Analytic and numerical analysis of the cosmic 21cm signal

Claude Jean-Paul Schmit
Astrophysics Group

Department of Physics
Imperial College London

Thesis submitted for the Degree of Doctor of Philosophy to
Imperial College London

· 2018 ·

Abstract

Cosmology in the 21st century has matured into a precision science. Measurements of the cosmic microwave background, galaxy surveys, weak lensing studies and supernovae surveys all but confirm that we live in a geometrically flat Universe dominated by a dark energy component where most of the matter is dark. Yet, challenges to this model remain as well as periods in its evolution unobserved at present. The next decade will see the construction of a new generation of telescopes poised to answer some of these remaining questions and peer into unseen depths. Because of the technological advances of the previous decades and the scale of the new generation of telescopes, for the first time, cosmology will be constrained through the observation of the cosmic 21cm signal emitted by hydrogen atoms across the Universe. Being the ubiquitous element present throughout the different evolutionary stages of the Universe, neutral hydrogen holds great potential to answer many of the remaining challenges which face cosmology today. In the context of 21cm radiation, we identify two approaches which will increase the information gain from future observations, a numerical as well as an analytic approach. The numerical challenges of future analyses are a consequence of the data rates of next generation telescopes, and we address this here introducing machine learning techniques as a possible solution. Artificial neural networks have gained much attention in both the scientific and commercial world, and we apply one such network here as a way to emulate numerical simulations necessary for parameter inference from future data. Further, we identify the potential of the bispectrum, the Fourier transform of the three-point statistic, as a cosmological probe in the context of low redshift 21cm intensity mapping experiments. This higher order statistical analysis can constrain cosmological parameters beyond the capabilities of CMB observations and power spectrum analyses of the 21cm signal. Lastly, we focus on a fully 3D expansion of the 21cm power spectrum in the natural spherical basis for large angle observations, drawing on the success of the technique in weak lensing studies.

Declaration and Copyright

The work presented in this thesis is my own. Exceptions are explicitly indicated in the text.

Claude Schmit
November 2018

Acknowledgments

I would like to thank my supervisors Alan Heavens and Jonathan Pritchard for their continued support during the ups and downs of the last few years. I do want to especially thank Alan for convincing me that I could do the PhD even after initially getting cold feet. I am still utterly struck by the depths of their scientific knowledge, their good humour, and the mathematical wizardry they have exposed to me during my time at Imperial. Simply put, without them I would not be where I am today and for that, I am forever grateful.

Then, I would like to thank my fellow students Josh Greenslade, Rhys Barnett, Ciarán Coneely, and Charlotte Norris for sharing this journey with me. From whiskey to board game nights, and from tea breaks to rants on the bus home, you've made the last four years a lot of fun, and something I won't forget.

For scientific input, besides that provided by my supervisors, I want to especially thank Catherine Watkinson, Suman Majumdar, Emma Chapman, Thomas Binnie and Adélie Gorce. Our weekly meetings have helped me to keep in touch with the rest of the field and the interesting things that you were working on. In the same breath, I would like to thank the rest of the Imperial College astrophysics group for providing a friendly, fun and inspiring place to work. You have all contributed to this, and I thank you for it!

I would then like to thank the Luxembourg Research Fund for making the last four years possible with their financial support under the AFR grant: 'Analytic and numerical analysis of the 21cm signal'.

Naturally, I would like to thank my family for believing in me and supporting me to this point. *Mamm, Papp a Philippe, ouni iech wier dat hei net méiglech gewiescht an dowéinst soen ech iech Merci! Merci datt der ëmmer un mech gegleeft hutt an der ëmmer e gudd Wuert parat hat wann ech et gebraucht hunn. Wann mer eng kéier Zäit hunn, dann erklären ech iech och wat hei dran steet, et ass alles nëmmen hallef sou komplizéiert wei et sech ugesäit!*

Next, I would like to say thank you to my friends around the world for making me feel at home wherever I go! I cannot understate how much this has contributed to me finishing my thesis, as you have given me ample opportunity to clear my head when I needed to. Special thanks to my friends in Luxembourg, you have made every visit feel like I had never left.

Finally, thank you Claire for giving me the push I needed to complete the thesis. *Claire, le soutien que tu m'as donné pendant les derniers mois menant à l'accomplissement de ma thèse est inestimable, merci pour être là pour moi.*

Publications

- ***Paper I:***
  C. J. Schmit and J. R. Pritchard. *Emulation of reionization simulations for Bayesian inference of astrophysics parameters using neural networks*. MNRAS, vol. 475, 1213-1223, March 2018 .

- ***Paper II:***
  C. J. Schmit, A. F. Heavens, and J. R. Pritchard. *The gravitational and lensing-ISW bispectrum of 21cm radiation*. Submitted for publication to MNRAS, October 2018.

Contents

List of Tables

List of Figures

Part I

Preface

"It is better, I think, to grab at the stars than to sit
flustered because you know you cannot reach them...
At least he who reaches will get a good stretch, a good view,
and perhaps even a low-hanging apple for his efforts."

– Drizzt Do'Urden, Sojourn, by Robert A. Salvatore.

Chapter 1

Analytic and numerical analysis of the cosmic 21cm signal

1.1 Introduction

Over the past few decades, technological and methodological advances have pushed the limits of humanity's understanding of our Universe further than ever before. The last few years have seen sensational discoveries from the discovery of gravitational waves (Abbott et al. 2016; Abbott et al. 2016; 2017a;b;c), over the discovery of new worlds (Gillon et al. 2016; 2017), to the precision measurements of our cosmological model (Planck Collaboration et al. 2016b; 2018a;b). These and other discoveries have transpired because of the transition from a science bereft of data to the data driven science that astronomy has become today. We have, for example, progressed to where once individual parallax measurements had to be painstakingly taken, missions like the GAIA (Gaia Collaboration et al. 2018) space-telescope readily measure the positions of billions of stars within the Milky Way. Moreover, current and future ground based telescopes will collect vast amounts of information over the coming decades. The Square Kilometre Array (SKA[1]), one of the most ambitious projects of the astronomical community, will consist of upwards of 130,000 antennas and dishes deployed across two continents and will perform frontier science for a host of fields (Bull et al. 2018). With the capability of outputting multiple times the global internet traffic in data per year, new techniques for data management, storage and analysis will be required and this will pose one of the major challenges for the community in upcoming years. The availability of large data sets has become a global phenomenon, and has brought with it an increasing interest in machine learning as a new way to deal with these data sets. The fields of astronomy and cosmology have thus

[1] https://pos.sissa.it/215/

unsurprisingly seen a surge of machine learning applications in recent times. Ranging from emulation of large scale simulations (Heitmann et al. 2009; Heitmann et al. 2013; Heitmann et al. 2014; 2016) to using Gaussian processes to separate foregrounds from observed data (Mertens et al. 2018), machine learning has already proven to be a useful tool, and is likely to play a critical role in solving major challenges faced in the near future.

Besides the challenge of large data, recent cosmological experiments have lead to the near exhaustion of the cosmic microwave background (CMB) radiation as a cosmological probe, as detailed in Planck Collaboration et al. (2018b), warranting a shift in focus for cosmological measurements. We now have an exquisite understanding of the conditions during the time of recombination. However, as precise as our knowledge of the Universe at redshift $z \sim 1100$ is, some of the most important epochs following recombination are left unobserved to date. The cosmic Dark Ages present an important epoch during which the growth of structure transitions from linear to non-linear structures, yet the complete absence of luminous sources has made any direct measurement of the epoch impossible. This transition importantly ushers in the Cosmic Dawn as the non-linear gravitational collapse leads to the formation of the first luminous structures in the Universe. Again, the enormous distance to these faint sources have left this crucial epoch undetected. Further, these first stars and galaxies start to produce highly ionizing radiation which leads to the second phase transition of the Universe as the hydrogen gas in the intergalactic medium completely ionizes over a relatively short period of time known as the Epoch of Reionization. High redshift quasar observations probe the tail end of this era, and the optical depth to the CMB can give us an estimate of the instantaneous redshift of this otherwise elusive phase transition. All of these epochs exhibit a major presence of neutral hydrogen gas in the intergalactic medium and a useful property of this gas is the existence of a ground state spin flip transition at a wavelength of 21cm. This easily excited transition holds the potential to open up a new frontier of cosmological measurement at all stages during the cosmic evolution. Although 21cm science has its own inherent theoretical and observational challenges, many insights gained from previous studies can be readily applied.

The focus of this thesis is therefore twofold. With the increasing need for numerical techniques to support cosmological data analysis, we study the application of machine learning techniques in the context of Epoch of Reionization simulations. Furthermore, analytical developments are needed to ensure the maximal information gain from future 21cm experiments. We thus focus on the development and adaptation of analysis techniques which have proven useful for the analysis of different cosmic probes to the measurement of the 21cm signal.

1.2 Thesis structure

The thesis is organised into five parts, each subdivided into a number of chapters. Here, we describe each part and the chapters contained within following this introductory part of the thesis.

Part II contains the background material for this thesis. Chapter 2 focusses on developing a basic understanding of the cosmological model used throughout the thesis. The chapter introduces the basic concepts of the standard cosmological model and characterizes the equations of motions which describe the dynamics of our expanding universe. The formation of structure counts towards the most important requirements for a model of the Universe and this chapter thus proceeds by focussing on the linear and non-linear evolution of the dark matter density field. The evolution of the dark matter forms the foundation for the formation of the structures we observe today. Once the conditions are right for star formation, the first light sources appear and bring about the cosmic dawn.

In chapter 3 we focus on the cosmic 21cm signal as a new avenue for cosmological observations at high and low redshifts. The chapter opens with an introduction of the atomic physics that produces the electron transition in atomic hydrogen resulting in the emission of a photon at 1420 MHz and introduces the spin temperature as the critical signal strength parameter. Most of the chapter then discusses the physics which controls the spin temperature of the 21cm signal and thus the evolution of the signal strength from early times until the Epoch of Reionization. During this epoch, high energy photons ionize the intergalactic medium thereby limiting the emission of the signal to isolated shielded regions of atomic hydrogen after reionization has ended. We then proceed by introducing the statistical signal from fluctuations in the global signal and finish the chapter with a look at current and future observational methods.

Part III contains chapter 4 which focusses on the numerical analysis of the 21cm signal in the context of machine learning. The chapter contains the findings of ***Paper I*** and thus opens with a description of the necessity for new computational efforts in the age of big data astronomy. A selection of supervised and unsupervised machine learning techniques are discussed in this context. The analysis presented in this chapter aims to increase the data reduction efficiency for Epoch of Reionization observations by emulating the model evaluation in an MCMC analysis using an artificial neural network emulator. The analysis aims to identify the amount of training samples needed for an accurate representation of the model emulated as well as the optimal choice for the selection of the training samples. We find that a training set of as few as 100 samples selected in such a way that no two samples share a common coordinate results in reliable inference of the observed parameters by the emulating network.

Part IV of the thesis contains two chapters focussing on the derivations of analytic results aimed to amplify the information gain from future 21cm observations. Chapter 5 introduces the work published in ***Paper II***. The chapter principally describes the 21cm bispectrum introduced on the signal by the non-linear gravitational collapse of structure and analyses its detectability and information content in the context of low redshift ($z \sim 1 - 3$) intensity mapping experiments. Besides this primary contribution, the chapter describes an additional contribution to the 21cm bispectrum in the form of the lensing-ISW bispectrum. A Fisher forecast model which compares the bispectrum to 21cm power spectrum observations at the same redshifts is build up and evaluated in this chapter. The main findings of this analysis is a significant decrease in the marginalised errors of the cosmological parameters when the power spectrum and bispectrum analysis is combined as compared to the current best estimate values of the cosmological parameters obtained from CMB observations.

Chapter 6 contains unpublished work on the expansion of the 21cm power spectrum in the 3D spherical harmonic spherical Bessel function basis. Here we introduce the highly successful 3D formalism from Heavens (2003) to 21cm cosmology. Similar to the signal in weak lensing studies, the 21cm signal is an inherently three dimensional signal which is subject to discontinuous sampling due to observational constraints. The natural basis in which to expand the 21cm signal when large areas of the sky are surveyed is that of a spherical harmonic spherical Bessel function basis, and we derive the 21cm power spectrum in this basis in this chapter. Additionally we derive the form of the power spectrum when redshift space distortions are present in the observations.

Part V concludes the thesis by drawing together the insights gained by the various analyses introduced throughout the thesis and provides a scientific outlook for studies based on the material presented in this thesis.

Part II

Background

"Habe nun, ach! Philosophie,
Juristerei und Medizin,
Und leider auch Theology!
Durchaus studiert, mit heißem Bemühn.
Da steh ich nun ich armer Tor!
Und bin so klug als wie zuvor."

– Heinrich Faust, Faust, by Johann W. von Goethe.

Chapter 2

Cosmology

2.1 History of physical cosmology

Humanity's universal desire for a fundamental understanding of the cosmos is self evident from the plethora of creation myths and cosmologies that are found in religions and traditions throughout the world. Cultures throughout history have built elaborate cosmological frameworks within which they could reconcile their observations of the natural world. Most of the ancient cosmologies, such as those of the ancient Egyptians, the Babylonians and the early Greeks were intricately linked to their respective myths and legends, having gods and other supernatural beings at the centre. Tragically, despite early observations of the heavens, progress towards a deeper understanding of the phenomena they saw was impossible for many early cultures as they too often believed all things to be simply at the whim of the gods. It was the Greeks who for the first time subjected everything to the flow of time, implicating that there are powers greater than the divine. This along with the cultural exchange that followed on from trade and wars over the first millenium B.C., allowed the Greeks to use reason to evolve and demystify their cosmology. The birth of deductive reasoning culminated in the first theory of physical cosmology dating back to the 4th century B.C., a time when Aristotle argued for the circular motion of the heavenly bodies around the Earth. Although many of the assertions used to develop his theory are flawed, Aristotle used observations as the basis of his deductions and was able to predict the positions of the planets using his model of the Universe. In the second century A.D., Ptolemy picked up on Aristotle's ideas and formalised them in his *Almagest* which would become the authority on Astronomy until the Copernican Revolution in the 16th century.

After about 10 centuries of stifled scientific progress during the Middle Ages in the West, Nicolaus Copernicus famously rejected the then accepted geocentric model of the Universe. With this he created the foundation for Johannes Kepler and later Isaac Newton to formalise the heliocentric model in which the planets move on ellipses around the

Sun. Although these ideas were only slowly accepted at the time, this marks the perhaps most important moment in the history of cosmology as it no longer places mankind at the centre of the Universe. After Galileo Galilei introduced the telescope as an astronomical instrument in the 17th century, steady progress was possible using observations as the scientific mode of inquiry.

The modern era of physical cosmology began in the late part of the 19th and the 20th century with the foundations for and ultimately the formulation of Einstein's general theory of relativity. Limitations of Newtonian mechanics had become obvious primarily due to the Michelson-Morley experiment, which revealed the lack of a universal frame of reference and allowed for a relative definition of time. Einstein's formulation of special relativity fundamentally linked space and time and also equated energy and mass famously through $E = mc^2$. It subsequently took the inspired insight of the 'equivalence principle' for Einstein to dynamically link these two concepts into, perhaps, the most predictively powerful and conceptually important theory in the history of science; the general theory of relativity. With its help, general relativity allowed for a mathematical formulation of the dynamical properties of the whole Universe, and thus derive a number of powerful predictions assessing the nature of the Universe we live in. After a period of model crafting by people such as Willem de Sitter, Georges Lemaître and Alexander Friedmann in the early part of the 20th century, it was the observational confirmation of the expanding Universe by Vesto Slipher and Edwin Hubble that should steer the development of modern cosmology for the rest of the century and up to the present day.

The discovery of the almost isotropic cosmic microwave background (CMB) by Penzias & Wilson (1965), brought us the richest source of cosmological information yet. Over the past three decades a series of land-based as well as space-based experiments have measured the statistical fluctuations seen in the after-glow of the big bang, and shaped our theoretical understanding of the Universe on the largest scales. The culmination of this effort was the release of the 2018 data analysis of the *Planck* satellite (see Fig. 2.1), constraining model parameters on the 1% level, and setting a new high-precision standard for cosmological observations.

The study of physical cosmology has a long and interesting history, and we will summarize the modern standard cosmological model in this chapter. In section 2.2 we introduce the fundamental building blocks of the standard model of cosmology and develop the dynamics of the expansion of the Universe via the Friedmann equations. From section 2.3 onwards, we focus on structure formation using concepts of first linear perturbation theory, culminating in a form of the linear dark matter power spectrum, before discussing non-linear collapse in section 2.4, in addition to halo abundance models and the formation of the first stars and galaxies. In section 2.5 we take a look at what pri-

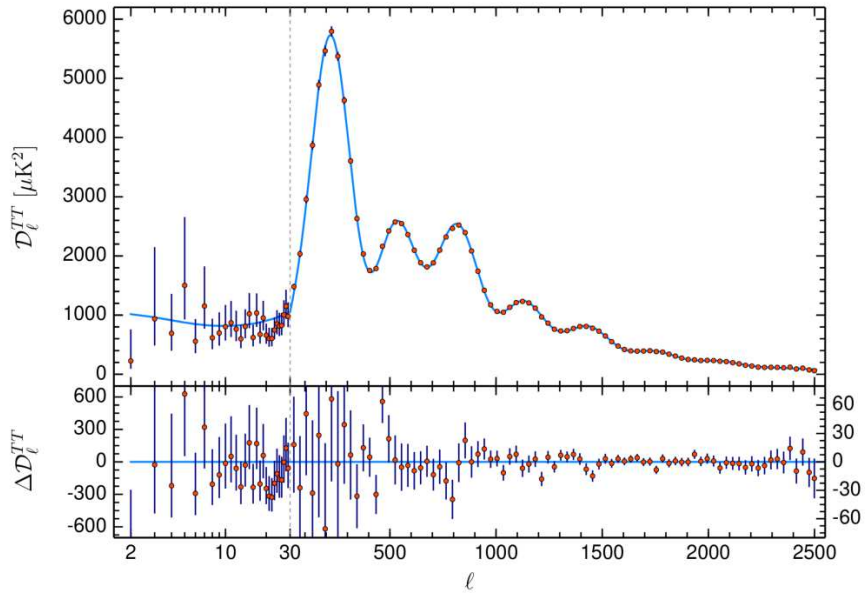

Figure 2.1: The top panel shows the temperature power spectrum as measured by Planck Collaboration et al. (2018b) with $\pm 1\sigma$ error bars. The best fit model (blue line) is a 6 parameter ΛCDM cosmology. At low ℓ the signal is dominated by uncertainties due to cosmic variance, however exquisite precision is achieved at higher ℓ values as seen in the residuals plotted in the lower panel. *Taken with permission from Planck Collaboration et al. (2018a).*

mordial non-Gaussianities may tell us about the inflationary period in the early Universe. Finally, we discuss two line of sight observables in sections 2.6 and 2.7. First we introduce weak lensing, and later the integrated Sachs-Wolfe effect as ways to observe the large scale structure of the Universe today.

2.2 The standard cosmological model

2.2.1 The cosmological principle

The standard cosmological model is fundamentally based upon the assumptions of large scale *homogeneity* and *isotropy*, sometimes referred to as the 'cosmological principle'. These two properties guarantee that structure is distributed uniformly (homogeneity) in all directions (isotropy) when viewed on sufficiently large scales, and allow us to test cosmological models independent of our particular location in the Universe. Although, these properties truly had to be taken as assumptions for most of the development of the cosmological model, and have thus come under severe criticism (Kragh 2012), we now have good evidence in favour of the isotropy and growing confidence in the large scale homogeneity of the Universe. Curiously, the cosmological principle seems to be valid

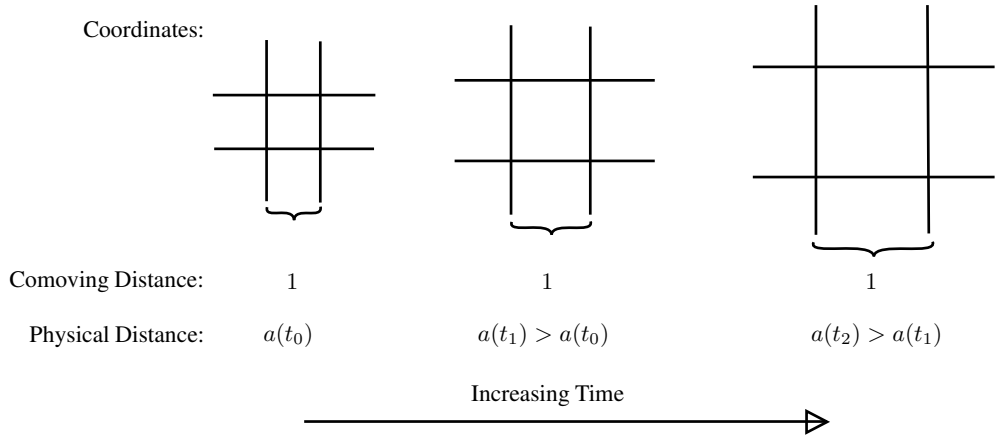

Figure 2.2: As time progresses, objects stay at the same coordinate separation when the physical separation increases due to the expansion of space. The scale factor, $a(t)$, measures the relation between physical and coordinate distance.

within the observable Universe when structure is coarse grained on scales of 100 Mpc, whereas inhomogeneities certainly exist on smaller scales. It is also impossible to know whether the Universe remains homogeneous and isotropic outside our observable patch, which results in a validity of the most fundamental principle in cosmology only over a finite range of scales.

