lewis and Sarah Bridle. Cosmological parameters from CMB and other data: A Monte Carlo approach. *Phys.Rev.*, D66:103511, 2002.

[78] Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of cosmic microwave background anisotropies in closed friedmann-robertson-walker models. *The Astrophysical Journal*, 538(2):473, 2000.

[79] JiXia Li, ShiFan Zuo, FengQuan Wu, YouGang Wang, JuYong Zhang, ShiJie Sun, YiDong Xu, ZiJie Yu, Reza Ansari, YiChao Li, Albert Stebbins, Peter Timbie, YanPing Cong, JingChao Geng, Jie Hao, QiZhi Huang, JianBin Li, Rui

Li, DongHao Liu, YingFeng Liu, Tao Liu, John P. Marriner, ChenHui Niu, Ue-Li Pen, Jeffery B. Peterson, HuLi Shi, Lin Shu, YaFang Song, HaiJun Tian, GuiSong Wang, QunXiong Wang, RongLi Wang, WeiXia Wang, Xin Wang, KaiFeng Yu, Jiao Zhang, BoQin Zhu, JiaLu Zhu, and XueLei Chen. The Tianlai Cylinder Pathfinder array: System functions and basic performance analysis. *Science China Physics, Mechanics, and Astronomy*, 63(12):129862, September 2020.

[80] Weitian Li, Haiguang Xu, Zhixian Ma, Ruimin Zhu, Dan Hu, Zhenghao Zhu, Junhua Gu, Chenxi Shan, Jie Zhu, and Xiang-Ping Wu. Separating the eor signal with a convolutional denoising autoencoder: a deep-learning-based method. *Monthly Notices of the Royal Astronomical Society*, 485(2):2628–2637, 2019.

[81] E. Lifshitz. On the Gravitational stability of the expanding universe. *J.Phys.(USSR)*, 10:116, 1946.

[82] E.M. Lifshitz and I.M. Khalatnikov. Investigations in relativistic cosmology. *Adv.Phys.*, 12:185–249, 1963.

[83] Adrian Liu and J Richard Shaw. Data analysis for precision 21 cm cosmology. *Publications of the Astronomical Society of the Pacific*, 132(1012):062001, 2020.

[84] Abraham Loeb and J Stuart B Wyithe. Possibility of precise measurement of the cosmological power spectrum with a dedicated survey of 21 cm emission after reionization. *Physical Review Letters*, 100(16):161301, 2008.

[85] Chung-Pei Ma and Edmund Bertschinger. Cosmological perturbation theory in the synchronous and conformal newtonian gauges. *arXiv preprint astro-ph/9506072*, 1995.

[86] Chung-Pei Ma and Edmund Bertschinger. Cosmological perturbation theory in the synchronous and conformal Newtonian gauges. *Astrophys.J.*, 455:7–25, 1995.

[87] T Lucas Makinen, Lachlan Lancaster, Francisco Villaescusa-Navarro, Peter Melchior, Shirley Ho, Laurence Perreault-Levasseur, and David N Spergel. deep21: a deep learning method for 21 cm foreground removal. *Journal of Cosmology and Astroparticle Physics*, 2021(04):081, 2021.

[88] G. Mangano, G. Miele, S. Pastor, and M. Peloso. A precision calculation of the effective number of cosmological neutrinos. 2001.

[89] Gianpiero Mangano, Gennaro Miele, Sergio Pastor, Teguayco Pinto, Ofelia Pisanti, and Pasquale D Serpico. Relic neutrino decoupling including flavour oscillations. *Nuclear Physics B*, 729(1-2):221–234, 2005.

[90] Y. Mao, M. Tegmark, M. McQuinn, M. Zaldarriaga, and O. Zahn. How accurately can 21cm tomography constrain cosmology? *PRD*, 78(2):023529, July 2008.

[91] K. W. Masui, E. R. Switzer, N. Banavar, K. Bandura, C. Blake, L.-M. Calin, T.-C. Chang, X. Chen, Y.-C. Li, Y.-W. Liao, A. Natarajan, U.-L. Pen, J. B. Peterson, J. R. Shaw, and T. C. Voytek. Measurement of 21 cm brightness fluctuations at z 0.8 in cross-correlation. *The Astrophysical Journal Letters*, 763(1):L20, 2013.