2.2.2 The expanding Universe

Any model of the Universe must reconcile basic observed truths, and thus the observed expansion of the Universe lies at the heart of the standard model. The expansion history is described by a scale factor $a(t)$[1], which increases as time progresses, pictured in Fig. 2.2. It is useful to define a coordinate system which expands with the space in which it is embedded. Thus the comoving distance between coordinate points remains unchanged while the physical distance between them increases with the scale factor, such that

$$\boldsymbol{r} = a(t)\boldsymbol{x}, \tag{2.1}$$

where $\boldsymbol{r}$ denotes the proper distance and $\boldsymbol{x}$ is the comoving distance.

The expansion rate observed by Hubble can thus be related to the scale factor (Hubble 1929),

$$H(t) \equiv \frac{\dot{a}}{a}. \tag{2.2}$$

[1]We use a normalised scale factor whose current value is defined to be $a(t_0) = 1$.

The hubble parameter, $H(t)$, which was discovered as a proportionality constant between the recession velocity and the distance of galaxies, $v = H(t)d$, thus measures the rate of change of the scale factor over time.

2.2.3 Geometrical features

The next step in building a reliable cosmological model, is to assure that two observers at different points in space-time can reliably determine the distance between themselves in an expanding Universe. We therefore need to give these observers a way to transform coordinate distances into some invariant physical distance measure. The connecting matrix for transforming between coordinates in any given system to the physical length is known as the metric, and we can generally write the invariant distance measure in a 4 dimensional space-time, such as our Universe, as

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu, \tag{2.3}$$

where we have used the Einstein summation convention, and repeated indices are then summed over. The μ and ν indices in the above equation range from 0 to 3, and the 0 index connects to the time dimension, such that $dx^0 = cdt$. The metric, $g_{\mu\nu}$, thus connects coordinates to the invariant length element ds, known as the proper time, this is the fundamental distance in a 4 dimensional space-time, which all observers can agree on. The form of this metric depends on the kind of universe the observers live in and their choice of coordinate system in which they choose to measure it. For example, a flat, static universe, fully described by the Minkowski space-time, has a metric, $g_{\mu\nu} = \eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$, when expressed in cartesian coordinates. In order to describe our own Universe, the metric has to obey the cosmological principle, as well as incorporate the expansion of space as a function of time according to Hubble's law. The Friedmann - Lemaître - Robertson - Walker (FLRW or FRW) metric is the unique metric which obeys these constraints, and can be most usefully expressed in spherical polar coordinates as

$$g_{\mu\nu} = \mathrm{diag}\left[-1, a^2(t), a^2(t) S_k^2(r), a^2(t) S_k^2(r) \sin^2\theta\right]. \tag{2.4}$$

The resulting line element is then

$$ds^2 = -c^2 dt^2 + a^2(t)\left[dr^2 + S_k^2(r)\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right], \tag{2.5}$$

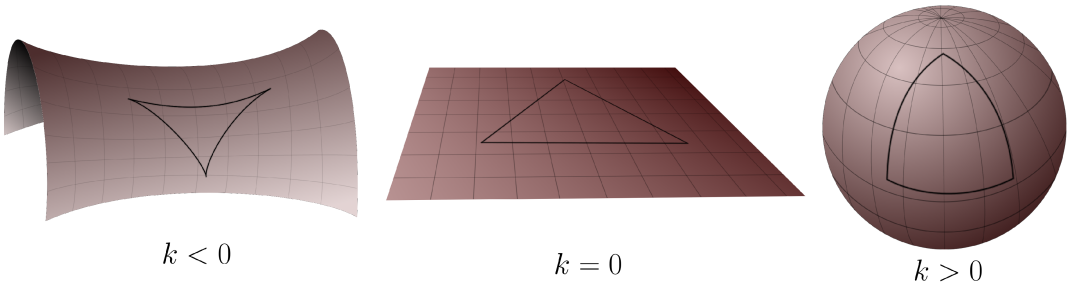

Figure 2.3: Possible geometries of the Universe as a function of k. From left to right, $k < 0$ implies a hyperbolic geometry where the sum of the angles of a triangle is less than $\pi/2$, $k = 0$ implies Euclidean geometry where the angles of a triangle always add to $\pi/2$, and $k > 0$ implies spherical geometry where triangles have the sum of their angles add to more than $\pi/2$.

where $S_k(r)$ allows for the geometrical uncertainty we have of our Universe,

$$S_k(r) = \begin{cases} \sqrt{k}^{-1} \sin\left(r\sqrt{k}\right), & k > 0, \text{closed}, \\ r, & k = 0, \text{flat}, \\ \sqrt{|k|}^{-1} \sinh\left(r\sqrt{|k|}\right), & k < 0, \text{open}. \end{cases} \tag{2.6}$$

The curvature constant k, which is related to the energy contents of the Universe, as we will see later, sets the geometrical properties of the metric. A positive curvature leads to a closed universe with a spherical geometry, a negative curvature leads to an open universe with a hyperbolic geometry, and zero curvature leads to a flat universe. The three cases are shown in Fig. 2.3.

2.2.4 Cosmological redshift

Although we have yet to define the dynamics of our model of the Universe, we can already use the FRW metric to compute the trajectories on which light propagates. Defining an observer to be at the origin, light travels on radial null geodesics (trajectories with zero proper time). The FRW metric thus reduces to the following for photons,

$$ds^2 = 0 = -c^2dt^2 + a^2(t)dr^2. \tag{2.7}$$

As two observers, or galaxies, move apart in an expanding universe, relative to one, the other is receding. Light emitted from a receding source is subject to a redshift z. The redshift of an object is directly related to the change of frequency between the emitted and observed frequency of a beam of light. It is then useful to consider two beams of light emitted a short amount of time after another, and realise that both beams travel the same

comoving distance r, in order to relate the frequency shift to a change in the scale factor,

$$\frac{\nu_{\rm emit}}{\nu_{\rm obs}} \equiv 1 + z = \frac{a(t_{\rm obs})}{a(t_{\rm emit})}. \tag{2.8}$$

2.2.5 Dynamics of the expanding Universe

For a dynamical description of the Universe, we require the use of general relativity (GR). The equation of motion for a particle in a curved space-time can be obtained by applying the principle of least action to the FRW metric and results in the geodesic equation,

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} = 0, \tag{2.9}$$

where $\Gamma^\mu_{\alpha\beta}$ represents the affine connection, and τ is the conformal time. Further, it took Einstein's insight to relate the dynamical properties of the Universe's content to the metric, via his field equations,

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu}. \tag{2.10}$$

The *lhs* of this equation consists of the Ricci tensor $R_{\mu\nu}$, which describes the degree of curvature of a topological manifold, and the Ricci scalar R, which describes the difference in volume of a ball in the curved space to that of a ball in Euclidean space. The *rhs* of equation (2.10) describes the contents of the Universe via the energy-momentum tensor $T_{\mu\nu}$. Famously summarised by John Wheeler, these equations are described by the following two statements:

- Space-time tells matter how to move,
- Matter tells space-time how to curve.

When combining the FRW metric with Einstein's field equations we obtain a set of three equations that describe the expansion of space in a GR context. The Friedmann equations, named after Alexander Friedmann, are

$$H^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda}{3}, \tag{2.11}$$

and,

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3\frac{P}{c^2}\right) + \frac{\Lambda}{3}, \tag{2.12}$$

which can be combined to form a third,

$$\dot{\rho} + 3H\left(\rho + \frac{P}{c^2}\right) = 0. \tag{2.13}$$

From these equations, we can derive some conditional parameters which allow insight into the dynamics of the Universe. Evaluating the first Friedmann equation today, we see that

$$H_0^2 = \frac{8\pi G}{3}\rho_0 - kc^2 + \frac{\Lambda}{3}. \tag{2.14}$$

Considering a flat universe without cosmological constant, $k = \Lambda = 0$, allows us to determine a critical value for the current density, ρ_0, for which the universe has zero curvature,

$$\rho_c = \frac{3H_0^2}{8\pi G}. \tag{2.15}$$

Any $\Lambda = 0$ universe whose energy density differs from ρ_c will thus express non-zero curvature. It is then useful to define a density parameter which measures the density relative to the critical density of the universe,

$$\Omega_0 = \frac{\rho_0}{\rho_c} = \frac{8\pi G\rho_0}{3H_0^2}. \tag{2.16}$$

Various sources contribute to the energy content of the universe and so ρ_0 is the sum of energy densities of these components. Using the first Friedmann equation and expressing it in terms of the critical density one finds

$$H^2(a) = H_0^2\left[\Omega_M(a) + \Omega_R(a) + \Omega_k(a) + \Omega_\Lambda\right], \tag{2.17}$$

where we have introduced the matter and radiation contributions to the energy density and defined

$$\Omega_k(a) = -\frac{kc^2}{H_0^2 a^2} \quad \text{and} \quad \Omega_\Lambda(a) = \frac{\Lambda}{3H_0^2}, \tag{2.18}$$

as the density parameters for curvature and dark energy respectively. An appropriate time evolution for the density parameters can be found by analysing the third Friedmann equation in the matter dominated and radiation dominated limits. For the *matter dominated* case, we can assume a near pressureless fluid as most matter in the Universe is dark and only interacts gravitationally. Then the function that solves equation (2.13) is,

$$\rho_M \propto a^{-3}, \tag{2.19}$$

and as a consequence[2],

$$a \propto t^{2/3}. \tag{2.20}$$

[2]From the Friedmann equation: $\dot{a} \propto a^{-1/2} \Rightarrow \int a^{1/2} da \propto \int dt \Rightarrow a \propto t^{2/3}$.

For the *radiation dominated* case, we use the equation of state, $P = 3\rho_{\rm R}c^2$, and thus

$$\rho_{\rm R} \propto a^{-4}. \tag{2.21}$$

Similarly for the radiation dominated era one finds[3]

$$a \propto t^{1/2}. \tag{2.22}$$

The fact that the scale factor and the energy densities evolve at different rates during matter and radiation dominated epochs is a fundamental result which has implications on the growth of structure as we will see in section 2.3.

We can then express $H(a)$ as a function of the current values for the density parameters,

$$H^2(a) = H_0^2\left(\Omega_{\rm M,0}a^{-3} + \Omega_{\rm R,0}a^{-4} + \Omega_{k,0}a^{-2} + \Omega_{\Lambda,0}\right). \tag{2.23}$$

Today, there is almost no contribution from radiation in terms of the energy content of the Universe. The CMB, which makes up the vast majority of the photons in the Universe, is a blackbody with a temperature of $T_{\rm CMB} \simeq 2.75$K, which compared to the critical density results in $\Omega_{\rm R,0} \simeq 5 \times 10^{-5}$. However, as the radiation density goes as a^{-4}, at early times, photons were the dominant constituent of the Universe. Observational efforts over the last three decades, experiments such as COBE, WMAP and *Planck*, have pinned the values of the other density parameters at (Planck Collaboration et al. 2018a)

$$\Omega_{k,0} = 0.001 \pm 0.002, \quad \Omega_{\rm M,0} = 0.315 \pm 0.007 \quad \text{and} \quad \Omega_{\Lambda,0} = 0.685 \pm 0.007. \tag{2.24}$$

We can thus see that the dominant component is that of dark energy through a cosmological constant, gravitational matter only constitutes about a third of the total energy budget today, and the Universe is consistent with having no curvature. Furthermore, the expansion rate inferred from the CMB is

$$H_0 = 100h \text{ km/s/Mpc} \quad \text{and} \quad h = 0.6727 \pm 0.0060. \tag{2.25}$$

Of note is that there remain tensions between the Hubble rate as inferred from the CMB, giving the value presented above, and other measurements. Riess et al. (2018b) have used standard candles to calibrate the local distance ladder and thus measured the Hubble parameter locally to find a somewhat faster expansion rate of $h = 73.48 \pm 1.66$. Similarly, the *H0LiCOW* collaboration (Bonvin et al. 2017; Birrer et al. 2019) have consistently

[3]From the Friedmann equation: $\dot{a} \propto a^{-1} \Rightarrow \int ada \propto \int dt \Rightarrow a \propto t^{1/2}$.

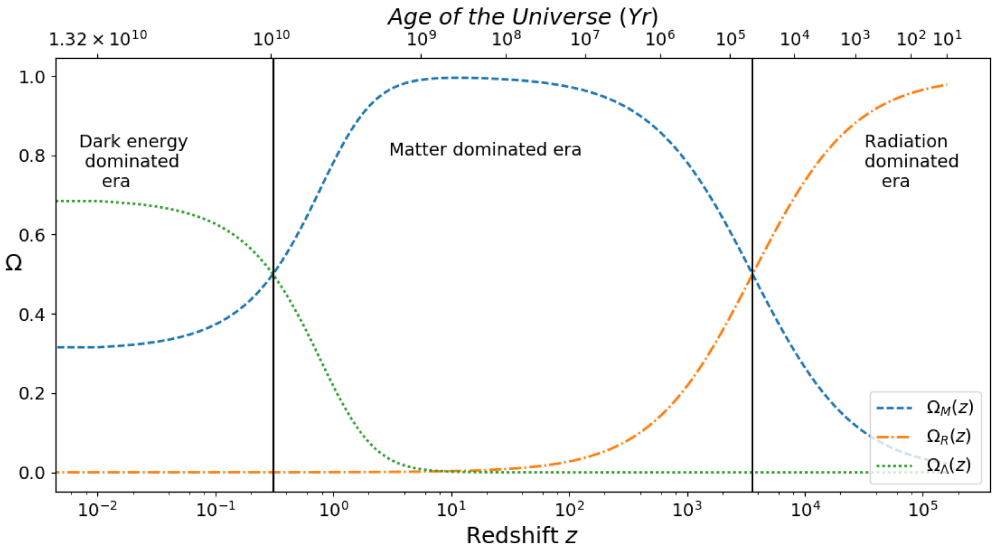

Figure 2.4: Evolution of the energy density components as a function of age of the Universe (top axis) and redshift (bottom axis). We show that the contributions of matter (blue), $\Omega_{\rm M}$, radiation (orange), $\Omega_{\rm R}$, and dark energy (green), Ω_Λ, dominate the energy budget of the Universe at different times.

found high values of the Hubble constant using strong lensing time delays. Recently, Macaulay et al. (2018) have inferred a value of $h = 0.6777 \pm 0.0130$, consistent with the *Planck* measurements, from BAO measurements. The determination of the expansion rate is thus far from settled and analyses of the systematics of different methods are needed to reconcile these results.

In Fig. 2.4, we show the evolution of the density parameters as a function of redshift and highlight the different eras in which each component was dominant, according to the model defined by equations (2.24) and (2.25). It is important to realize that the Universe was matter dominated for the majority of cosmic history, and we will study it in more detail when talking about structure formation in section 2.3. In the recent past, since $z \sim 1$, the matter density has decreased and the dark energy component has become the dominant contributor to the energy budget of the Universe. The model predicts that this trend will continue on into the future until $\Omega_\Lambda \approx 1$, and all other forms of energy are dominated completely by dark energy. This would ultimately result in the "heat death" of the Universe, a state where the Universe expands indefinitely, continuously increasing the distances between stars and galaxies, creating an environment inadequate for star formation.

2.2.6 Age and distance measures in the expanding Universe

Possibly one of the most important aspects of the Big Bang model is the fact that the Universe has only existed for a finite amount of time. We can compute the age of our expanding Universe, t_0, by integrating a time element from the Big Bang to the present day. Differentiating the redshift relation (2.8), we can change variables such that the integral can be written as

$$t_0 = H_0^{-1} \int_0^\infty \frac{dz}{(1+z)E(z)}, \tag{2.26}$$

where $E(z)$ is derived from equation (2.23) and defines the cosmological model,

$$E(z) = \left[\Omega_{\mathrm{M},0}(1+z)^3 + \Omega_{\mathrm{R},0}(1+z)^4 + \Omega_{k,0}(1+z)^2 + \Omega_{\Lambda,0}\right]^{1/2}. \tag{2.27}$$

We can thus see that the age is related to the Hubble time, H_0^{-1}, and a correction factor which is determined by the energy content of the Universe. A simple matter dominated universe can be shown to result in a correction factor of exactly $2/3$. However, as seen in the previous section, a significant contribution from a cosmological constant implies that the Universe is much older than without it. The best observational constraints come from the *Planck* satellite (Planck Collaboration et al. 2018a) and put the age at $t_0 = 13.800 \pm 0.024$ Gyrs.

Due to the expansion of the Universe, distance measures in cosmology can be ambiguous as the distance between two points is constantly changing (Hogg 1999). There exist a number of useful distance definitions which are applicable in a variety of contexts. The proper distance along the line of sight defines the physical distance we would measure if we froze the expansion of the Universe at the time of observation. At low redshifts, the proper distance is that which we would intuitively observe. This distance depends on the expansion history and thus on the factor $E(z)$ introduced in equation (2.27), such that the proper distance between us and an object at redshift z is

$$r = cH_0^{-1} \int_0^z \frac{dz'}{E(z')}. \tag{2.28}$$

For observational purposes, the most important distance measures are that of the angular diameter distance,

$$D_\mathrm{A} = (1+z)^{-1} D_\mathrm{M}, \tag{2.29}$$

and the luminosity distance,

$$D_\mathrm{L} = (1+z) D_\mathrm{M}, \tag{2.30}$$

which both depend on the transverse comoving distance,

$$D_{\mathrm{M}} = cH_0^{-1} S_k(r). \tag{2.31}$$

The angular diameter distance measures the ratio of an objects physical size to its angular size subtended on the sky. This ratio has a turn-over at $z \sim 1$ such that objects at high redshifts appear larger than at low redshifts. The luminosity distance measures the ratio between the bolometric flux observed and the luminosity of a distant object and thus measures how far away the object is based on the amount of light arriving at the observer.

2.3 Linear perturbation theory

Up to this point we have discussed the expansion history and energy contents of a universe adhering completely to the cosmological principle and especially to homogeneity. However, our existence is prime proof for this not being the complete picture. In a perfectly homogeneous universe, no structure can ever form as the gravitational pull on matter at any point cancels exactly due to the uniform pull from every direction. The theoretical framework which is most widely accepted and discussed here is that of primordial density fluctuations which grow over time and eventually form the galaxies and clusters we observe today. As long as these density perturbations are small, linear perturbation theory, as developed in this section, completely describes the evolution of structure formation. Once perturbations become large, a non-linear description for structure formation becomes necessary (see section 2.4).

Considering a universe with a homogeneous density distribution, let us assume a spherical region which is slightly overdense as compared to the surrounding, background density,

$$\delta(\boldsymbol{x}, t) \equiv \frac{\rho(\boldsymbol{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}, \tag{2.32}$$

where $\rho(\boldsymbol{x}, t)$ denotes the density at a comoving location $\boldsymbol{x}$ at some time t, and $\bar{\rho}(t)$ denotes the density of the background at that time. This overdensity generates an increased gravitational pull on the surrounding material, such that more and more matter will fall into the overdensity and thus creating a growing perturbation. At the same time, the Hubble expansion will exert pressure on the matter in and surrounding the perturbation which will suppress or inhibit its growth. A sufficiently dense perturbation decouples from the Hubble expansion and can in principle grow indefinitely. Regions with underdensities, where $\delta < 0$, exhibit the opposite and material is drawn out from them more quickly than

from the surrounding areas. Over time, this process leads to the emergence of voids in the structure distribution we see today.

Linear perturbation theory studies the evolution of these density fluctuations when they are small, $|\delta| \ll 1$, in a regime where the equations of motion for the perturbed quantities can be linearized. This regime is especially important as all fluctuations start out in this regime such that the crucial beginning era of structure formation is modelled by linear perturbation theory. Perturbation amplitudes appear to be scale dependent and have higher amplitudes on small scales, such that these scales become non-linear earlier than larger scales. This leads to hierarchical structure formation, where differing scales are linear at different epochs, and thus the description developed in this section becomes important for different scales throughout time. Today, scales smaller than $\sim 10h^{-1}$Mpc have become highly non-linear, yet the linear regime still applies for the largest scales.

2.3.1 Equations of motion

The main assumption of linear perturbation theory is that the gravitating contents of the Universe can be modelled as a fluid on the scales at which the theory is applied. This is certainly a valid assumption on the largest scales where the discrete nature of baryonic and dark matter can be ignored. The evolution of a fluid in a gravitational field is then subject to the following three equations;

The continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}) = 0. \tag{2.33}$$

The Euler equation:

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\frac{1}{\rho}\nabla P - \nabla \Phi. \tag{2.34}$$

The Poisson equation:

$$\nabla^2 \Phi = 4\pi G \rho. \tag{2.35}$$

We will proceed to expand these equations in terms of perturbed quantities as well as comoving coordinates.

2.3.1.1 Continuity equation

The continuity equation (2.33) simply guarantees that the change in mass in a volume is equal to the flux of material into or out of the volume. Introducing comoving coordinates, $\boldsymbol{x} = \boldsymbol{r}/a$, both temporal and spatial derivatives need to be adjusted. The spatial derivative

transforms in a straightforward manner,

$$\nabla_{\boldsymbol{r}} = \frac{\partial}{\partial r^i} = \frac{1}{a}\frac{\partial}{\partial x^i} = \frac{1}{a}\nabla_{\boldsymbol{x}}. \tag{2.36}$$

For the temporal derivative, consider a function $f(t, r)$, where $r = a(t)x$,

$$\begin{aligned} df &= \left.\frac{\partial f}{\partial t}\right|_r dt + \left.\frac{\partial f}{\partial r}\right|_t dr \\ &= \left.\frac{\partial f}{\partial t}\right|_r dt + \left.\frac{\partial f}{\partial r}\right|_t (\dot{a}x dt + a dx) \\ &= \left[\left.\frac{\partial f}{\partial t}\right|_r + Hr\left.\frac{\partial f}{\partial r}\right|_t\right] dt + a\left.\frac{\partial f}{\partial r}\right|_t dx \end{aligned} \tag{2.37}$$

In 3 dimensions, we thus find that the time derivative at a fixed $\boldsymbol{r}$ translates into comoving coordinates via

$$\left.\frac{\partial}{\partial t}\right|_{\boldsymbol{r}} = \left.\frac{\partial}{\partial t}\right|_{\boldsymbol{x}} - H\boldsymbol{x}\cdot\nabla_{\boldsymbol{x}}. \tag{2.38}$$

Further, we note that in an unperturbed universe, matter simply expands with the Hubble flow, such that

$$\boldsymbol{v}_{\mathrm{b}} = H\boldsymbol{r}. \tag{2.39}$$

Deviations from the background flow due to local changes in the gravitational field will result in peculiar velocities,

$$\boldsymbol{u} = \frac{d\boldsymbol{r}}{dt} = \boldsymbol{v}_{\mathrm{b}} + a\dot{\boldsymbol{x}}, \tag{2.40}$$

where we define the peculiar velocity, ie. the velocity deviation from the Hubble flow, as

$$\boldsymbol{v}_{\mathrm{p}} = a\dot{\boldsymbol{x}}. \tag{2.41}$$

Using the results derived up to this point, the continuity equation becomes

$$\left.\frac{\partial\rho}{\partial t}\right|_{\boldsymbol{x}} - H(\boldsymbol{x}\cdot\nabla_{\boldsymbol{x}})\rho + \frac{1}{a}\nabla_{\boldsymbol{x}}\cdot\left[\rho\left(\boldsymbol{v}_{\mathrm{b}} + \boldsymbol{v}_{\mathrm{p}}\right)\right] = 0. \tag{2.42}$$

Expanding the spatial derivative and introducing the density perturbations via equation (2.32) one obtains

$$\frac{\partial\bar{\rho}}{\partial t} + 3H\bar{\rho} + \delta\left[\frac{\partial\bar{\rho}}{\partial t} + 3H\bar{\rho}\right] + \bar{\rho}\left\{\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla_{\boldsymbol{x}}\left[(1+\delta)\boldsymbol{v}_{\mathrm{p}}\right]\right\} = 0. \tag{2.43}$$

At 0^{th} order in the perturbation, the first two terms sum to zero, thus the term multiplied by δ equally vanishes, leaving the perturbed continuity equation

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla_{\boldsymbol{x}}\left[(1+\delta)\boldsymbol{v}_{\mathrm{p}}\right] = 0. \tag{2.44}$$

Of note is that this is a second order equation in perturbed quantities δ and $\boldsymbol{v}_{\mathrm{p}}$, and in the linear regime, the second order term can be ignored, resulting in the linearized equation

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla_{\boldsymbol{x}} \cdot \boldsymbol{v}_{\mathrm{p}} = 0. \tag{2.45}$$

2.3.1.2 Euler equation

The Euler equation (2.34) is a consequence of momentum conservation and embodies Newton's second law for the fluid components of the Universe. It describes the change in the velocity field sourced by the forces acting on the fluid. Those forces are the force due to pressure in the fluid and the gravitational attraction by matter in the Universe. Using the results derived for the continuity equation, the Euler equation becomes

$$\frac{\partial \boldsymbol{u}}{\partial t} - H(\boldsymbol{x}\cdot\nabla_{\boldsymbol{x}})\boldsymbol{u} + \frac{1}{a}(\boldsymbol{v}_{\mathrm{b}}\cdot\nabla_{\boldsymbol{x}})\boldsymbol{u} + \frac{1}{a}(\boldsymbol{v}_{\mathrm{p}}\cdot\nabla_{\boldsymbol{x}})\boldsymbol{u} = -\frac{1}{a\rho}\nabla_{\boldsymbol{x}}P - \frac{1}{a}\nabla_{\boldsymbol{x}}\Phi. \tag{2.46}$$

The second and third term cancel and we expand $\boldsymbol{u} = \boldsymbol{v}_{\mathrm{b}} + \boldsymbol{v}_{\mathrm{p}}$, while at the same time introducing the perturbations in the pressure and gravitational field. Considering only fluctuating terms and using the fact that $(\boldsymbol{v}_p \cdot \nabla_{\boldsymbol{x}})\boldsymbol{x} = \boldsymbol{v}_p$, we obtain

$$\frac{\partial \boldsymbol{v}_{\mathrm{p}}}{\partial t} + H\boldsymbol{v}_{\mathrm{p}} + \frac{1}{a}(\boldsymbol{v}_{\mathrm{p}}\cdot\nabla_{\boldsymbol{x}})\boldsymbol{v}_{\mathrm{p}} = -\frac{1}{a\bar{\rho}(1+\delta)}\nabla_{\boldsymbol{x}}\delta P - \frac{1}{a}\nabla_{\boldsymbol{x}}\phi. \tag{2.47}$$

Similar to the continuity equation, linearizing the equation yields the useful form,

$$\frac{\partial \boldsymbol{v}_{\mathrm{p}}}{\partial t} + H\boldsymbol{v}_{\mathrm{p}} = -\frac{1}{a}\nabla_{\boldsymbol{x}}\phi, \tag{2.48}$$

where we have ignored the pressure term as during the matter dominated era, when most of the linear structure growth occurs, the most important component of material in the Universe is pressureless dark matter.

2.3.1.3 Poisson equation

The Poisson equation as expressed in (2.35) ignores the pressure term due to radiation pressure as we are mostly interested in the matter dominated era. The equation thus shows that the gravitational potential Φ is sourced by the density distribution of matter in

the Universe. Considering fluctuations in the density distribution, it is easy to see that the gravitational potential will be perturbed as well,

$$\Phi = \Phi_{\rm b} + \phi. \tag{2.49}$$

Transforming to comoving coordinates, the first order perturbation to the Poisson equation becomes

$$\nabla_{\boldsymbol{x}}^2 \phi = 4\pi G a^2 \bar{\rho} \delta. \tag{2.50}$$

2.3.2 Matter dominated Universe

As seen in Fig. 2.4, the matter dominated era spans from beyond $z_{\rm rec}$ to the 'recent' past, $z \sim 0.3$, and thus comprises the vast majority of the structure formation in it. As a consequence, we solve the equations of motion for the density fluctuation in the pressureless limit defining this epoch. The equation governing the evolution of density perturbations in a flat matter dominated universe can be obtained by combining the divergence of equation (2.48) with the linearised continuity equation (2.45) and substituting equation (2.50), one finds

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G \bar{\rho}_{\rm M} \delta. \tag{2.51}$$

The *rhs* of this equation becomes $4\pi G\bar{\rho}_{\rm M}\delta = \frac{3}{2}H^2\delta$, when $\Omega_{\rm M} \simeq 1$. We have seen that during matter domination, $\bar{\rho}_{\rm M} \propto a^{-3}$ and $a \propto t^{2/3}$, using these facts with the Ansatz $\delta = a^n$, one finds two solutions to the equation of motion,

$$\delta \propto a \quad \text{and} \quad \delta \propto a^{-3/2}. \tag{2.52}$$

The negative exponent solution constitutes the decaying mode for density fluctuations which will decrease over time and is subdominant to the positive exponent solution, the growing mode. One can thus see that throughout matter domination, density perturbations grow with the scale factor, $\delta \propto a$. This result encodes the main reason why we see structure around us today, even though the only departure from homogeneity in the early Universe were minute density fluctuations. From the Poisson equation (2.50), one can easily see that $\phi \propto \delta/a$ and thus the gravitational potential does not grow during matter domination. The depth of the potential fluctuations are thus frozen into place until dark energy takes over at later times.

2.3.3 Radiation dominated Universe

At early times, the Universe is dominated by a radiation component, with a minute matter content, $\bar{\rho}_{\rm M}/\bar{\rho}_{\rm R} \sim 0$. Note that the total density, $\rho = \rho_{\rm M} + \rho_{\rm R}$, sources the gravitational potential fluctuations. Whereas special relativistic fluid equations are necessary to analyse perturbations in the radiation component, it can be easily shown that the fluctuations in the dominant component oscillate rapidly during this time, which implies that the average density contrast at any location will be zero (see Padmanabhan et al. 2016). Qualitatively the oscillatory nature of the radiation component can be understood as it experiences radiation pressure during this time which acts against the inward gravitational pull. Schematically the equation governing the evolution of the perturbations reads,

$$\ddot{\delta} + [\text{Pressure} - \text{Gravity}]\delta = 0, \tag{2.53}$$

where oscillations occur if the pressure term is large. It is important to note that during this epoch, the perturbations in the dominant radiation component influences the gravitational potential, which in turn influences the dark matter perturbations (see Dodelson 2003; for derivation.).

Here, we are interested in how the matter component evolves during this time and can thus look at equation (2.51) setting the *rhs* to zero as both $\delta_{\rm R} \sim 0$ and $\bar{\rho}_{\rm M} \sim 0$,

$$\ddot{\delta} + \frac{1}{t}\dot{\delta} = 0, \tag{2.54}$$

where we have used the fact that $H = 1/2t$ during radiation domination. This equation has two solutions,

$$\delta = \text{const.} \quad \text{and} \quad \delta \propto \ln t \propto \ln a. \tag{2.55}$$

Matter fluctuations therefore grow slowly, logarithmically, during this period, before their growth speeds up once matter becomes the dominant component. This is a direct consequence of the oscillations in the radiation fluctuations to which the density perturbations are coupled.

2.3.4 Dark energy dominated Universe

Today, we observe a significant dark energy component, which drives the accelerated expansion of our Universe. This dark energy does not cluster, such that in this regime, the perturbation equation takes the familiar form,

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G\bar{\rho}_{\rm M}\delta, \tag{2.56}$$

where H remains roughly constant. In that case, we assume that the matter component is small, such that

$$\ddot{\delta} + 2H\dot{\delta} \simeq 0. \tag{2.57}$$

Trying an exponential solution one finds that this equation, again, has two solutions,

$$\delta = \text{const.} \quad \text{and} \quad \delta \propto e^{-2Ht}. \tag{2.58}$$

One thus sees that at late times the density fluctuations stop growing. The Poisson equation then shows that due to the accelerated expansion of the Universe, the gravitational potential decays as a^{-1}.

2.3.5 Scales beyond the horizon

When considering superhorizon scales, a full general relativistic analysis of the perturbation equations is necessary for a rigorous derivation of the evolution of perturbations on those scales. However, by considering a perturbation on a superhorizon scale itself as a slightly overdense universe, embedded in a flat universe, the main results can be shown by a simple argument. Consider a flat background universe, then the Friedmann equation (2.11) gives

$$H^2 = \frac{8\pi G}{3}\rho_0, \tag{2.59}$$

where ρ_0 denotes the density which results in zero curvature. A slightly overdense universe, with $\rho_1 > \rho_0$, will thus expose a curvature term in the same Friedmann equation,

$$H^2 = \frac{8\pi G}{3}\rho_1 - \frac{kc^2}{a_1^2}. \tag{2.60}$$

We consider that both equations really represent two regions in the same universe, and thus have the same expansion rate, we find

$$\delta \equiv \frac{\rho_1 - \rho_0}{\rho_0} = \frac{3kc^2}{8\pi G\rho_0 a_1^2}. \tag{2.61}$$

If the perturbation is small, ie. in the linear regime, then $a_1 \simeq a_0 \equiv a$. Now, from this result, we can see that during radiation domination, $\rho_0 \propto a^{-4}$, therefore $\delta \propto a^2$, whereas during matter domination, where $\rho_0 \propto a^{-3}$, the perturbations grow at the same rate outside the horizon as they do inside, namely $\delta \propto a$.

We then see an important distinction between those scales that enter the horizon during radiation domination, ie. small scales, and those that enter during matter domination, ie. larger scales. Those that enter during RD are hindered in their growth by the rapid

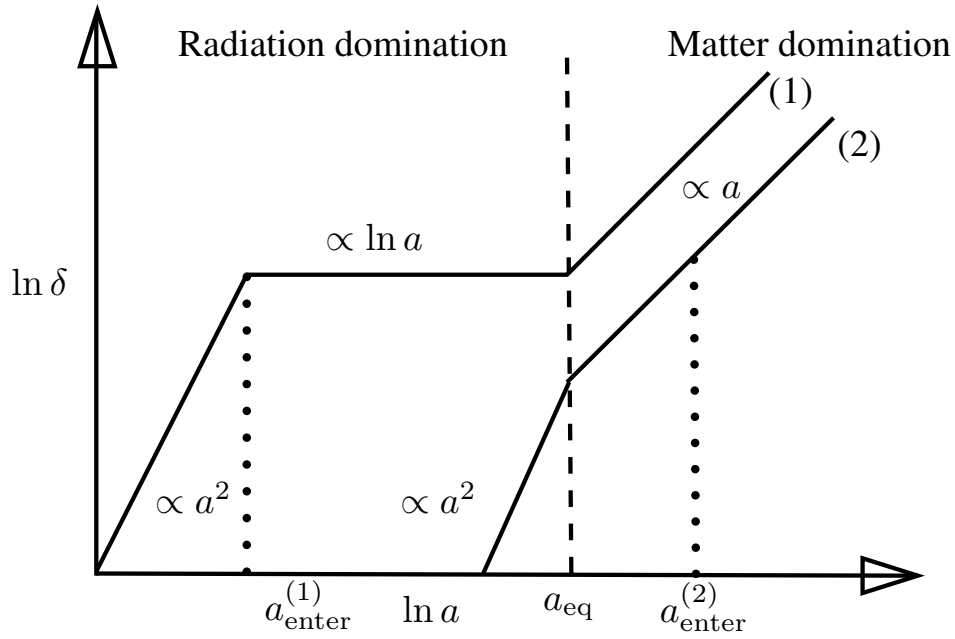

Figure 2.5: Summary of the growth of linear perturbations with time. We show the evolution of two different scales entering the horizon at two different times. A small scale perturbation, (1), entering during radiation domination, $a_{\rm enter}^{(1)}$, experiences stifled growth until the Universe becomes matter dominated at $a_{\rm eq}$. A larger scale, (2), perturbation enters the horizon later, during matter domination, $a_{\rm enter}^{(2)}$, and thus grows freely throughout.

expansion induced by the relativistic component at early times. Scales which are still outside the horizon at this time therefore freely grow with the hubble expansion ($\delta \propto a^2$), and do not experience a similar lag. As a consequence we should see a suppression of small scale structure up to a scale which is equal to the horizon scale at matter-radiation equality, as any scales larger than that will enter the horizon once matter has become the dominant component and will therefore never experience logarithmic growth. This effect is called the Mészáros effect and we illustrate it in Fig. 2.5.

2.3.6 The matter power spectrum

We have already seen that different effects affect the density perturbations on different scales, and it is therefore useful to consider these results in terms of the Fourier decomposition of the fluctuations. Furthermore, primordial density fluctuations are inherently random in amplitude so a statistical analysis of the fluctuations on different scales is useful. A useful measure is then the power spectrum $P(\boldsymbol{k})$, which is defined via

$$\left\langle \delta_{\boldsymbol{k}} \delta^*_{\boldsymbol{k}'} \right\rangle = (2\pi)^3 \delta^{\rm D}(\boldsymbol{k} - \boldsymbol{k}') P(\boldsymbol{k}), \tag{2.62}$$

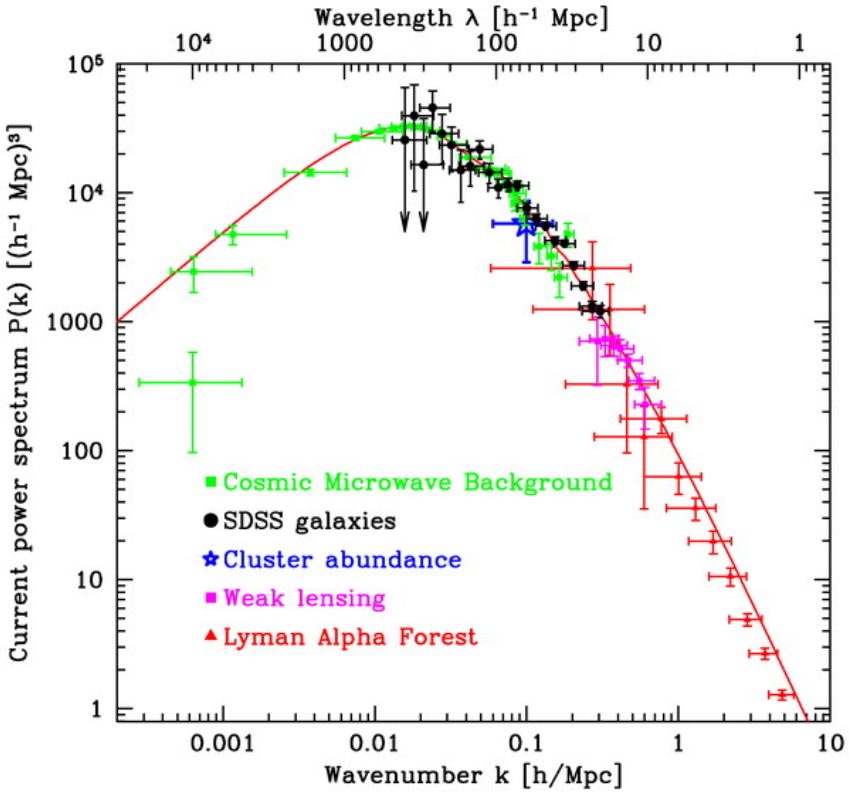

Figure 2.6: Summary of a variety of local matter power spectrum observations. The red line shows a standard flat, scale-invariant cosmology using $\Omega_{\rm M} = 0.28$, $h = 0.72$, $\Omega_{\rm b}/\Omega_{\rm M} = 0.16$, and $\tau = 0.17$. *Taken with permission from Tegmark et al. (2004).*

where $\delta^{\rm D}$ denotes the Dirac delta function and for an isotropic universe the power spectrum depends on the scale amplitude only, $P(|\boldsymbol{k}|) = P(k)$. Inflationary models produce close to a simple power law form for the power spectrum, with $P_{\rm prim.}(k) \propto k^{n_{\rm s}}$, and $n_{\rm s} \approx 1$, however, we have seen that the growth of density perturbations is affected by a stifled growth during radiation domination at small scales, resulting in a power spectrum proportional to $k^{n_{\rm s}-4}$. This scale dependence is generally separated out from the growth factor, $D(z)$, which encompasses the growth of perturbations as a function of time, such that the power spectrum can be written as

$$P(k,z) = T^2(k)\frac{D^2(z)}{D^2(z_{\rm eq})}P_{\rm prim.}(k), \tag{2.63}$$

where $T(k)$ denotes the transfer function for the scale dependence of the evolution of the perturbations. In a Universe containing baryons and neutrinos, the transfer function can become complicated, however standard fitting formulae for these exist (Bardeen et al. 1986; Eisenstein & Hu 1997). Fig. 2.6 shows observations of the matter power spectrum and clearly indicates the expected turn-over at $k = k_{\rm eq}$, which is the scale of the horizon

at matter-radiation equality, where smaller modes are suppressed and larger modes can grow freely according to the primordial spectrum.