[92] Kiyoshi Wesley Masui, Patrick McDonald, and Ue-Li Pen. Near-term measurements with 21 cm intensity mapping: Neutral hydrogen fraction and bao at z< 2. *Physical Review D*, 81(10):103527, 2010.

[93] KW Masui, ER Switzer, N Banavar, K Bandura, C Blake, L-M Calin, T-C Chang, X Chen, Y-C Li, Y-W Liao, et al. Measurement of 21 cm brightness fluctuations at z 0.8 in cross-correlation. *The Astrophysical Journal Letters*, 763(1):L20, 2013.

[94] G. Montani, M. V. Battisti, and G. R. Benini. *Imponente, Primordial Cosmology, ISBN: 978-981-4271-00-4.* 2011.

[95] M. F. Morales. HI Structure Observations of Reionization and Dark Energy. In R. Minchin and E. Momjian, editors, *The Evolution of Galaxies Through the Neutral Hydrogen Window*, volume 1035 of *American Institute of Physics Conference Series*, pages 82–86, August 2008.

[96] Miguel F Morales and Jacqueline Hewitt. Toward epoch of reionization measurements with wide-field radio observations. *The Astrophysical Journal*, 615(1):7, 2004.

[97] Viatcheslav F. Mukhanov, H.A. Feldman, and Robert H. Brandenberger. Theory of cosmological perturbations. Part 1. Classical perturbations. Part 2. Quantum theory of perturbations. Part 3. Extensions. *Phys.Rept.*, 215:203–333, 1992.

[98] Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, Paul Torrey, Shy Genel, Mark Vogelsberger, Ruediger Pakmor, Federico Marinacci, Rainer Weinberger, et al. The illustristng simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):1–29, 2019.

[99] Laura B Newburgh, Graeme E Addison, Mandana Amiri, Kevin Bandura, J Richard Bond, Liam Connor, Jean-François Cliche, Greg Davis, Meiling Deng, Nolan Denman, et al. Calibrating chime: a new radio interferometer to probe

dark energy. In *Ground-based and Airborne Telescopes V*, volume 9145, page 91454V. International Society for Optics and Photonics, 2014.

[100] L.B. Newburgh et al. HIRAX: A Probe of Dark Energy and Radio Transients. *Proc. SPIE Int. Soc. Opt. Eng.*, 9906:99065X, 2016.

[101] Roger Christopher OBrient. *A log-periodic focal-plane architecture for cosmic microwave background polarimetry.* University of California, Berkeley, 2010.

[102] AR Offringa, AG De Bruyn, M Biehl, S Zaroubi, G Bernardi, and VN Pandey. Post-correlation radio frequency interference classification methods. *Monthly Notices of the Royal Astronomical Society*, 405(1):155–167, 2010.

[103] AR Offringa, JJ Van De Gronde, and JBTM Roerdink. A morphological algorithm for improving radio-frequency interference detection. *Astronomy & astrophysics*, 539:A95, 2012.

[104] Gregory Paciga, Tzu-Ching Chang, Yashwant Gupta, Rajaram Nityanada, Julia Odegova, Ue-Li Pen, Jeffrey B. Peterson, Jayanta Roy, and Kris Sigurdson. The gmrt epoch of reionization experiment: a new upper limit on the neutral hydrogen power spectrum at z 8.6. *Monthly Notices of the Royal Astronomical Society*, 413(2):1174–1183, Mar 2011.

[105] Hamsa Padmanabhan, Alexandre Refregier, and Adam Amara. A halo model for cosmological neutral hydrogen: abundances and clustering h i abundances and clustering. *Monthly Notices of the Royal Astronomical Society*, 469(2):2323–2334, 2017.

[106] T. Padmanabhan. Accelerated expansion of the universe driven by tachyonic matter. *Phys. Rev. D*, 66:021301, Jun 2002.

[107] PJE Peebles. Recombination of the primeval plasma. *The Astrophysical Journal*, 153:1, 1968.