2.4 Non-linear structure formation

Astrophysical objects such as stars and galaxies are highly non-linear and cannot be described with the same tools as introduced in the previous section. Here we focus on describing spherical collapse of overdense regions that have become non-linear which leads to halo formation in the dark matter component and star formation when baryons are involved. A statistical description of the dark matter distribution can then be achieved using the excursion set formalism, and we introduce this in the context of the Press-Schechter formalism.

2.4.1 Spherical collapse

We first consider that for some time $t > t_{\rm nl}(\lambda)$, the scale λ has become non-linear and the linear result that modes evolve independently breaks down. To shine light on the non-linear evolution of a perturbation at such a scale, we consider the simplified situation of a spherically symmetric perturbation of constant mass. We follow the evolution of the density inside a mass shell of radius r_i at some initial time t_i. The density perturbation inside this shell is δ_i and the background density is $\bar{\rho}_i$. As small scales become non-linear first, we can assume that $\lambda \ll d_{\rm H}$, so the use of the Newtonian limit is appropriate. The dynamics of the mass shell is thus dictated by the gravitational potential and for a pressureless component, using the Friedmann equations, we find

$$\phi_{\rm tot}(r,t) = \frac{2\pi}{3} G\bar{\rho} r^2 + \delta\phi(r,t). \tag{2.64}$$

The motion of the shell is thus given by

$$\frac{d^2\boldsymbol{r}}{dt^2} = -\nabla\phi_{\rm tot} = \left(-\frac{G\bar{M}}{r^2} - \frac{G\delta\bar{M}}{r^2}\right)\hat{\boldsymbol{r}}, \tag{2.65}$$

which can be rewritten as

$$\frac{d^2 r}{dt^2} = -\frac{GM}{r^2}, \tag{2.66}$$

with

$$M = \frac{4\pi}{3}\bar{\rho} r_i^3 (1+\bar{\delta}_i) \quad \text{and} \quad \bar{\delta}_i = \frac{3}{4\pi r_i^3}\int_0^{r_i} \delta_i(r) 4\pi r^2 dr. \tag{2.67}$$

The integral, with respect to t, of the equation of motion yields

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 - \frac{GM}{r} = E, \tag{2.68}$$

and the constant of integration E is the specific energy of the shell. If $E > 0$, then $\dot{r} > 0$ as r increases, so the shell expands forever. However, when $E < 0$, as r increases, $\dot{r}$ will decrease to zero, eventually becoming negative, and the shell collapses. A parametric solution to the equation of motion in this case is

$$\begin{aligned} r &= A(1 - \cos\theta), \\ t &= B(\theta - \sin\theta), \\ A^3 &= GMB^2. \end{aligned} \tag{2.69}$$

From this we can see that the overdensity expands until it hits a radius of $r_{\rm ta} = 2A$ at time $t_{\rm ta} = B\pi$. After this turn-around, the overdensity starts to collapse to zero radius at time $t_{\rm coll} = 2t_{\rm ta}$. Considering the energy of the mass fluctuation at t_i and $t_{\rm ta}$, it can be shown that the turn-around radius solely depends on the initial density of the perturbation and not on the total mass perturbed. We can then determine what the critical density for a spherical region is to turn-around and collapse. By applying equation (2.69) to the expressions of the density enclosed in the spherical overdensity and the mean background density, the non-linear overdensity can be derived as

$$1 + \delta_{\rm nl} = \frac{\rho}{\bar{\rho}} = \frac{9(\theta - \sin\theta)^2}{2(1 - \cos\theta)^3}. \tag{2.70}$$

It is instructive to compare this to the linear result as linear theory can be fully described, while non-linear collapse can only be discussed in a simplified regime. So, for $\delta_i \ll 1$, using the appropriate Taylor expansion for the parametric solution, the linear density can be related to the initial perturbation in a flat, matter dominated Universe,

$$\delta_{\rm lin} = \delta_i \left(\frac{t}{t_i}\right)^{2/3} = \frac{3}{20}(6\pi)^{2/3}\left(\frac{t}{t_{\rm ta}}\right)^{2/3}. \tag{2.71}$$

The turn-around density can thus be found to be $\delta_{\rm nl} \approx 5.55$, at which point the linear theory predicts $\delta_{\rm lin} \approx 1.06$. At $t_{\rm coll}$, the linear density has grown to $\delta_{\rm lin} \approx 1.68$, but the spherical collapse model diverges. Objects do not collapse to zero radius, but instead particle interactions and angular momentum virialize the material and the spherical collapse result in the formation of virialized dark matter halos. Accounting for virial equilibrium in the spherical collapse model shows that after virialization, the non-linear density of the

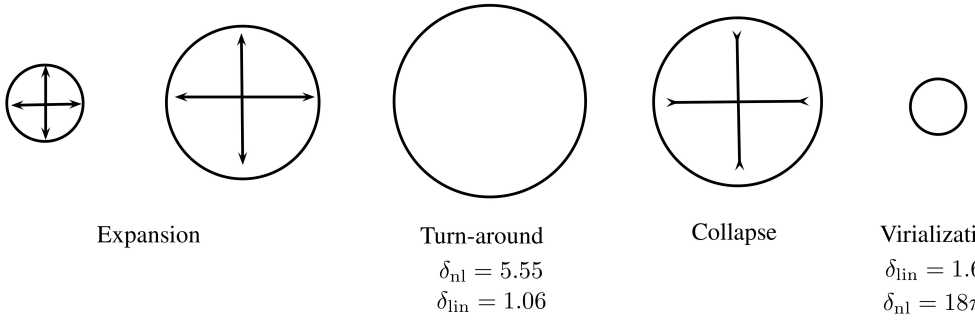

Figure 2.7: Summary of the spherical collapse model for non-linear growth. Density perturbations, once becoming non-linear, grow until the gravitational self-attraction exceeds the internal pressure and objects collapse and ultimately virialize into dark matter halos.

perturbation has risen to $\delta_{\rm nl} = 18\pi^2$, and objects stabalize at half their turn-around radius. We illustrate these results in Fig. 2.7.

2.4.2 Halo abundance

While the spherical collapse model determines the conditions for an individual density fluctuation to collapse and form a virialized dark matter halo, it says nothing about the statistical distribution of these objects. Determining the abundance of dark matter halos is a crucial step towards predicting the number and distribution of galaxies and galaxy clusters in the Universe. An important result from the previous section is that any region with a linear density of above $\delta_{\rm c} = 1.68$ should have collapsed and formed a virialized halo. Using this result and separating out the growth factor from the density field, the condition for halo formation becomes

$$\delta(\boldsymbol{x}) > \frac{\delta_{\rm c}}{D(t)} \equiv \delta_{\rm c}(t), \tag{2.72}$$

which can be seen as a static density field with a decreasing critical density boundary, where any region exceeding the boundary is part of a collapsed halo.

In order to characterize the mass of the dark matter halos formed, we consider the density field smoothed on some scale R, using a top-hat window function $W(\boldsymbol{x}; R)$ normalized such that $\int W(\boldsymbol{x}; R) d^3\boldsymbol{x} = 1$,

$$\delta(R) = \delta(\boldsymbol{x}; R) \equiv \int \delta(\boldsymbol{x}') W(\boldsymbol{x} - \boldsymbol{x}'; R) d^3\boldsymbol{x}'. \tag{2.73}$$

The convolution implies that density and window function are separable in Fourier space, and thus we can define the variance of the smoothed density field in terms of the linear

power spectrum,

$$\sigma^2(R) = \left\langle \delta^2(R) \right\rangle = \frac{1}{2\pi^2} \int P(k) \tilde{W}^2(kR) k^2 dk, \tag{2.74}$$

where $\tilde{W}$ denotes the Fourier transform of the window function. The filter scale R and mass M for a top-hat window function are related via

$$M = \frac{4\pi R^3}{3} \bar{\rho}, \tag{2.75}$$

such that we can interchange labels and discuss the halo abundance in terms of the mass variance $\sigma^2(M) = \sigma^2(R)$ of the smoothed density field.

2.4.2.1 Peak Statistics

The question to be answered is that given any region which satisfies equation (2.72) should be part of a collapsed dark matter halo, what is the mass associated with the halo, and what is the number density of halos as a function of mass, ie. the halo mass function. We can associate a mass with a collapsed object by considering the density field smoothed on mass scale M, $\delta_M = \delta(R)$, then, at all $\boldsymbol{x}$ where $\delta_M = \delta_c(t)$ a halo of mass M has formed. An initial consideration (Bardeen et al. 1986) is to directly relate the number density of halos with mass larger than M to the density of peaks in the smoothed density field which exceed the boundary $\delta_c(t)$. Although this approach seems intuitive, when peaks persist at multiple smoothing scales, this may lead to counting halos residing inside larger halos, thus overestimating the total number of condensed structures.

2.4.2.2 Press - Schechter formalism

An alternative approach was proposed by Press & Schechter (1974), who postulated an equivalence between the probability that $\delta_M > \delta_c(t)$ and the mass fraction contained in halos with mass larger than M at time t. For a Gaussian distributed density field, this probability takes the following analytic form

$$\mathcal{P}\left[\delta_M > \delta_c(t)\right] = \frac{1}{\sqrt{2\pi}\sigma(M)} \int_{\delta_c(t)}^{\infty} \exp\left[-\frac{\delta_M^2}{2\sigma^2(M)}\right] d\delta_M = \frac{1}{2}\mathrm{erfc}\left[\frac{\delta_c(t)}{\sqrt{2}\sigma(M)}\right]. \tag{2.76}$$

This result reveals that only half of all particles can at most contribute to structure formation, as $\lim_{M\to 0} \sigma(M) = \infty$ implies that $\mathcal{P}\left[\delta_{M\to 0} > \delta_c(t)\right] = 1/2$. This is unphysical as small underdense regions surrounded by larger overdensities should be part of larger collapsed objects. Press & Schechter (1974) solved this by introducing an *ad hoc* factor of

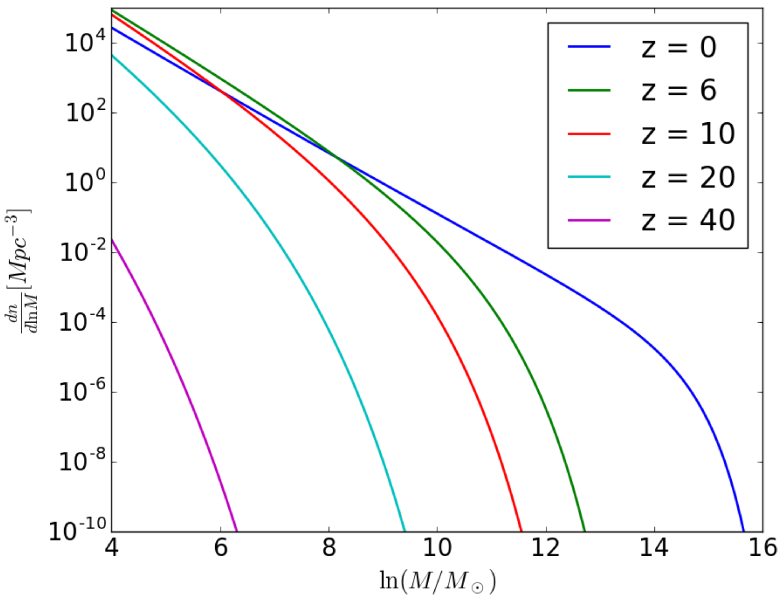

Figure 2.8: The Press-Schechter halo mass function at several redshifts: from right to left we have z = 0 (blue), z = 6 (green), z = 10 (red), z = 20 (teal), and z = 40 (pink).

2, resulting in

$$F(>M,t) = \mathrm{erfc}\left[\frac{\delta_c(t)}{\sqrt{2}\sigma(M)}\right]. \tag{2.77}$$

A natural explanation for this additional factor is achieved through the excursion set formalism and will be described in the following section.

The halo mass function $n(M,t)dM$ represents the number of halos with mass within a range $[M, M+dM]$ per comoving volume at time t and is thus obtained by the derivative of the mass fraction contained in halos of mass M or above per volume, such that

$$n(M,t)dM = \frac{\bar{\rho}}{M}\frac{\partial F(>M,t)}{\partial M}dM, \tag{2.78}$$

which is customarily written as

$$n(M,t)dM = \frac{\bar{\rho}}{M}f_{\mathrm{PS}}(\nu_c)\left|\frac{d\ln\nu_c}{d\ln M}\right|dM, \tag{2.79}$$

where $\nu_c \equiv \delta_c(t)/\sigma(M)$, and the multiplicity function is given by

$$f_{\mathrm{PS}}(\nu) = \sqrt{\frac{2}{\pi}}\nu e^{-\nu^2/2}. \tag{2.80}$$

The Press - Schechter mass function gives an analytic point of comparison for numerical simulations, and we show its form in Fig. 2.8.

2.4.2.3 Excursion set formalism

The occurrence of the additional factor of 2 in the Press - Schechter formalism can be explained using an excursion set formalism (Bond et al. 1991), and the combination is referred to as the Extended Press - Schechter formalism. We start by reformulating the Press - Schechter condition in terms of the variance of the density field at a particular smoothing mass M. In a hierarchical theory of structure formation, such as our CDM cosmology, $S \equiv \sigma^2(M)$ is a monotonically decreasing function in M. Considering the density field at a given point smoothed over decreasing mass scales with a sharp k-space filter, such as the top-hat used in section 2.4.2, results in a Markovian trajectory, as illustrated in Fig. 2.9. Although a sharp k-space filter implies that different mass scales are independent, which is unphysical, such a filter generates random walks as trajectories (eg. **?**). In this formulation a trajectory for which $\delta_S > \delta_c(t)$ indicates that the mass element resides in a dark matter halo with mass larger than M. The missing factor of 2 arises from not counting trajectories such as A in Fig. 2.9. At point S_2, trajectory A indicates that the corresponding mass element does not reside in a halo of mass $M > M_2$, however, when smoothed over a larger mass scale M_1, the same element is found to reside in a halo of mass $M > M_1 > M_2$. In order to account for the missed trajectories, trajectories, such as A', which exceed the boundary at S_2 need to be double counted. Due to the random nature of the trajectories, these trajectories occur with the same probability as those reflected by the boundary, and would normally be missed. This double counting then accounts for the missing factor in the Press-Schechter formalism.

A reworded ansatz for the excursion set formalism which results in the same mass function as the Press - Schechter formalism without the need of the unnatural factor of 2, is that the fraction of trajectories with first upcrossing of the collapse barrier $\delta_S = \delta_c(t)$ at $S > S_1$ is equal to the mass fraction condensed into halos with mass $M < M_1$ at time t.

2.4.3 Collapse of baryonic matter

Although the dark matter largely dictates the large scale distribution of structure in the Universe, the most interesting objects, such as stars and galaxies, are made out of baryons. Due to the gravitational force exerted on the baryonic gas by the underlying dark matter field, the baryons follow the dark matter distribution. As the density contrast between a gas cloud and the cosmic environment grows, the interplay between gravity and pressure determines the evolution of the gas cloud, and ultimately the formation of bound objects. Let us assume a spherical gas cloud which is starting to collapse under its own self gravity. The free-fall time can be defined as the characteristic time-scale over which this collapse would occur. The pressure resistance to this collapse can be expressed in terms of the

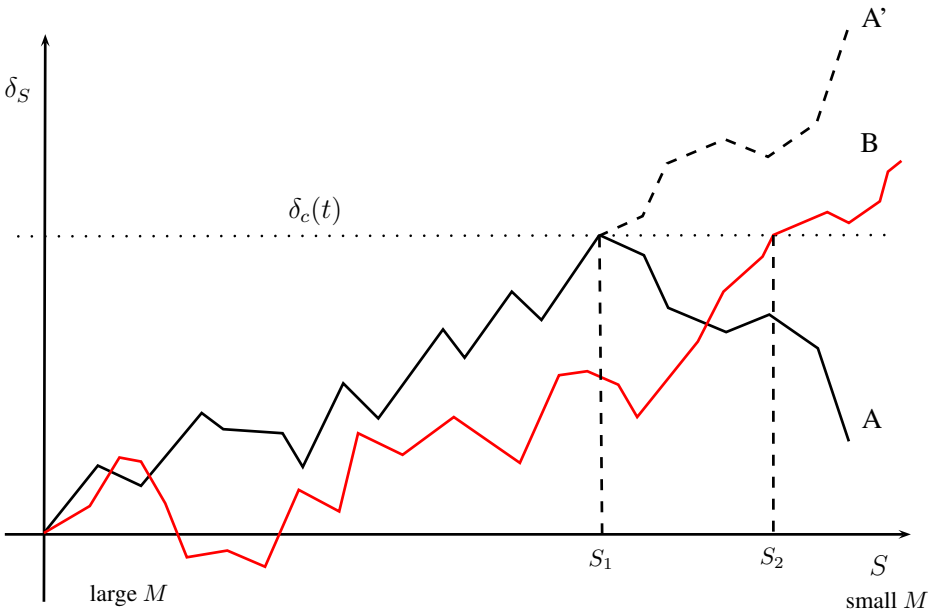

Figure 2.9: Example trajectories for the excursion set formalism as a function of smoothing scale. When computing the fraction of trajectories for which $\delta_S > \delta_c(t)$, trajectories exceeding the boundaries at large mass smoothing scales but not at lower mass smoothing scales are accounted for by double counting trajectories which mirror the trajectory at low mass scales.

time that a pressure wave requires to cross the diameter of the cloud, known as the sound crossing time. The ratio of the sound crossing time to the free-fall time then determines whether collapse occurs, in the case of a shorter free fall time, or whether the pressure in the gas is high enough to resist the collapse, when the free-fall time exceeds the sound crossing time. The acceleration felt by a particle on the boundary of a spherical gas cloud of mass M is

$$a = -\frac{GM}{R^2}. \tag{2.81}$$

From Newton's second law, the speed of the test particle can be derived as

$$|v| = \sqrt{\frac{8\pi G\rho R_0^2}{3}\left(\frac{R_0}{R} - 1\right)}, \tag{2.82}$$

where R_0 denotes the initial radius of the gas cloud and ρ its density. Using this result, the total collapse time is found to be

$$t_{\text{coll}} = \sqrt{\frac{3\pi}{32G\rho}}. \tag{2.83}$$

The sound crossing time is based on the sound speed through the cloud,

$$c_\mathrm{s} = \sqrt{\frac{\gamma k_\mathrm{B} T}{m}}, \tag{2.84}$$

where γ is the adiabatic index of the gas, k_B is the Boltzmann constant, T is the gas temperature, and m is the mass of the gas particles. The sound crossing time is then just given as the time required by a sound wave to traverse the gas cloud,

$$t_\mathrm{cross} = \frac{R}{c_\mathrm{s}} = \sqrt{\frac{m}{\gamma k_\mathrm{B} T}} R. \tag{2.85}$$

When $t_\mathrm{coll} > t_\mathrm{cross}$, the pressure in the gas resists the collapse. The turn-over, $t_\mathrm{cross}/t_\mathrm{coll} = 1$, defines the condition for collapse to occur, and the radius at which this occurs is defined as the Jeans length,

$$R_\mathrm{J} = c_\mathrm{s} \sqrt{\frac{3\pi}{32 G \rho}}. \tag{2.86}$$

We define the mass within a sphere of radius R_J as the Jeans mass,

$$M_\mathrm{J} = \frac{4\pi}{3} \rho R_\mathrm{J}^3. \tag{2.87}$$

This collapse conditions crucially depends on the gas temperature through the sound speed and thus as the temperature decreases, so does the Jeans mass and collapse of smaller objects becomes possible. For stars to form out of large gas clouds, those clouds thus need to fragment and cool such that the Jeans mass reduces to the mass scale of individual stars. Gas cooling in clouds inside the Milky Way is mainly controlled by the abundance of heavy elements. However, the first stars had to rely on atomic and molecular hydrogen in order to bring the Jeans mass down to enable star formation.

2.5 Inflation and primordial non-Gaussianities

In the previous sections, we have described the growth of structure using the concept of primordial density fluctuations sourced by quantum fluctuations in the density field. The mechanism by which these quantum fluctuations grow into macroscopic perturbations is that of a short period of exponential expansion in the early Universe, known as inflation. During this period, the Universe became shortly dominated by a negative pressure inflaton field, which induced the stretching of a Hubble size patch by about 60 e-foldings. Any initial curvature or inhomogeneities were inflated away which resulted in the flatness and homogeneity of our observable patch. A number of predictions about the nature

of primordial density fluctuations can be made from such an inflationary stage in the evolution of the Universe, and these can be tested against observations of the large scale structure. These observables give us a direct way to probe the detailed behaviour of the inflationary epoch as well as the properties of the inflaton field which sourced it.

One of the prime observables able to distinguish between different inflationary models is that of primordial non-Gaussianities (Maldacena 2003; Acquaviva et al. 2003; Bartolo et al. 2004; Weinberg 2005; Chen 2010; Komatsu 2010). Consider the temperature fluctuations, $\Delta T(\hat{\boldsymbol{n}}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{\boldsymbol{n}})$, measured on the CMB and assume statistical homogeneity and isotropy. If the temperature signal is perfectly Gaussian, then its probability density function (PDF) takes the following form,

$$P_{\rm g}(a) = \prod_{\ell m} \frac{e^{-|a_{\ell m}|^2/2C_\ell}}{\sqrt{2\pi C_\ell}}, \tag{2.88}$$

where C_ℓ denotes the angular temperature power spectrum and is defined through

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = C_\ell \delta^K_{\ell\ell'} \delta^K_{mm'}. \tag{2.89}$$

However, if non-Gaussianities in the temperature fluctuations exist, but the signal is very close to Gaussian, then a valid description is to Taylor expand the distribution around a Gaussian PDF (Komatsu 2010; Fergusson et al. 2012),

$$\begin{aligned} P(a) = P_{\rm g}(a) \times \Bigg\{ 1 + \frac{1}{6} \sum_{\text{all } \ell_i m_i} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \Big[(C^{-1}a)_{\ell_1 m_1} (C^{-1}a)_{\ell_2 m_2} (C^{-1}a)_{\ell_3 m_3} \\ - 3 (C^{-1})_{\ell_1 m_1, \ell_2 m_2} (C^{-1}a)_{\ell_3 m_3} \Big] \Bigg\}. \end{aligned} \tag{2.90}$$

The leading order deviation from a perfectly Gaussian field, which would be fully described by its power spectrum, can then be seen to be the angular bispectrum, $\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle$. Thus, any primordial non-Gaussianities introduced during inflation would be observable through the bispectrum, and more precisely through its amplitude. One of the important distinctions between the various inflationary models is their difference in predicted bispectrum amplitude and shape. The amplitude of the primordial bispectrum is often parametrized through the $f_{\rm NL}$ parameter, defined by,

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle = \mathcal{H}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \sum_i f^{(i)}_{\rm NL} b^{(i)}_{\ell_1 \ell_2 \ell_3}, \tag{2.91}$$

where the sum is taken over the different bispectrum models introduced shortly, $b^{(i)}_{\ell_1 \ell_2 \ell_3}$ is the reduced bispectrum, which defines the shape of the bispectrum, and $\mathcal{H}$ denotes the

Gaunt integral, enforcing the triangle condition and is defined through

$$\mathcal{H}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} = \int d^2\hat{\boldsymbol{n}} Y_{\ell_1 m_1}(\hat{\boldsymbol{n}}) Y_{\ell_2 m_2}(\hat{\boldsymbol{n}}) Y_{\ell_3 m_3}(\hat{\boldsymbol{n}}). \tag{2.92}$$

In spherical harmonic space this factor can be compared to the $\delta^D(\sum \boldsymbol{k}_i)$ arising when constructing the Fourier space bispectrum, and ensures that the coefficients form a closed triangle on the sphere. Inflation models separate into three categories in terms of their predictions for the angular bispectrum. The 'local form' bispectrum originates from the local perturbations in the curvature field and peaks in the squeezed triangle limit. Single field inflation models cannot give rise to a large local form bispectrum and thus any detection of $f_{\rm NL}^{\rm local} \gtrsim 1$ would be fatal for such models and a strong indication for multi-field inflation (Maldacena 2003; Acquaviva et al. 2003; Komatsu 2010). The 'equilateral form' bispectrum, which peaks in the equilateral triangle configuration, and the 'orthogonal form' bispectrum, which appears in a mixing of the other two, each arise from a variety of single-field inflation models, and would rule in or out a number of models if large values for $f_{\rm NL}^{\rm equi.}$ or $f_{\rm NL}^{\rm orth.}$ were to be detected. The best current constraints come from Planck Collaboration et al. (2016a) with measurements combining the temperature and polarization data to obtain $f_{\rm NL}^{\rm local} = 0.8 \pm 5.0$, $f_{\rm NL}^{\rm equi.} = -4 \pm 43$, and $f_{\rm NL}^{\rm orth.} = -26 \pm 21$ with 1σ errors. Due to the large errors on these measurements, they do not currently hold much constraining power on the various inflationary models. However, using new techniques, enabled by the 21cm line from neutral hydrogen, $\sigma_{f_{\rm NL}} \sim 0.03 - 1$ measurements could be achieved (Pillepich et al. 2007; Muñoz et al. 2015).

2.6 Gravitational lensing

One of the most famous predictions made by Einstein's theory of general relativity is the bending of light by massive objects (see Renn et al. 1997). The first observational evidence of the phenomenon was gathered during the 1919 solar eclipse where starlight passing close to the Sun was shown to bend due to its gravitational influence (Dyson et al. 1920; Stanley 2003). A first observation of extragalactic lensing was later made by Walsh et al. (1979) and thrusted the field into the forefront as a cosmological tool. Although gravitational lensing is visually most impressive in the strong regime, where image multiplication or even the formation of an Einstein ring can occur, cosmological lensing due to the accumulated effect of matter distorting photon paths along the line of sight presents a direct way to probe the statistical distribution of matter, and thereby probe the matter power spectrum and the cosmological parameters.

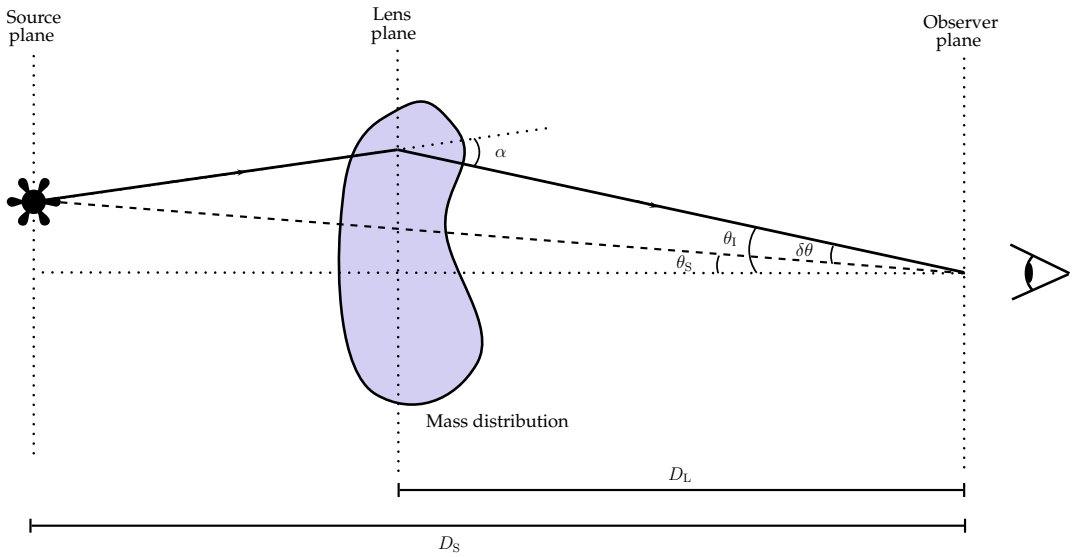

Figure 2.10: Sketch of gravitational lensing of a source on the source plane by matter on the lens plane observed on the observer plane.

Consider the situation depicted in Fig. 2.10, where the light emitted by an object on the source plane at a distance D_S from an observer is deflected by a mass at a distance D_L. The deflection caused by the mass distribution materializes in a small angular deviation, $\delta\theta$, to the real source location θ_S, such that an incident angle of $\theta_\mathrm{I} = \theta_\mathrm{S} + \delta\theta$ is measured by the observer. Considering a weak perturbation to the metric, one can derive the equation of motion for the transverse coordinates of a photon in a flat universe as (see Appendix A.1)

$$\frac{d^2\boldsymbol{x}}{d\eta^2} = -\frac{2}{c^2}\nabla\Phi, \tag{2.93}$$

where $\boldsymbol{x}$ denotes the comoving transverse coordinates of the photon, η is the conformal time, and Φ is the peculiar gravitational potential, related to the matter density field through the Poisson equation (2.35). A solution to the equation of motion can be obtained in the Born approximation, where photon paths are unperturbed and photons travel on null geodesics, and thus the metric yields $dr = -d\eta$. Using this result, equation (2.93) can be integrated twice, such that

$$x_i = r\theta_i - \frac{2}{c^2}\int_0^r dr'(r - r')\frac{\partial\Phi}{\partial x_i'}, \tag{2.94}$$

where the constant of integration θ_i has been set to be the incident angle of the photon. Considering the angular deviation of two nearby photons, we may write down the change of their relative positions after being affected by matter along their path of travel by Taylor

expanding the derivative of the gravitational potential,

$$\Delta x_i = r\Delta\theta_i(\delta^{\rm K}_{ij} - \phi_{ij}), \tag{2.95}$$

where $\delta^{\rm K}_{ij}$ is the Kronecker delta function and we have defined

$$\phi_{ij} \equiv \frac{2}{c^2}\int_0^r \frac{r-r'}{rr'}\frac{\partial^2\Phi}{\partial\theta_i\partial\theta_j}. \tag{2.96}$$

We can then write down the lensing potential which fully describes the weak lensing distortions induced by the gravitational potential,

$$\phi(r,\hat{\boldsymbol{n}}) = -\frac{2}{c^2}\int_0^r dr\frac{S_k(r-r')}{S_k(r)S_k(r')}\Phi(r',\hat{\boldsymbol{n}}), \tag{2.97}$$

where we have generalised to include non-flat universes. To see what the distortions in the image are, one writes down the ratio of image areas which is given by the Jacobian (Munshi et al. 2008),

$$\frac{\partial\boldsymbol{\theta}_{\rm S}}{\partial\boldsymbol{\theta}_{\rm I}} = (\delta^{\rm K}_{ij} - \phi_{ij})^{-1} = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}, \tag{2.98}$$

which defines the convergence κ and complex shear $\gamma = \gamma_1 + i\gamma_2$. The convergence

$$\kappa \equiv \frac{1}{2}(\phi_{11} + \phi_{22}), \tag{2.99}$$

describes the magnification of the image. The shear and its components

$$\gamma_1 = \frac{1}{2}(\phi_{11} - \phi_{22}) \quad \text{and} \quad \gamma_2 = \phi_{12}, \tag{2.100}$$

is related to the change in ellipticity of the galaxies which are weakly lensed,

$$e \simeq e_{\rm S} + 2\gamma, \tag{2.101}$$

where e and $e_{\rm S}$ are the observed and source ellipticities of the galaxies. The statistical correlation of the ellipticities of galaxies in a sample can thus be used as a probe for the shear correlation which is connected to the underlying cosmology.

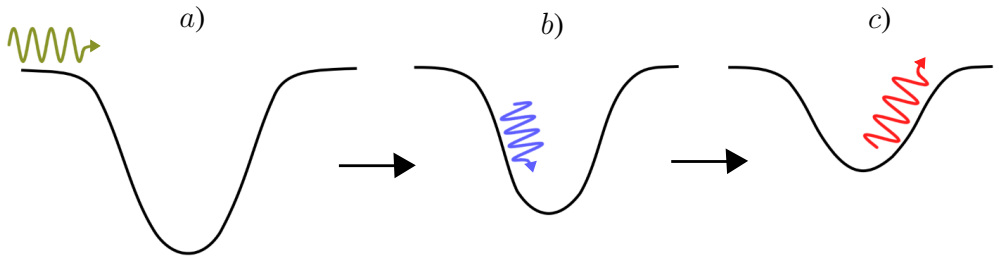

Figure 2.11: Sketch of the ISW effect. a) A photon enters the potential well of a galaxy cluster. b) As the photon is falling into the well, it gains energy and experiences a blue-shifting of its frequency. At the same time, the late time accelerated expansion of the Universe means that the potential decays as the photon is falling in. c) The photon leaving the potential well requires less energy than it gained upon entering, which results in an overall blue-shift of the photon's frequency.

2.7 The integrated Sachs-Wolfe effect

Another observable of the density perturbations in the Universe is the introduction of anisotropies in the CMB from structure along the light travel path and its effect on the gravitational potential that photons are travelling through. This effect was first predicted by Sachs & Wolfe (1967) and is known as the integrated Sachs-Wolfe (ISW) effect. As the potential is frozen during Matter domination, only a small amplitude contribution on small scales due to the non-linear collapse of structure affects the CMB photons (Rees & Sciama 1968). However, once the Universe transitions into the dark energy dominated regime, the potential decays as a^{-1} which gives rise of the late time ISW effect (Crittenden & Turok 1996). As the potential decays with time, photons will have to effectively climb out of a shallower potential than they fell into, which results in an overall frequency boost for the photon, as illustrated by Fig. 2.11.

To see the form this effect takes, one considers a small perturbation to the metric and solves the temporal part of the geodesic equation to first order (Sachs & Wolfe 1967; Martinez-Gonzalez et al. 1990; Dodelson 2003; Nishizawa 2014). The observed temperature fluctuations over the sky can then be found to be (Nishizawa 2014)

$$\frac{\delta T_0}{T}(\hat{\boldsymbol{n}}) = \frac{1}{4}\delta_\gamma(\tau_*) + \Phi(\tau_*) - \Phi_0 + \hat{\boldsymbol{n}}\cdot(\boldsymbol{v}_0 - \boldsymbol{v}) + \int_{\tau_*}^{\tau_0} d\tau(\Phi' + \Psi'). \tag{2.102}$$

Each term on the *rhs* represents a different physical effect which introduces anisotropies into the distribution of CMB photons. The first term represents the intrinsic temperature perturbation at recombination. The second and third term represent the Sachs-Wolfe effect from recombination, where, due to the perturbations in the density field, photons 'start' at different depths in the gravitational potential when the Universe becomes transparent.

The fourth term is the Doppler effect induced by the relative motion of the observer to the CMB. The final term represents the integrated Sachs-Wolfe effect, which is sourced by the time variation of the gravitational potential. In general relativity, we equate the two Newtonian potentials Φ and Ψ, such that the frequency shift, related to the temperature fluctuations, due to the ISW effect can be written as

$$\left.\frac{\Delta\nu}{\nu}\right|_{\mathrm{ISW}}(r,\hat{\boldsymbol{n}}) = \frac{2}{c^3}\int_0^r dr' \frac{\partial\Phi(r',\hat{\boldsymbol{n}})}{\partial t}. \tag{2.103}$$

As mentioned before, during most of the Universe's history, the gravitational potential is constant and thus the ISW effect vanishes. However, the potential changes during two important phases of cosmic evolution. At early times, when radiation dominates, an appreciable ISW effect is to be expected[4]. Then, a late time ISW contribution is evidence for a dark energy component, as such a Universe would experience a decay of the gravitational potential due to the accelerated expansion during dark energy domination.

2.8 Summary

The most precise measurements of the properties of our Universe suggest that we live in a geometrically flat Universe which largely consists of dark energy in the form of a cosmological constant with most of the gravitational matter being cold dark matter. The Universe in this model begins its expansion about 14 billion years ago in an inflationary period where space-time is stretched exponentially, and quantum fluctuations in the dark matter density field are amplified to a macroscopic level. After this time, the gravitational pull from matter in overdense regions causes the unrelenting growth of these density perturbations first in a linear, then a non-linear regime. The resulting structures formed by this process are virialized dark matter halos, which ultimately give rise to the stars and galaxies we observe today.

In this chapter, we have reviewed the basic cosmological model as well as structure formation in the linear regime and introduced non-linear collapse. The statistical distribution of dark matter is thus a predicted observable of the cosmological model and described by the matter power spectrum. Although the statistics of the matter density field can be gathered directly through the galaxy distribution, we discuss two other effects which couple to the density field and through which the parameters of the model are accessible: Weak lensing and the integrated Sachs-Wolfe effect.

[4] As $z_{\mathrm{eq}} \gg z_{\mathrm{rec}}$, this is usually seen as part of the primary temperature fluctuations at recombination.

Chapter 3

The Cosmic 21cm Signal

3.1 Observing neutral hydrogen

In the previous chapter, we have broadly summarized our theoretical understanding of the standard cosmological model with regard to structure formation. However, most of our theoretical understanding, especially at large scales, is based only on observations of the CMB, and although it has proven to contain a great amount of invaluable cosmological information, it largely only shows the conditions at the time of recombination, $z_{\rm rec} \approx 1100$. This deficiency has lead to a number of discrepancies between its predictions and low-redshift observations, which will require additional probes along the line of sight to remedy. For example, there remain tensions between local measurements of the Hubble parameter, h, and the value derived from *Planck* (Bennett et al. 2014; Riess et al. 2018a;b). Low redshift weak lensing surveys also indicate less matter clumping than expected from extrapolation of the CMB results (Heymans et al. 2013; MacCrann et al. 2015; Raveri 2016; Joudaki et al. 2017; Köhlinger et al. 2017).

Further, an important observational gap between the last scattering at $z_{\rm rec}$ and the formation of the first stars at $z \approx 30$ exists. Hence the crucial transition between linear and non-linear structure remains unobserved at present. The CMB photons also only weakly constrain the timing of the Epoch of Reionization (EoR), when high energy UV sources ionized the entire intergalactic medium (IGM), and give little to no information about the topology of the process. Observations of this high redshift regime represents the next frontier for observational cosmology and holds the potential to deepen our understanding of the transition from a dark into a light Universe.

The common thread throughout the evolution of the early Universe down to the present day is the ubiquitous presence of atomic hydrogen (HI). Direct observation of the hydrogen distribution throughout the history of the Universe represents a promising avenue to bridge the gap between the CMB and low redshift observations. Due to the hy-

perfine structure of atomic hydrogen, these atoms emit at a rest-frequency of 1420 MHz or a rest-wavelength of 21cm. The 21cm line from atomic hydrogen thus represents the only source of light which is continuously emitted from the formation of hydrogen at recombination to the present day. The potential for this line to deliver valuable cosmological information has been known for nearly 60 years (Field 1958; 1959; Sunyaev & Zeldovich 1972; Hogan & Rees 1979), however only recent technological advances have made an observational detection tangible. A new generation of radio observatories has thus reignited interest in this elusive hydrogen line and it is today widely regarded as the most promising avenue into the EoR and the high redshift Universe.

In this chapter we will introduce the theoretical background of the 21cm line in the context of cosmological observations. We begin in section 3.2 by defining the brightness temperature of the 21cm transition as a function of the relative abundance of atoms in the two hyperfine ground states. Section 3.3 details the various physical processes contributing to the signal and section 3.4 introduces the ionization history as an additional signal parameter. In section 3.5 we summarize the behaviour of the globally average signal as a function of time, before we focus on the fluctuations in the signal in section 3.6. Finally, in section 3.7 we introduce the various observational strategies to detect the cosmological 21cm signal.

3.2 The cosmic 21cm signal

The 21cm emission line originates from the hyperfine structure of the neutral hydrogen atom, which is due to the interaction of the magnetic moments of the proton and electron. Whereas the consideration of relativistic effects and spin-orbit coupling leads to the fine structure of the hydrogen atom[1], including the effects of the spin-spin coupling into the perturbed Hamiltonian splits the ground state into a low energy singlet and a high energy triplet state, lifting the spin degeneracy in the hydrogen atom. The energy difference between the ground state singlet and triplet states[2], $\Delta E_{10} = 5.9 \times 10^{-6}$eV, corresponds to a photon emission at a frequency of 1420 MHz, or equivalently a wavelength of 21cm (e.g. Griffiths 2005).

In order to talk about the cosmological significance of this result, one has to realize that transitions between the two hyperfine ground states are forbidden by quantum mechanical selection rules. This results in an extremely low spontaneous emission rate of

[1]This has the effect of breaking the degeneracy in the azimuthal quantum number.
[2]The subscripts 1 and 0 denote the triplet and singlet states respectively.

$A_{10} = 3 \times 10^{-15}\mathrm{s}^{-1}$ (Furlanetto et al. 2006), and thus a large half life of

$$\tau_{1/2} = A_{10}^{-1} = 3.5 \times 10^{14}\mathrm{s} \approx 11\ \mathrm{Myr}. \tag{3.1}$$

There are however three main arguments to be made that motivate the potential use of the 21cm transition in a cosmological context. Firstly, the energy difference of the hyperfine states is exceedingly small and corresponds to a temperature of

$$T_* = \frac{h\nu_{10}}{k_\mathrm{B}} \approx 0.0682\mathrm{K}. \tag{3.2}$$

The gas at any epoch in the cosmic history will therefore contain enough thermal energy to excite HI atoms into the triplet state. Secondly, given that most of the baryonic matter in the early Universe is HI, some fraction of it will be in the triplet state, and given that the half life $\tau_{1/2}$ is very much lower than the age of the Universe at all relevant epochs, spontaneous 21cm emission should not be rare. Thirdly, the optical depth of HI gas is small at frequencies associate with the 21cm line, leading to a large transmittance (Pritchard & Loeb 2012). Photons emitted at 1420 MHz can thus pass unhindered from their source into our telescopes, carrying information about the source and the intervening matter distribution.