[108] Jeffrey B Peterson, Kevin Bandura, and Ue Li Pen. The hubble sphere hydrogen survey. *arXiv preprint astro-ph/0606104*, 2006.

[109] Anh Phan, Santanu Das, Albert Stebbins, Peter Timbie, Reza Ansari, Shifan Zuo, Jixia Li, Trevor Oxholm, Fengquan Wu, Xuelei Chen, et al. Algoscr: An algorithm for solar contamination removal from radio interferometric data. *Monthly Notices of the Royal Astronomical Society*, 512(3):3520–3537, 2022.

[110] Vivian Poulin, Tristan L. Smith, Tanvi Karwal, and Marc Kamionkowski. Early dark energy can resolve the hubble tension. *Phys. Rev. Lett.*, 122:221301, Jun 2019.

[111] Jonathan R Pritchard and Abraham Loeb. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901, 2012.

[112] Bharat Ratra and P. J. E. Peebles. Cosmological Consequences of a Rolling Homogeneous Scalar Field. *Phys. Rev.*, D37:3406, 1988.

[113] Alexandre Refregier, Lukas Gamper, Adam Amara, and Lavinia Heisenberg. Pycosmo: An integrated cosmological boltzmann solver. *Astronomy and computing*, 25:38–43, 2018.

[114] William J. Riley. Handbook of frequency stability analysis. Technical report, US Department of Commerce, National Institute of Standards and Technology, 2008. Special Publication (NIST SP) - 1065.

[115] J. A. Rubiño-Martín, J. Chluba, W. A. Fendt, and B. D. Wandelt. Estimating the impact of recombination uncertainties on the cosmological parameter constraints from cosmic microwave background experiments. *Monthly Notices of the Royal Astronomical Society*, 403(1):439–452, 03 2010.

[116] George B. Rybicki and Alan P. Lightman. *Radiative Processes in Astrophysics.* 1986.

[117] Pablo F. de Salas and Sergio Pastor. Relic neutrino decoupling with flavour oscillations revisited. 2016.

[118] Mário G Santos, Asantha Cooray, and Lloyd Knox. Multifrequency analysis of 21 centimeter fluctuations from the era of reionization. *The Astrophysical Journal*, 625(2):575, 2005.

[119] S. Seager, D. D. Sasselov, and D. Scott. A New Calculation of the Recombination Epoch. *ApJLett*, 523:L1–L5, September 1999.

[120] Sara Seager, Dimitar D. Sasselov, and Douglas Scott. How exactly did the universe become neutral? 1999.

[121] Uros Seljak and Matias Zaldarriaga. A Line of sight integration approach to cosmic microwave background anisotropies. *Astrophys.J.*, 469:437–444, 1996.

[122] J Richard Shaw, Kris Sigurdson, Ue-Li Pen, Albert Stebbins, and Michael Sitwell. All-sky interferometry with spherical harmonic transit telescopes. *The Astrophysical Journal*, 781(2):57, 2014.

[123] J. Richard Shaw, Kris Sigurdson, Ue-Li Pen, Albert Stebbins, and Michael Sitwell. All-Sky Interferometry with Spherical Harmonic Transit Telescopes. *Astrophys. J.*, 781:57, 2014.

[124] J Richard Shaw, Kris Sigurdson, Michael Sitwell, Albert Stebbins, and Ue-Li Pen. Coaxing cosmic 21 cm fluctuations from the polarized sky using m-mode analysis. *Physical Review D*, 91(8):083514, 2015.

[125] Anže Slosar et al. Packed ultra-wideband mapping array (puma): A radio telescope for cosmology and transients. Technical report, Brookhaven National Lab.(BNL), Upton, NY (United States); SLAC National . . . , 2019.

[126] N Suzuki, D Rubin, C Lidman, G Aldering, R Amanullah, K Barbary, LF Barrientos, J Botyanszki, M Brodwin, N Connolly, et al. The hubble space telescope cluster supernova survey. v. improving the dark-energy constraints above z> 1 and building an early-type-hosted supernova samplebased in part on observations made with the nasa/esa hubble space telescope, obtained from the data archive at the space telescope institute. stsci is operated by the association of universities for research in astronomy, inc., under the nasa contract nas 5-26555. the observations are associated with program go-10496. *The Astrophysical Journal*, 746(1):85, 2012.