The principal observable for radio telescopes, aiming to observe this redshifted 21cm line, is the brightness temperature of the signal on the sky, T_b. The brightness temperature corresponds to the temperature of a black body emitting at an observed intensity I_ν at a given frequency ν,

$$I_\nu \equiv B_\nu(T_\mathrm{b}) = \frac{2\nu^2 k_\mathrm{B} T_\mathrm{b}}{c^2}, \tag{3.3}$$

where the Rayleigh-Jeans law has been used as an approximation to the Planck law for black body radiation, as this is a good approximation for the relevant frequencies. The signal will be redshifted by the expansion of the Universe and one thus distinguishes between the brightness temperature at the source, T'_b, which depends on the rest frame frequency of the emitted photons, ν_0, and the observed brightness temperature, which depends on the redshifted frequency, $\nu = \nu_0/(1+z)$. The observed brightness temperature is then related to the rest-frame quantity via

$$T_\mathrm{b}(\nu) = \frac{T'_\mathrm{b}(\nu_0)}{1+z}. \tag{3.4}$$

Using this result and the radiative transfer equation for photons travelling through a cloud of neutral hydrogen, the following can be derived (e.g. Loeb & Furlanetto 2013; see

Appendix B.1)

$$T'_{\rm b}(\nu) = T_{\rm S}(1 - e^{-\tau_\nu}) + T'_{\rm R}(\nu)e^{-\tau_\nu}, \tag{3.5}$$

where the optical depth, $\tau_\nu \equiv \int \alpha_\nu ds$, is the integral of the absorption coefficient α_ν along the light ray through the cloud, and $T'_{\rm R}(\nu)$ is the brightness temperature of the background radiation field incident on the cloud along the ray. The brightness temperature depends also on the relative occupation number of the two hyperfine ground states, which is quantified by the spin temperature $T_{\rm S}$,

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{-T_*/T_{\rm S}}, \tag{3.6}$$

where $g_1 = 3$ and $g_0 = 1$ are the statistical weights of the energy levels. The most important application considered here is when the background radiation field is the CMB. In this case, using the fact that the optical depth is small at the relevant frequencies, allowing us to expand the exponentials in equation (3.5), the difference between the observed 21cm brightness temperature and the CMB temperature T_γ is

$$\delta T_{\rm b}(\nu) \approx \frac{T_{\rm S} - T_\gamma(z)}{1+z}\tau_{\nu_0} \tag{3.7a}$$

$$\begin{aligned} &\approx 27 x_{\rm HI}(1+\delta_{\rm b})\left(\frac{\Omega_{\rm b}h^2}{0.023}\right)\left(\frac{0.15}{\Omega_{\rm m}h^2}\frac{1+z}{10}\right)^{1/2} \\ &\quad \times \left(\frac{T_{\rm S} - T_\gamma}{T_{\rm S}}\right)\left[\frac{\partial_r v_r}{(1+z)H(z)}\right]^{-1} \text{mK}, \end{aligned} \tag{3.7b}$$

where $x_{\rm HI}$ denotes the neutral fraction of hydrogen, $\delta_{\rm b}$ is the fractional overdensity in baryons, and $\partial_r v_r$ is the velocity gradient along the line of sight. As seen from this result, the observability of the signal hinges on the spin temperature, only if it is different from the T_γ can the signal be seen against the CMB. It is therefore important to understand the physics that determine $T_{\rm S}$ as well as possible, in order to make credible predictions of the signal. Further, the signal depends on the ionization history of hydrogen gas, which makes it an ideal probe for reionization.

3.3 Contributions to the spin temperature

The spin temperature is determined by three main mechanisms:

(a) Absorption of CMB photons and stimulated emission,

(b) collisions with other HI atoms, free electrons, and protons,

(c) and Lyman-α / UV scattering involving intermediate excited states.

If these mechanisms are parametrized by the transition rates C_{10}, C_{01} for collisions and P_{10}, P_{01} for UV coupling, the spin temperature in equilibrium is determined by (Furlanetto et al. 2006),

$$n_1(C_{10} + P_{10} + A_{10} + B_{10}I_{\rm CMB}) = n_0(C_{01} + P_{01} + B_{01}I_{\rm CMB}). \tag{3.8}$$

This conservation equation says that in equilibrium the number of particles in the triplet state times the rate away from the triplet state is equal to the number in the singlet state times the rate toward the triplet state. B_{10} and B_{01} are Einstein coefficients and $I_{\rm CMB}$ the energy flux of CMB photons. The equation can be written as (see Appendix B.2)

$$T_{\rm S}^{-1} = \frac{T_\gamma^{-1} + x_{\rm c}T_{\rm K}^{-1} + x_\alpha T_{\rm C}^{-1}}{1 + x_{\rm c} + x_\alpha}, \tag{3.9}$$

where $x_{\rm c}$ and x_α are the coupling coefficients for collisional and Lyman-α coupling respectively, $T_{\rm K}$ is the gas temperature, and $T_{\rm C}$ is the colour temperature of the Lyman-α radiation field. The coupling coefficients determine whether collisions or Lyman-α photons contribute to the signal at a certain time and determine how strongly the spin temperature depends on these mechanisms to produce 21cm photons. Equation (3.9) then gives a straight forward relation defining the dependencies of the spin temperature. Being able to calculate the five parameters in the above relation gives the full evolution of the spin temperature and thus most of the theoretical evolution of the 21cm signal as seen in equation (3.7). We explore the physics leading to each of the contributing mechanisms in the following sections.

3.3.1 Collisional coupling

Collisions between particles can induce spin-flips between the two hyperfine ground states and thus directly effect the spin temperature at early epochs when the particle densities were high. There are three main channels of collisions present in the early Universe which all happen through *elastic* scattering, H-H (Zygelman 2005), H-e^-(Furlanetto & Furlanetto 2007) and H-p (Furlanetto & Furlanetto 2007). The coupling coefficient $x_{\rm c}$ can be written so as to include the three channels,

$$x_{\rm c} = x_{\rm c}^{HH} + x_{\rm c}^{eH} + x_{\rm c}^{pH} = \frac{T_*}{A_{10}T_\gamma}\left(C_{10}^{HH} + C_{10}^{eH} + C_{10}^{pH}\right). \tag{3.10}$$

Then, expressing the de-excitation rates C_{10}^i in terms of the rate coefficients for spin de-excitation in collision with species i, κ^i, and the number density n_i of the species, one

obtains

$$x_{\rm c} = \frac{T_*}{A_{10}T_\gamma}\left(n_H\kappa_{10}^{HH} + n_e\kappa_{10}^{eH} + n_p\kappa_{10}^{pH}\right). \tag{3.11}$$

This is to say that the coupling coefficient is defined by the rate at which the 21cm radiation is produced through stimulated emission by CMB photons, modulated by the number of particle collisions which enhance this emission as a function of the particle species colliding and their collision frequencies. The collision rates κ^i (in units of $\mathrm{cm^3 s^{-1}}$) are obtained by solving the Schrödinger equation for the relevant particle interactions. This leads to a common form for all three collision channels ,

$$\kappa_{10}^i = \sqrt{\frac{8k_{\rm B}T_{\rm K}}{\pi M_i}}\bar{\sigma}_i, \tag{3.12}$$

where the mean spin transition cross section is

$$\bar{\sigma}_i = \frac{1}{(k_{\rm B}T_{\rm K})^2}\int_0^\infty dE\sigma_i(E)Ee^{-\frac{E}{k_{\rm B}T_{\rm K}}}, \tag{3.13}$$

depending on the collision channel $i = HH, eH, pH$, the cross-section of the relevant scattering process as a function of energy, $\sigma_i(E)$, and the reduced mass M_i of the colliding two-particle system. The prefactor in equation (3.12) can be interpreted as the mean collision velocity for each channel. The collision rates for each case are then integrated numerically in Zygelman (2005); Furlanetto & Furlanetto (2007); Furlanetto & Furlanetto (2007) and we show the results in Fig. 3.1. Although the rate coefficients for the p and e channels are higher than that for H-H collisions, the total rates also include the number densities of the species, so that the effect of these collisions depends on the ionisation fraction $\bar{x}_{\rm i}$. Numerical integration of the full Boltzmann equation places the value of this parameter at around $\bar{x}_{\rm i} \sim 2\times10^{-4}$ during the dark ages (Dodelson 2003), meaning that for each free proton and electron pair there are around 5×10^3 hydrogen atoms. Therefore, H-H collisions dominate collisional coupling. H-e^- collisions are the next important channel and are dominating only when the Universe becomes sufficiently ionized. Proton collisions and other species collisions are even less important. Collisional coupling is however only dominant during the dark ages and all together dominated by the Lyman-α coupling as soon as the gas is heated up sufficiently such that Lyman-α photons are produced. The Lyman-α domination is caused by the fact that UV-photons scatter $\sim 10^5$ times before their energy is redshifted enough not to effect the spin temperature anymore and only about 10% of particle collisions actually lead to a spin flip (Furlanetto et al.

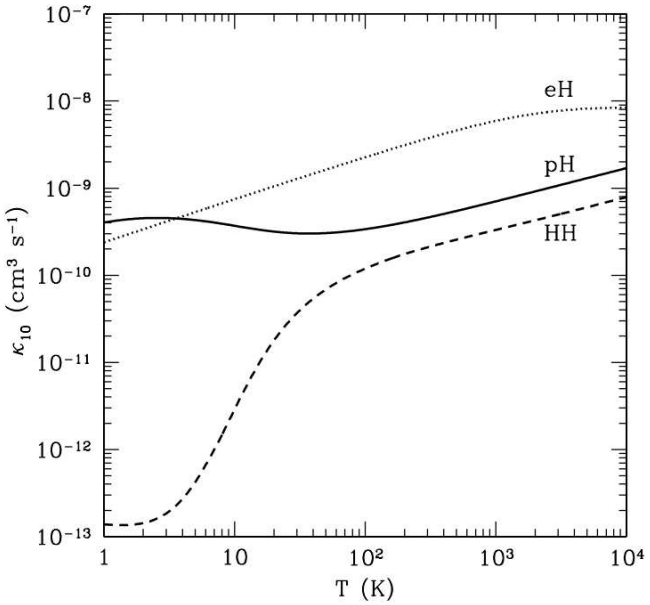

Figure 3.1: Evolution of the rate coefficients Rate coefficients for $p - H$ collisions (solid curve), $e - H$ collisions (dotted curve), and $H - H$ collisions (dashed curve). *Reproduced with permission from Fig 2. Furlanetto & Furlanetto (2007), by permission of Oxford University Press on behalf of the Royal Astronomical Society, available online at: https://academic.oup.com/mnras/article/379/1/130/1132800?searchresult=1. This figure is not included under the Creative Commons license of this publication. For permissions, please contact journals.permissions@oup.com.*

2006). An important consequence is that the high and low temperature regimes[3] in Fig. 3.1 are of little interest.

3.3.2 Lyman-α coupling, the Wouthuysen-Field effect

The Wouthuysen-Field effect describes the coupling of the spin temperature to Lyman-α, ie. UV, photons. The electric dipole selection rule allows for transitions only when $\Delta F = 0, \pm 1$, with the exception that there cannot be a transition from $F = 0$ to $F = 0$. So, upon absorbing a Lyman-α photon, a ground state hydrogen atom can be excited to either the $2\,{}_1P_{1/2}$ or the $2\,{}_1P_{3/2}$ state. From there, the aforementioned rule only allows the atom to de-excite down to the triplet state $1\,{}_1S_{1/2}$, thus resulting in a population mixing between the singlet and the triplet state. Excitations to the $2\,{}_0P_{1/2}$ and $2\,{}_2P_{3/2}$ are forbidden by the rule. The relevant transitions are shown in Fig. 3.2.

[3]Low temperatures are not reached by the IGM as heating starts when $z \sim 30$.

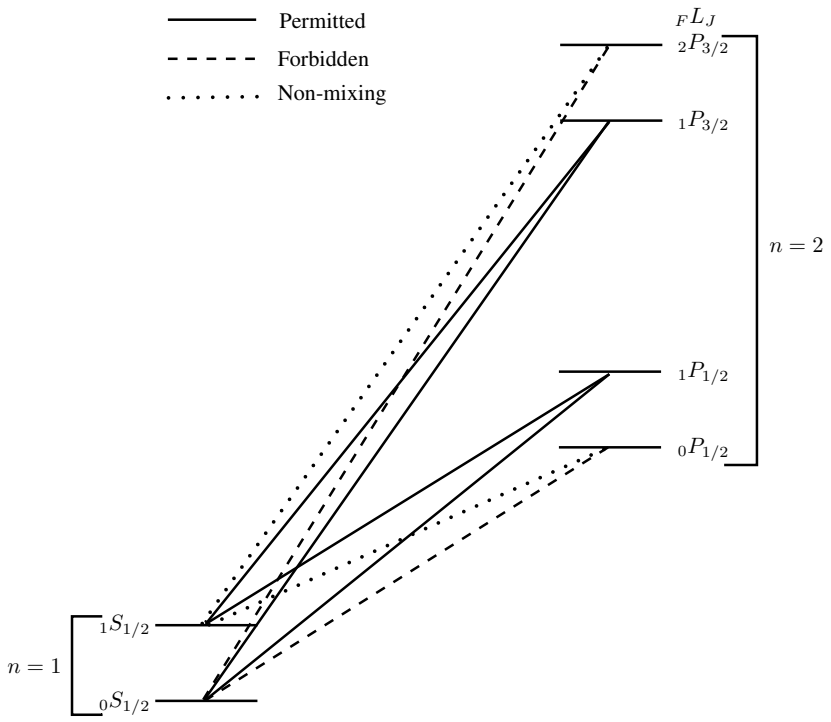

Figure 3.2: Relevant energy levels for the Wouthuysen-Field effect. The dashed lines are forbidden transitions and their Einstein coefficients can be taken to be zero. The dotted lines show allowed transitions that do not however mix the ground state singlet and triplet state. The solid lines show allowed transitions. *Adapted with permission from Field (1958).*

The coupling coefficient x_α depends on the total rate per atom at which Lyman-α photons are scattered within a gas,

$$P_\alpha = 4\pi\chi_\alpha \int d\nu J_\nu(\nu)\phi_\alpha(\nu), \tag{3.14}$$

where $\sigma_\nu = \chi_\alpha\phi_\alpha(\nu)$ is the local absorption cross section, $\chi_\alpha = \pi e^2 f_\alpha/m_e c$ is the oscillation strength of the Lyman-α transition, ϕ_α is the Lyman-α absorption profile, and J_ν is the angle-averaged specific intensity of the background radiation field (by number). It can be shown that $P_{10} = 4P_\alpha/27$ (e.g. Deguchi & Watson 1985; Meiksin 2000; see Appendix B.3), thus one can write the coupling coefficient in terms of this total scattering rate,

$$x_\alpha = \frac{4P_\alpha}{27A_{10}}\frac{T_*}{T_\gamma}. \tag{3.15}$$

Defining a suppression factor that describes the photon distribution close to the Lyman-α resonance,

$$S_\alpha = \int dx\phi_\alpha(x)J(x), \tag{3.16}$$

where x has been used to represent the relative frequency to the Lyman-α line centre. Combining equations (3.14) and (3.15), an expression for the coupling coefficient can be found,

$$x_\alpha = \frac{16\pi^2 T_* e^2 f_\alpha}{27A_{10}T_\gamma m_e c}S_\alpha J_\alpha = S_\alpha\frac{J_\alpha}{J_\alpha^{\rm C}}, \tag{3.17}$$

simply using the definition for χ_α, evaluating J_ν at the Lyman-α frequency to obtain J_α and setting $J_\alpha^{\rm C} = 1.165\times10^{-10}(1+z)/20{\rm cm}^{-2}{\rm s}^{-1}{\rm Hz}^{-1}{\rm sr}^{-1}$.

In order to determine the Lyman-α background, the Hubble flow across the resonance frequency as well as energy loss induced by atomic recoil have to be taken into account. As Lyman-α photons scatter off hydrogen atoms, some of the energy will be converted into kinetic energy as the atoms recoil, resulting in a net loss in energy of the re-emitted photon. This effect adds an additional drift term to the Hubble flow across the Lyman-α resonance. The Hubble flow and atomic recoil are both captured in the evolution parameter of the Lyman-α flux across the line centre, $J(x)$, in equation (3.16). Solving the radiative transfer equation for J_ν in an expanding universe, parametrised via γ, with an additional recoil parameter η gives the result

$$\delta_J(x) = 2\eta\int_0^\infty dy\exp\left[-2\eta y - 2\gamma\int_{x-y}^{x}\frac{dx'}{\phi_\alpha(x')}\right], \tag{3.18}$$

which is written in terms of $\delta_J \equiv (J_\infty - J)/J_\infty$. Details for this derivation are given in Furlanetto & Pritchard (2006), and numerical solutions for photons injected into the line

centre and for photons redshifting in from infinity are shown in Fig. 3.3. This determines

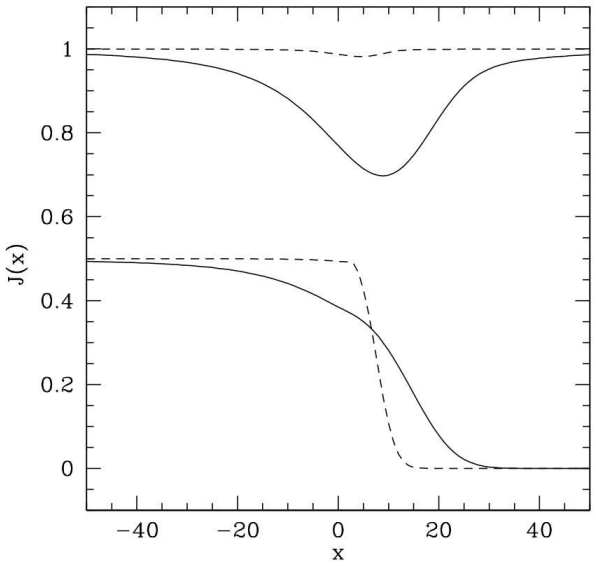

Figure 3.3: Background radiation field near the Lyman-α resonance at $z = 10$. Upper and lower panels are for photons redshifting in from infinity and for photons injected into the line centre respectively. The solid and dashed lines are for $T_{\rm K} = 10$K and 1000K respectively. *Reproduced with permission from Fig 1. Furlanetto & Pritchard (2006), by permission of Oxford University Press on behalf of the Royal Astronomical Society, available online at: https://academic.oup.com/mnras/article/372/3/1093/973192. This figure is not included under the Creative Commons license of this publication. For permissions, please contact journals.permissions@oup.com.*

the behaviour of the Lyman-α background near the resonance, and is naturally related to the suppression factor,

$$S_\alpha = \int_{-\infty}^{\infty} dx \phi_\alpha(x) J(x) \approx 1 - \delta_J(0) \leq 1. \tag{3.19}$$

A useful approximation for the correction factor, which is accurate to a few percent at all $T_{\rm K} \gtrsim 1$K, can be derived as (Furlanetto et al. 2006)

$$S_\alpha \sim e^{-0.803 T_{\rm K}^{-2/3} \left(\frac{10^{-6}}{\gamma}\right)^{1/3}}, \tag{3.20}$$

where $\gamma^{-1} = \tau_{\rm GP} \sim 3 \times 10^5 \bar{x}_{\rm HI} [(1+z)/7]^{3/2}$ is the mean Lyman-α optical depth experienced by a photon that redshifts across the entire resonance.