[127] ER Switzer, KW Masui, K Bandura, L-M Calin, T-C Chang, X-L Chen, Y-C Li, Y-W Liao, A Natarajan, U-L Pen, et al. 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, 2013.

[128] Eric R. Switzer and Christopher M. Hirata. Primordial helium recombination. i. feedback, line transfer, and continuum opacity. *Phys. Rev. D*, 77:083006, Apr 2008.

[129] S.V. Tassev. *Cosmological Perturbation Theory Beyond Linear Order*. Collections of the Harvard University Archives: Dissertations. Harvard University, 2011.

[130] Steven John Tingay, Robert Goeke, Judd D Bowman, David Emrich, Stephen M Ord, Daniel A Mitchell, Miguel F Morales, Tom Booler, Brian Crosse, Randall B Wayth, et al. The murchison widefield array: The square kilometre array precursor at low radio frequencies. *Publications of the Astronomical Society of Australia*, 30, 2013.

[131] Francisco Villaescusa-Navarro, Shy Genel, Emanuele Castorina, Andrej Obuljen, David N Spergel, Lars Hernquist, Dylan Nelson, Isabella P Carucci, Annalisa Pillepich, Federico Marinacci, et al. Ingredients for 21 cm intensity mapping. *The Astrophysical Journal*, 866(2):135, 2018.

[132] JL Weiland, RS Hill, D Larson, CL Bennett, J Dunkley, N Jarosik, L Page, DN Spergel, M Halpern, SS Meyer, et al. Five-year wilkinson microwave anisotropy probe (wmap) observations: Data processing, sky maps, and basic results. 2008.

[133] Steven Weinberg. *Cosmology*. Oxford University Press, 2008.

[134] Martin White and Emory F Bunn. The cobe normalization of cmb anisotropies. *arXiv preprint astro-ph/9503054*, 1995.

[135] Fengquan Wu, 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. The Tianlai dish pathfinder array: design, operation, and performance of a prototype transit radio interferometer. , 506(3):3455–3482, September 2021.

[136] CA Wuensche, BINGO-Collaboration, et al. The bingo telescope: a new instrument exploring the new 21 cm cosmology window. In *Journal of Physics: Conference Series*, volume 1269, page 012002. IOP Publishing, 2019.

[137] Yidong Xu, Xin Wang, and Xuelei Chen. Forecasts on the dark energy and primordial non-gaussianity observations with the tianlai cylinder array. *The Astrophysical Journal*, 798(1):40, 2014.

[138] Matias Zaldarriaga, Uros Seljak, and Lam Hui. Correlations across scales in the Lyman alpha forest: Testing the gravitational instability paradigm. *Astrophys.J.*, 551:48, 2001.

[139] Saleem Zaroubi. The epoch of reionization. *The First Galaxies*, pages 45–101, 2013.

[140] Jiao Zhang, Reza Ansari, Xuelei Chen, Jean-Eric Campagne, Christophe Magneville, and Fengquan Wu. Sky reconstruction from transit visibilities: Paon-4 and tianlai dish array. *Monthly Notices of the Royal Astronomical Society*, 461(2):1950–1966, 2016.

[141] Juyong Zhang, Jingxin Liu, Fengquan Wu, Xuelie Chen, Jixia Li, Peter T Timbie, Santanu Das, Ruibin Yan, Jiachen He, and Osinga Calvin. Beam measurements of the tianlai dish radio telescope using an unmanned aerial vehicle [antenna applications corner]. *IEEE Antennas and Propagation Magazine*, 63(6):98–109, 2021.

[142] Shifan Zuo, Jixia Li, Yichao Li, Das Santanu, Albert Stebbins, Kiyoshi W Masui, Richard Shaw, Jiao Zhang, Fengquan Wu, and Xuelei Chen. Data processing pipeline for tianlai experiment. *Astronomy and Computing*, 34:100439, 2021.