Further, a prediction of the detailed evolution of the Lyman-α flux is difficult as it depends strongly on the poorly constrained star formation history. Instead, an educated guess as to what the Lyman-α background looks like can be made. Photons contributing

to J_α are not inserted at line centre, because the photons redshift out of resonance very quickly after their creation and therefore only contribute to the coupling very close to their sources. Instead, the relevant photons are emitted at higher frequencies, in the UV, and redshift into resonance, or possibly redshift into Lyman-n resonance and cascade down to the Lyman-α transition as shown in Fig. 3.4. Lyman cascades ending in a Lyman-α transition are accounted for by the recycling fraction of the Lyman-n transition, $f_{\rm rec}(n)$. This factor is the probability that a Lyman-n photon will generate a Lyman-α photon and is obtained iteratively by computing the transition probabilities between different stages in the cascade,

$$f_{\rm rec}(n) = \sum_m P_{nm} f_{\rm rec}(m), \tag{3.21}$$

where

$$P_{mn} = \frac{A_{mn}}{\sum\limits_{n'} A_{mn'}}, \tag{3.22}$$

depending on the Einstein coefficients A_{mn} for spontaneous decays between initial state m and final state n. In the calculation of this factor it is assumed that direct Lyman-n decays into the ground state result into photons which are quickly reabsorbed and will not affect the population of excited states. This is achieved by setting $A_{n{\rm p}\to 1{\rm s}} = 0$. Values for the recycling fraction are tabulated in Pritchard & Furlanetto (2006), and show that $f_{\rm rec} \approx 0.36$ for large n. Therefore, the Wouthuysen-Field coupling to Lyman-n cascades is about a third as efficient as the coupling to Lyman-α photons directly. Two effects are thus taken into account when writing down an estimate of the Lyman-α background, cascades and emissivity of UV photons,

$$J_\alpha(z) = \sum_{n=2}^{n_{\rm max}} J_\alpha^{(n)}(z) \tag{3.23a}$$

$$= \frac{c}{4\pi} \sum_{n=2}^{n_{\rm max}} f_{\rm rec}(n) \int_z^{z_{\rm max}(n)} dz' \left|\frac{dt}{dz'}\right| \left(\frac{1+z}{1+z'}\right)^3 4\pi \frac{c}{H(z')} \epsilon(\nu'_n, z'). \tag{3.23b}$$

The emissivity, ϵ, depends on the star formation rate, and is also proportional to the rate at which matter collapses into galaxies. There are significant uncertainties in the prediction of these quantities, nevertheless one can write,

$$\epsilon(\nu, z) = f_* \frac{\rho_{\rm b}}{m_{\rm p}} N_{{\rm L}n}(\nu) \frac{df_{\rm coll}}{dt}, \tag{3.24}$$

where $N_{{\rm L}n}$ is the number of photons produced in the frequency interval $\nu \pm d\nu/2$ per baryon incorporated into stars, $\rho_{\rm b}$ is the baryon density and $m_{\rm p}$ the proton mass. Useful estimates for the values of N, assuming that stars are the dominant source of radiation

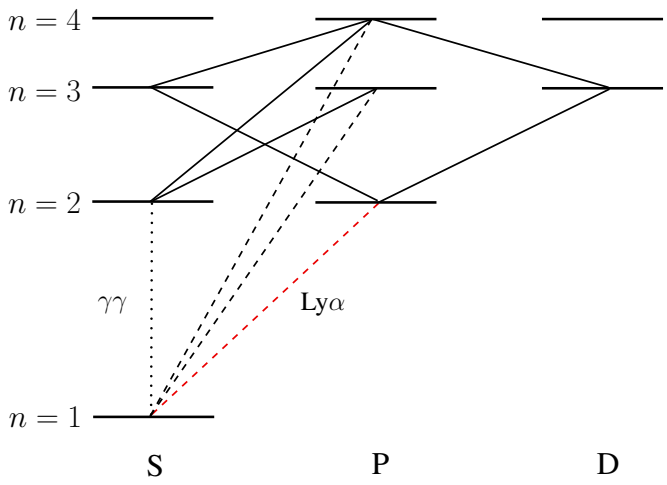

Figure 3.4: Ionizing photons redshift into Lyman-n resonance, and from there, they can cascade down via multiple possible decay chains (solid lines). Some decay chains lead to a Lyman-α emission, and thus contribute to the Lyman-α background. The figure shows decay chains for Lyman-β and Lyman-γ. Dashed lines denote Lyman-n transitions, the red-dashed line is the Lyman-α transition, and the dotted line shows the forbidden $2S \rightarrow 1S$ transition. *Reproduced with permission from Fig 2. Pritchard & Furlanetto (2006), by permission of Oxford University Press on behalf of the Royal Astronomical Society, available online at: https://academic.oup.com/mnras/article/367/3/1057/1039877. This figure is not included under the Creative Commons license of this publication. For permissions, please contact journals.permissions@oup.com.*

are the following: $N_\alpha = 9690$ for low-metallicity Population II stars and $N_\alpha = 4800$ for massive Population III stars (Barkana & Loeb 2005).

The Wouthuysen-Field effect couples the spin temperature to the colour temperature of the UV radiation field. T_C is a measure of the shape of the radiation field as a function of frequency, and defined by

$$\frac{h}{k_B T_C} = -\frac{d \log n_\nu}{d\nu}, \tag{3.25}$$

where $n_\nu = c^2 J_\nu / 2\nu^2$ is the photon occupation number. Given that in an optically thick medium Lyman-α photons scatter a large number of times, the shape of the Lyman-α profile will approach that of a black body at the gas temperature T_K (Wouthuysen 1952). Therefore $T_C \approx T_K$ for the frequencies of interest.

3.3.3 Thermal evolution of the IGM

The thermal evolution of the intergalactic medium (IGM), parametrized by the gas temperature T_K, is the next important ingredient to the spin temperature studied here. The IGM is modelled as follows (Furlanetto 2006),

$$\frac{dT_K}{dt} = -2H(z)T_K + \frac{2}{3}\sum_i \frac{\epsilon_i}{k_B n}. \tag{3.26}$$

The first term on the *rhs* is a cooling term originating from the Hubble expansion. The second term describes the heating of the IGM which is determined by the sum of the energies, ϵ_i, injected into the gas through mechanisms i. Various heating mechanisms play a roles at different epochs of the history of the Universe.

3.3.3.1 Heating of the IGM: Compton heating

At early times, before star formation has begun in earnest, the only effect heating the gas is Compton scattering between CMB photons and residual free electrons. Scattering increases the energy of the electrons which can then transfer the excess energy through collisions with other particles in the gas. The heating rate due to Compton scattering is calculated in Peebles (1993) and Seager et al. (1999) to be

$$\frac{2}{3}\frac{\epsilon_{\rm compton}}{k_B n} = \frac{\bar{x}_i}{1 + f_{\rm He} + \bar{x}_i}\frac{(T_\gamma - T_K)}{t_\gamma}, \tag{3.27}$$

where $t_\gamma \equiv 3m_e c/8\sigma_T u_\gamma$ is the Compton cooling time, $u_\gamma \propto T_\gamma^4$ is the energy density of the CMB, $f_{\rm He}$ is the Helium fraction (by number), $\bar{x}_i$ is the globally averaged ionization fraction, and $\sigma_T = 6.65 \times 10^{-25}{\rm cm}^2$ is the Thomson cross section. This effect is most im-

portant at high redshifts, when astrophysical sources are not yet prevalent in the Universe. Thus, early on, Compton heating couples the gas temperature to the CMB temperature and so the IGM cools at the same rate as the CMB, which is $T_{\rm K} \propto (1+z)$. At some later stage however, the IGM decouples from the CMB and thus cools down adiabatically with the expansion after that. The redshift at which this freeze-out occurs depends on the recombination rate and can be found to be $z_{\rm dec} \sim 200$.

3.3.3.2 Heating of the IGM: X-ray heating

As star formation begins, the Universe becomes filled with high intensity X-ray sources, such as Population III stars, stellar mass black holes, and quasars (Loeb & Furlanetto 2013). The X-ray photons emitted by those sources have very high mean free paths meaning that they can heat the IGM far away from the source thus ensuring homogeneous heating of the medium. The comoving mean free path of an X-ray photon is given in Furlanetto et al. (2006) to be

$$\lambda_{\rm X} \approx 4.9 \bar{x}_{\rm HI}^{1/3} \left(\frac{1+z}{15}\right)^{-2} \left(\frac{E}{300eV}\right)^{3} \text{Mpc}. \tag{3.28}$$

A conservative assumption for the poorly constrained X-ray background is that the correlation between the star formation rate, SFR, and the X-ray luminosity, $L_{\rm X}$, can be extrapolated to high redshifts, such that (Oh 2001),

$$L_{\rm X} = 3.4 \times 10^{40} f_{\rm X} \left(\frac{\rm SFR}{1 M_{\odot} \text{yr}^{-1}}\right) \text{erg s}^{-1}, \tag{3.29}$$

where $f_{\rm X}$ is an unknown renormalization factor. The determination of the redshift evolution of $f_{\rm X}$ hinges on our understanding of high redshift X-ray sources. In order to determine the energy input into the IGM, the mechanisms by which X-rays can heat the IGM need to be understood. The high energy X-ray photons ionize helium and hydrogen atoms, creating energetic free 'primary' electrons which heat the gas through three main channels;

- collisional ionization, producing more secondary electrons,
- collisional excitations of HeI and HI, which produce photons capable to ionize HI and a Lyman-α background respectively,
- Coulomb collisions with free electrons.

The fractions of the X-ray energies going into heating, ionization and excitation, parametrized by $f_{\rm X,h}$, $f_{\rm X,ion}$, and $f_{\rm X,coll}$ respectively, are determined exactly by the rela-

tive cross sections of these mechanisms but can be approximately related to the neutral fraction by (Chen & Kamionkowski 2004)

$$f_{\mathrm{X,h}} \sim \frac{1 + 2\bar{x}_{\mathrm{i}}}{3}, \tag{3.30a}$$

$$f_{\mathrm{X,ion}} \sim f_{\mathrm{X,coll}} \sim \frac{1 - \bar{x}_{\mathrm{i}}}{3}. \tag{3.30b}$$

Then, finally one can determine the energy input into the gas, by assuming that the SFR is proportional to the rate of gas collapse onto virialized halos (Furlanetto et al. 2006), df_{coll}/dt,

$$\frac{2}{3}\frac{\epsilon_{\mathrm{X}}}{k_{\mathrm{B}} n H(z)} = 10^3 \mathrm{K} f_{\mathrm{X}} \left(\frac{f_*}{0.1} \frac{f_{\mathrm{X,h}}}{0.2} \frac{df_{\mathrm{coll}}/dz}{0.01} \frac{1+z}{10} \right), \tag{3.31}$$

where f_* is the star formation efficiency. This result only takes the star formation rate into account and does not consider different X-ray sources such as quasars. These sources need to be studied in greater detail in order to determine the X-ray heating term to higher accuracy.

3.3.3.3 Heating of the IGM: Lyman-α heating

X-ray heating is expected to be the largest contributing heating mechanism, however a second order mechanism would be Lyman-α heating. The effect which leads to the suppression of the Lyman-α background, discussed in 3.3.2, deposits energy from atomic recoils into the IGM. The energy loss in the radiation background thus goes into heating the IGM. As before, continuum photons behave differently from photons injected at line centre. Whereas continuum photons heat the gas as expected via atomic recoil, at temperatures of interest here ($T_{\mathrm{K}} \gtrsim 4\mathrm{K}$), photons injected at line centre can cool the gas as more photons scatter on the red side of the line than on the blue. In this case, re-emitted photons have higher energies in the IGM frame, thus removing energy from it. The following total heating rate can be obtained as (Furlanetto & Pritchard 2006)

$$\frac{2}{3}\frac{\epsilon_\alpha}{k_{\mathrm{B}} n_{\mathrm{H}} H(z) T_{\mathrm{K}}} \approx \frac{0.8}{T_{\mathrm{K}}^{4/3}} \frac{x_\alpha}{S_\alpha} \left(\frac{10}{1+z} \right), \tag{3.32}$$

depending on the parameters discussed in section 3.3.2.

3.3.3.4 Heating of the IGM: Shock heating and dark matter heating

The effect of Lyman-α heating is already negligibly small compared to X-ray heating, so higher order effects are very often neglected in determining the gas temperature. However, effects like shock heating (Furlanetto et al. 2006) or dark matter heating (Evoli et al. 2014)

can influence the gas temperature and are being studied to gain further insight into the thermal evolution of the IGM.

3.4 Ionization history

The next step towards the evolution of the spin temperature and the brightness temperature is the ionization history, ie. the evolution of $\bar{x}_{\rm i}$. Most of the mechanisms driving the gas temperature depend on the ionization state of the Universe, as does the collisional coupling $x_{\rm c}$ as described in section 3.3.1. This starts varying away from not being neurtral around the time when X-rays start to heat up the gas (Oh 2001).

The usual assumption is that the production rate of ionizing photons is coupled to the star formation rate. An average ionizing efficiency is assigned to all galaxies, such that

$$\bar{x}_{\rm i} = \frac{\zeta f_{\rm coll}}{1 + \bar{n}_{\rm rec}}, \tag{3.33}$$

where $\bar{n}_{\rm rec}$ is the mean number of recombinations per ionized hydrogen atom and ζ is the ionisation efficiency given by

$$\zeta = A_{\rm He} f_* f_{\rm esc} N_{\rm ion}, \tag{3.34}$$

where $N_{\rm ion}$ denotes the mean number of ionizing photons produced per stellar baryon, and $A_{\rm He}$ is a correction factor to convert the number of ionizing photons per baryon in stars to the fraction of ionized hydrogen. At late stages of the ionization of the IGM, recombinations become important and must be taken into account as well. This generates a term that opposes ionization in the form of

$$\left(\frac{d\bar{x}_{\rm i}}{dt}\right)_{\rm rec} = -\alpha C(z, \bar{x}_{\rm i}) \bar{x}_{\rm i}(z) n_e(z), \tag{3.35}$$

where α is the recombination coefficient, $C \equiv \left\langle n_e^2 \right\rangle / \langle n_e \rangle^2$ is the clumping factor, and n_e is the average electron density in ionized regions. This gives the overall evolution of the neutral fraction as

$$\frac{d\bar{x}_{\rm i}}{dt} = \zeta(z) \frac{df_{\rm coll}}{dt} - \alpha C(z, \bar{x}_{\rm i}) \bar{x}_{\rm i}(z) n_e(z). \tag{3.36}$$

The terms in the above equation are complicated and depend on parameters which are poorly constrained by observations. A detailed analysis of the reionization history is done by Pritchard & Loeb (2012) and Furlanetto et al. (2006).

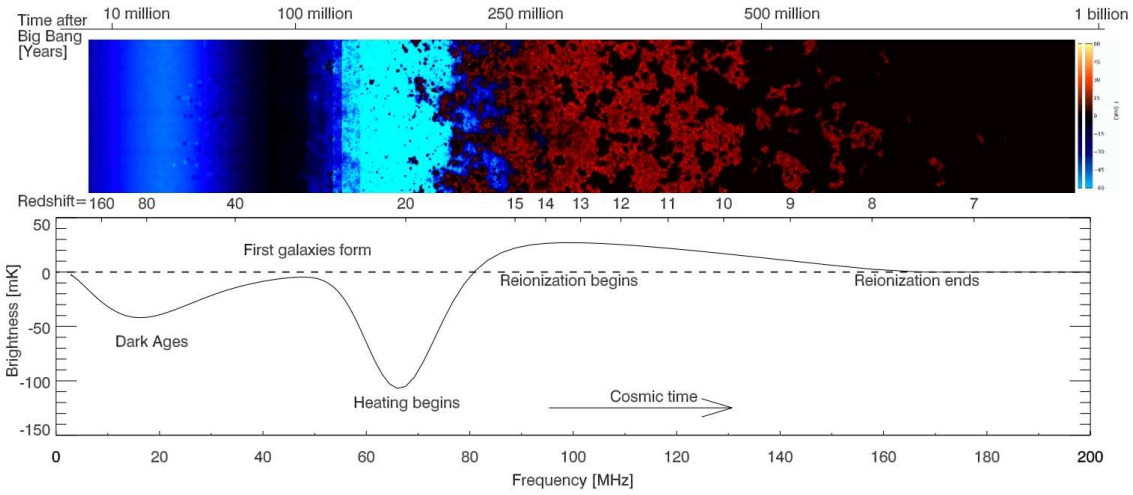

Figure 3.5: Time evolution of the global 21cm signal. The *upper* panel shows the 21cm brightness pieced together from redshift slices through a cosmic volume, coloured according to the signal strength. At early times, two absorption phases can be seen (purple and blue) separated by a (black) period where no signal is observed. When reionization begins, the signal transitions into emission (red) before disappearing (black) as the Universe fully ionizes. The *lower* panel shows the sky-averaged 21cm brightness temperature expected from this sample history. *Taken with permission from Pritchard & Loeb (2010).*

3.5 Global signal

We have introduced the physics determining the evolution of the spin temperature, gas temperature, ionization history, and thus the brightness temperature in the previous sections and a prediction of the global evolution of the 21cm signal can be made. The prediction of the 21cm signal is model dependent as large uncertainties on the necessary parameters exist. Nevertheless, useful information can be extracted and fundamentally separate regimes can be identified by using conservative model parameters. One example history is plotted in Fig. 3.5. Following Pritchard & Loeb (2012) and Mesinger et al. (2011), the important regimes are believed to be the following.

- $1100 \gtrsim z \gtrsim 200$: After recombination, high particle densities collisionally couple the spin and gas temperature to the CMB temperature, setting $T_S = T_K = T_{CMB} \propto (1 + z)$. No 21cm signal is expected, $\bar{T}_b = 0$.

- $200 \gtrsim z \gtrsim 40$: The Compton scattering rate eventually falls below the expansion rate, such that the gas decouples from the CMB, leading to adiabatic cooling of the gas, $T_K \propto (1 + z)^2$. At this point the spin temperature is only affected by collisional

coupling to the gas, it follows the gas temperature and the 21cm signal can first be seen in absorption against the CMB.

- $40 \gtrsim z \gtrsim z_*$: Eventually, as the Universe continues to expand, the gas density decreases so as to make collisional coupling more and more ineffective. The spin temperature couples again to the CMB temperature and the brightness temperature approaches 0 once more. Little to no 21cm signal is to be expected.

- $z_* \gtrsim z \gtrsim z_\alpha$: Star formation kicks in at around $z_* \sim 30$. These first sources emit Lyman-α as well as X-ray radiation, heating the gas significantly. The Wouthuysen-Field effect now couples the spin temperature back to the gas, and depending on how quickly the gas is heated above the CMB, this may lead to an absorption signal, $\bar{T}_\mathrm{b} < 0$, as it does in Fig. 3.5.

- $z_\alpha \gtrsim z \gtrsim z_\mathrm{h}$: As more star formation occurs, the Lyman-α coupling will eventually saturate at a redshift of z_α, when $x_\alpha \gg 1$. Fluctuations in the Lyman-α flux no longer affects the signal, instead, the brightness temperature fluctuations are sourced by the gas temperature fluctuations. The gas is heated via Lyman-α , X-rays and other mechanisms to eventually reach the CMB temperature at a redshift of z_h. During this time most of the 21cm signal is still seen in absorption against the CMB in conservative histories.

- $z_\mathrm{h} \gtrsim z \gtrsim z_\mathrm{T}$: After the gas is heated above the CMB temperature, $T_\mathrm{K} > T_\mathrm{CMB}$, the 21cm signal can be seen in emission. By this point, the ionization fraction of the Universe has likely risen to the 1% level, so that the 21cm signal is sourced by a mixture of fluctuations in the ionization fraction, density and gas temperature.

- $z_\mathrm{T} \gtrsim z \gtrsim z_\mathrm{r}$: Heating continues as more sources turn on and drive the gas temperature far above the CMB temperature, $T_\mathrm{K} \gg T_\mathrm{CMB}$. At some redshift z_T, the fluctuations in the gas temperature become unimportant, $T_\mathrm{S} \sim T_\mathrm{K} \gg T_\mathrm{CMB}$, and the dependence on the spin temperature in equation (3.7) can be ignored. Fluctuations in the ionization fraction dominate the signal at this stage.

- $z_\mathrm{r} \gtrsim z$: After reionization, the IGM is fully ionized and nothing contributes to a large scale 21cm signal anymore. Any residual 21cm signal would have to originate in self-shielded islands of neutral hydrogen, so-called damped Lyman-α systems.

The exact value of the redshifts stated above are model dependent as is their order. All this depends on the physics of the first sources, which is still insufficiently well understood. Measuring the global 21cm signal could thus constrain a large set of parameters.

3.6 The fluctuating 21cm sky

Whereas up to this point the discussion was mainly focussed on the sky-averaged signal, thus producing the lower panel of Fig. 3.5, a change of gear is in order for two reasons. The first reason is that sensitivity of upcoming radio experiments is not high enough to probe this smoothly varying monopole term (Lidz et al. 2008) and possibly space-based observations are needed (Burns et al. 2012). Foregrounds play a significant role here, as they too are predicted to be smoothly varying (Chapman et al. 2014b), thus requiring very high sensitivity to distinguish the sky-averaged signal from it. The second reason is that each component going into the 21cm signal fluctuates, thus making the 21cm signal a statistical quantity, as illustrated by upper panel in Fig. 3.5.

3.6.1 The 21cm power spectrum

A Fourier space power spectrum can be defined via

$$\left\langle \tilde{\delta}_{21}(\mathbf{k}_1)\tilde{\delta}^*_{21}(\mathbf{k}_2)\right\rangle \equiv (2\pi)^3\delta^{\mathrm{D}}(\mathbf{k}_1-\mathbf{k}_2)P_{21}(\mathbf{k}_1), \tag{3.37}$$

where $\tilde{\delta}_{21}$ is the Fourier transform of the brightness temperature fluctuation about the mean value. The fluctuation in the brightness temperature depends on the fluctuations of the parameters that determine it, so

$$\delta_{21} = \beta_{\mathrm{b}}\delta_{\mathrm{b}} + \beta_{\mathrm{x}}\delta_{\mathrm{x}} + \beta_{\alpha}\delta_{\alpha} + \beta_{\mathrm{T}}\delta_{\mathrm{T}} - \delta_{\partial v}, \tag{3.38}$$

where δ_i is the fractional variation in the particular quantity: δ_{b} for the baryon density, δ_{x} for the neutral fraction, δ_{α} for the Lyman-α coupling, δ_{T} for the gas temperature, and $\delta_{\partial v}$ for the peculiar velocities along the line of sight. The β coefficients are

$$\beta_{\mathrm{b}} = 1 + \frac{x_{\mathrm{c}}}{x_{\mathrm{tot}}(1+x_{\mathrm{tot}})}, \tag{3.39a}$$

$$\beta_{\mathrm{x}} = 1 + \frac{x_{\mathrm{c}}^{\mathrm{HH}} - x_{\mathrm{c}}^{e\mathrm{H}}}{x_{\mathrm{tot}}(1+x_{\mathrm{tot}})}, \tag{3.39b}$$

$$\beta_{\alpha} = \frac{x_{\alpha}}{x_{\mathrm{tot}}(1+x_{\mathrm{tot}})}, \tag{3.39c}$$

$$\beta_{\mathrm{T}} = \frac{T_{\gamma}}{T_{\mathrm{K}}-T_{\gamma}} + \frac{1}{x_{\mathrm{tot}}(1+x_{\mathrm{tot}})}\left(x_{\mathrm{c}}^{e\mathrm{H}}\frac{d\ln\kappa_{10}^{e\mathrm{H}}}{d\ln T_{\mathrm{K}}} + x_{\mathrm{c}}^{\mathrm{HH}}\frac{d\ln\kappa_{10}^{\mathrm{HH}}}{d\ln T_{\mathrm{K}}}\right), \tag{3.39d}$$

where $x_{\mathrm{tot}} = x_{\mathrm{c}} + x_{\alpha}$. The Fourier transform of δ_{21} can be written down analogous to (3.38), such that the 21cm power spectrum contains terms $P_{\delta_i\delta_j}$ for each combination

of i and j. As the β terms can be large[4], second order terms are not necessarily small. Further, all of the components contributing to the 21cm power are isotropic apart from the peculiar velocity contribution, which introduces an angular dependence to the signal known as the Kaiser effect (Kaiser 1987). This further separates the 21cm power spectrum into isotropic and anisotropic contributions and so one can write (Pritchard & Loeb 2008; Loeb & Furlanetto 2013)

$$P_{21}(k,\mu) = P_{\mu^0}(k) + \mu^2 P_{\mu^2}(k) + \mu^4 P_{\mu^4}(k) + P_{f(k,\mu)}(k,\mu), \tag{3.40}$$

where μ is the cosine of the angle between the wave-vector and the line of sight. This is useful because it separates out some of the fluctuations into different terms, eg. $P_{\mu^4}(k)$ is a function of the density field only. By measuring different components of the 21cm power, one can thus isolate the various contributing factors. Moreover, as some contributions to the 21cm signal saturate or are negligible during the cosmic history, measuring its power at different redshifts allows the measurement of different components to the signal. So for example, at early times fluctuations in the ionization fraction may be negligible, rendering terms including δ_x zero.

3.6.2 Power spectrum estimations

The evolution of the 21cm power spectrum is inherently linked to the state of matter in the Universe, making it a prime probe for the Epoch of Reionization (EoR) at redshifts $z \sim 6 - 15$. The shape and magnitude of the power spectrum can inform us on how reionization proceeded, how ionized bubbles grew (Watkinson & Pritchard 2014), what the dominating ionization mechanisms are (Mesinger et al. 2013), and when it started and ended (Pritchard & Loeb 2008; Mesinger et al. 2013). As observations of the 21cm signal are likely to still be a few years away, numerical simulations are necessary to compare models and predict the signal. Detailed radiative transfer simulations are computationally expensive and fast semi-numerical analyses tend to introduce errors at a level lower than current astrophysical uncertainties (Zahn et al. 2011). 21cm power spectrum estimations currently rely strongly on such semi-numerical techniques (Mesinger et al. 2011). An example of such a simulation is shown in Fig. 3.6 leading into the EoR. The left panel shows an example for what a direct imaging observation of the 21cm signal at $z = 10 - 30$ might look like and depicts the impact that some of the crucial evolutionary stages in the early Universe have on the intergalactic medium. Lyman-α photons emitted by first generation stars, starting at around $z \sim 30$, couple the spin temperature to the gas

[4]Note that β_b and β_x are of order unity.

temperature leading to a strong absorption feature, and thus $\delta T_b < 0$. Once X-rays heat the gas, the signal transitions into emission, shown by the $\delta T_b > 0$ structure in the $z = 20$ simulation here. The spin temperature saturates around $z \sim 18$ in this model as seen by the homogenisation of the signal in emission. Finally, the fluctuations are dominated by reionization and the holes in the signal it creates. The left panel thus indicates the fluctuating nature of the signal field and need for its statistical characterisation through the power spectrum. The right panel then shows the accompanying signature of the power spectrum at each epoch. The results of radiative transfer simulations are shown on the left in Fig. 3.7. The left panel here showing the evolution of the power spectrum during the EoR and the right panel plotting the power on particular scales through most of the cosmic history up to the dark ages. The 3D nature of the signal complicates the illustrations such that either only particular scales or particular redshifts can be shown.

Various features, can be identified in the power spectra plotted in Fig. 3.6 and Fig. 3.7. These features include;

- The rise and fall of the amplitude of the power spectrum before the EoR. Lyman-α emissions couple the signal to the gas temperature. Most of the gas is still in absorption ($T_S < T_\gamma$), however, regions close to the first X-ray emitters are rapidly heated into emission ($T_S \gg T_\gamma$), resulting in a large fluctuation amplitude with strong contrasts between emitting and absorbing regions, generating a large power spectrum amplitude. Once the gas is heated everywhere and the signal decouples from the gas (and before reionization has started in earnest), only the density field contributes to the signal, decreasing fluctuation amplitudes and thus the power.

- The flattening of the spectrum during the EoR. As reionization proceeds, ionized bubbles of hydrogen which cannot emit a 21cm signal anymore grow. Eventual overlap of these bubbles necessarily decreases the power on large scales, thus flattening the spectrum.

- The dramatic drop in power at large scales and persistent power at low scales. At the end of the EoR, almost all of the hydrogen in the Universe has been ionized except for pockets of self-shielded HI regions. The only left over 21cm signal is generated in those pockets, so only power on the smaller scales survives.

3.7 Observing the 21cm line

The 21cm line holds the potential to follow the formation of the large scale cosmological structures back to the earliest moments after the recombination of hydrogen atoms at

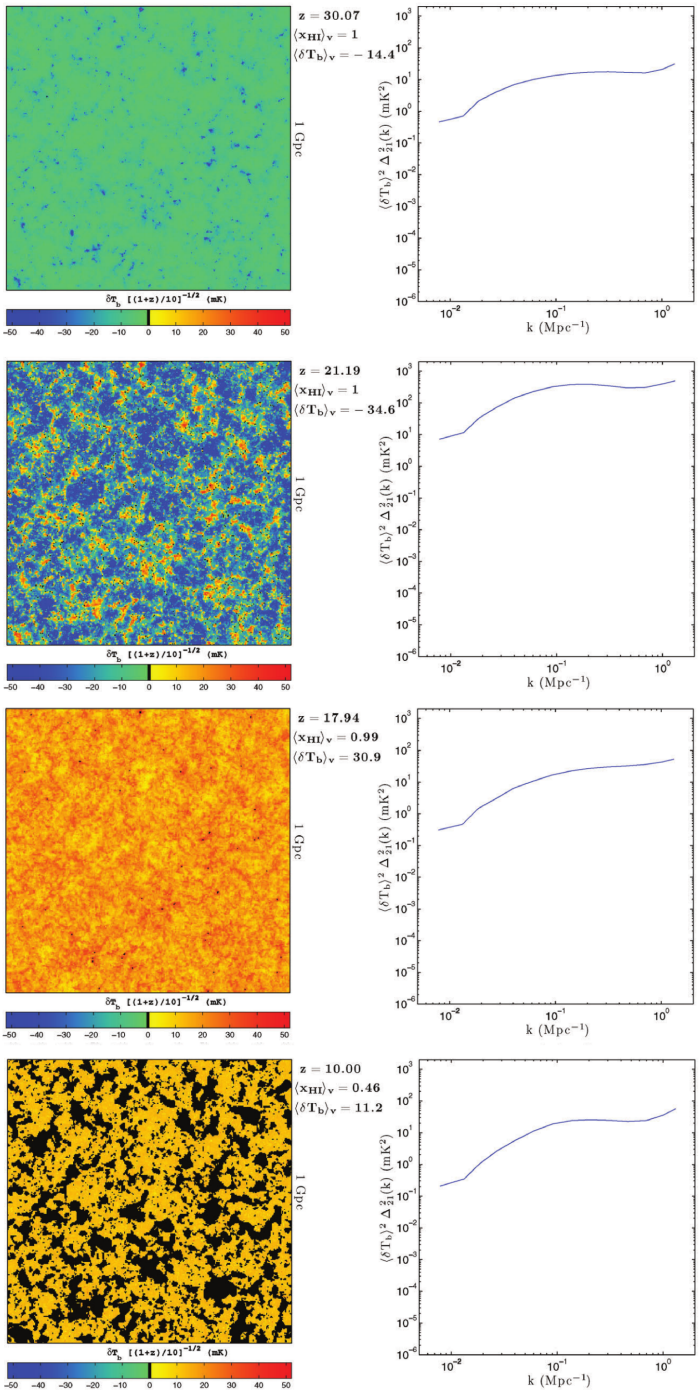

Figure 3.6: The left column shows slices through a $\delta T_{\rm b}$ simulation box of 1Gpc per side with a depth of 3.3 Mpc. The colouration indicates the signal brightness. The corresponding 21cm spherically averaged power spectrum is plotted in the right column. Four redshifts are shown, $z = 30.07, 21.19, 17.94$ and 10.00 from top to bottom. These redshifts are chosen to highlight various epochs in the 21cm signal (Compare with Fig. 3.5): the Cosmic Dawn and with it the start of Lyman-α coupling, the start of X-ray heating, saturation of the spin temperature and decoupling from the gas temperature, and the midpoint of reionization (black regions indicate ionized HII bubbles) are shown from top to bottom. *Reproduced with permission from Fig 12. Mesinger et al. (2011), by permission of Oxford University Press on behalf of the Royal Astronomical Society, available online at: https://academic.oup.com/mnras/article/411/2/955/1273635. This figure is not included under the Creative Commons license of this publication. For permissions, please contact journals.permissions@oup.com.*

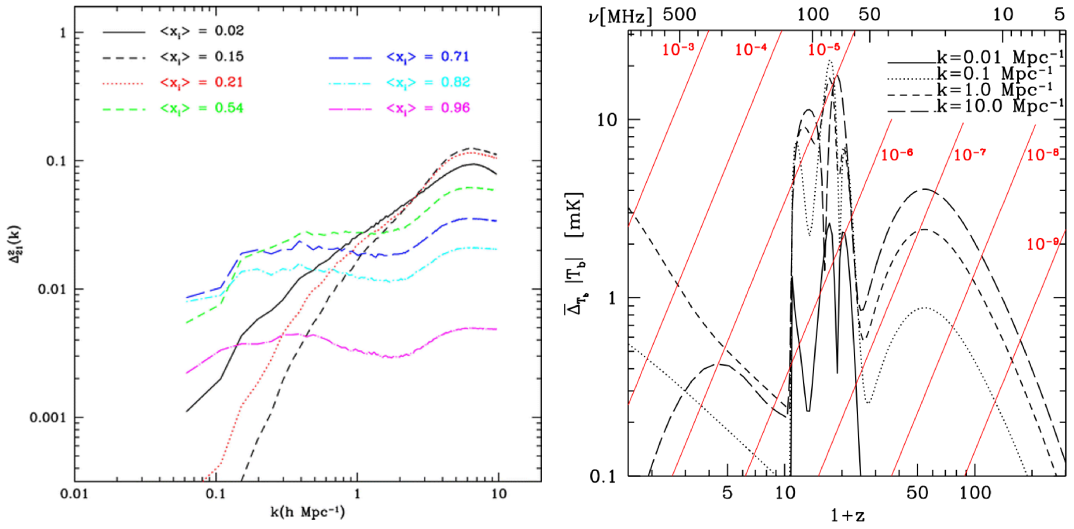

Figure 3.7: On the *left*, the evolution of the dimensionless spin-flip background throughout the EoR is shown for varying average ionization fractions. To get the 21cm signal, one needs to multiply this quantity by the mean brightness temperature in a fully ionized medium. Keeping the ionization fraction fixed and changing the redshift does not significantly change this power, such that the ionization fraction is used as a redshift proxy here. On the *right*, the redshift evolution of the 21cm power-spectrum, of a model where reionization ends at $z \sim 10$, through most of the cosmic history is plotted for a number of different scales. The diagonal lines show contours of a fixed fraction of the sky brightness as a function of frequency. *The left panel is taken with permission from Lidz et al. (2008), the right panel is taken with permission from Pritchard & Loeb (2008).*

$z \sim 1100$. The prospect of following the linear evolution of structure through the EoR until today has made this line a popular target for current and future experiments. As the expansion of the Universe elongates the emission wavelength of the photons as they travel toward us, the observational efforts to detect this signal operate at radio frequencies. However, a number challenges need to be overcome before a detection is possible. Firstly, the frequency range targeted by many EoR and Cosmic Dawn experiments is heavily used by radio and TV stations as well as other communication channels, resulting in a significant component of radio frequency interference (RFI), which needs to be understood, removed, or avoided for a first detection. Secondly, ionospheric dynamics affects incoming photons and requires new calibration techniques in order to avoid diffractive delays and image distortions. Thirdly, the 21cm line is not the only signal at these low frequencies, as the galactic synchrotron emission as well as extra galactic sources like AGN or star forming regions obfuscate the hydrogen signature. Foreground removal is thus of paramount importance for cosmic 21cm observations (Chapman et al. 2014a; Chapman et al. 2015; 2016). Experiments are generally split into a number of categories, those that observe the sky-averaged, global signal, those observing 21cm brightness temperature fluctuations at

high redshifts during the EoR and beyond, and those mapping the statistical 21cm intensity at low redshifts. Here we will discuss these various observational techniques and the progress made over the last few years.

3.7.1 Global experiments

Global experiments target the 21cm monopole, in essence the signal shown in the lower panel of Fig. 3.5, and can thus in principle be observed using a single dipole. The large beam of a radio dipole can be used to effectively detect the sky-averaged signal. Current experiments include EDGES (Bowman et al. 2009), SARAS (Patra et al. 2013), LEDA [5], SCI-HI (Voytek et al. 2014), and the proposed lunar orbiter DARE (Burns et al. 2012). The EDGES collaboration have recently reported the detection of a strong 21cm absorption signal at $z \sim 17$ (Bowman et al. 2018a). With an absorption signal twice as deep as conventional model predictions, the implications of the detection are potentially great. Among other implications, the detection could signify evidence for dark matter-baryon interactions (Barkana 2018; Barkana et al. 2018; Kovetz et al. 2018) which would lead to cooling of the IGM prior to the EoR, and it requires enhanced star formation rates at high redshifts (Mirocha & Furlanetto 2018). Although the EDGES detection is exciting in terms of its potential for new physics (e.g. Barkana 2018; Fraser et al. 2018; Muñoz et al. 2018; Aristizabal Sierra & Fong 2018; Houston et al. 2018) confirmation of the detection from comparable experiments, as well as further investigation into the systematics are required. Recently, a number of publications call into question the analysis of the EDGES data (Hills et al. 2018; Bradley et al. 2018; Bull et al. 2018). Hills et al. (2018) show that the signal model used by Bowman et al. (2018a) implies unphysical foreground parameters and the flattened trough model is not a unique fit to the data, such that using alternative signal templates can fit the data without presenting a large absorption trough and thus no evidence for exotic physics. Further, Bradley et al. (2018) report that subtle systematic errors due to the instrument calibration can have large effects on the interpretation of the data. Although, some of the concerns have been addressed in Bowman et al. (2018b), future data released as well as independent detections will have to confirm their observations. A detection of the global signal well within the dark ages will however require complete shielding from terrestrial RFI, and the proposed DARE lunar orbiter is currently our best hope to achieve this (Harker et al. 2012).

[5] `http://www.tauceti.caltech.edu/leda/`

3.7.2 Epoch of Reionization experiments

Telescopes targeting the spatial fluctuations of the EoR include LOFAR (Patil et al. 2017), MWA (Dillon et al. 2015), HERA (DeBoer et al. 2017), PAPER (Parsons et al. 2010), and eventually the low frequency part of the SKA[6]. Although many observational challenges need to be overcome for a statistical detection of the cosmic signal, these experiments are beginning to set upper limits for the 21cm power spectrum during the EoR. The upper limits, summarized in Fig. 3.8, are still an order of magnitude above even some exotic model predictions, as total integration time for many of these experiments are low, and instrument calibration is generally not understood to a level where systematic effects would become minimal (Bull et al. 2018). It remains questionable whether pre-SKA interferometers will indeed measure the EoR signal, however these observational efforts will aid the understanding of signal, foregrounds and instrument when SKA sees its first light in the coming decade. The cosmological implications for an EoR detection are potentially wide-reaching. The 21cm signal is sensitive to the distribution of hydrogen gas in the early Universe and can be used to infer the underlying density field. Furthermore, due to its sensitivity to the temperature and ionization state of the IGM, the signal can be used to identify exotic energy injection mechanism.

3.7.3 21cm intensity mapping

After the EoR, the remaining HI is found in self-shielded regions in galaxies and galaxy clusters, and thus the 21cm emission serves as a tracer for the galaxy distribution. The galaxy distribution is in turn a biased tracer of the underlying dark matter density field, such that observations of the galaxy distribution allow measurements of the matter power spectrum. By integrating the emission of these hydrogen clouds over large portions of the sky, the 21cm emission can thus be used as a new avenue for large scale structure observations (Maartens et al. 2014). This low-resolution intensity mapping of unresolved galaxies provides promising prospects for precision measurements of the cosmological parameters (Bull et al. 2014). Next generation experiments such as CHIME[7], TIANLAI (Chen 2015), BINGO (Battye et al. 2016) and the mid-frequency part of the SKA [8], have all planned to carry out intensity mapping surveys in the coming decade. The information gain by this technique can be credited to two important properties of these type of observations. Firstly, the new generation of 21cm experiments able to perform these intensity mapping surveys will vastly increase the amount of comoving volume surveyed. SKA-

[6] https://www.skatelescope.org/lfaa/
[7] https://chime-experiment.ca/
[8] https://www.skatelescope.org/mfaa/

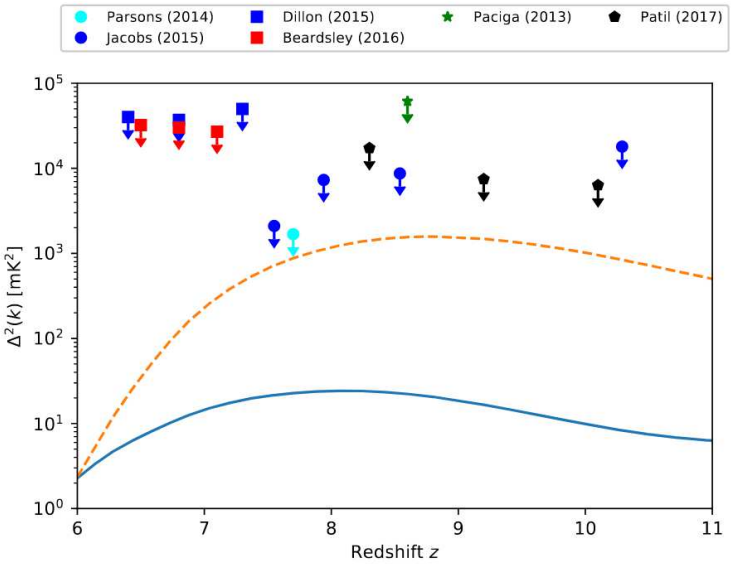

Figure 3.8: Current constraints on the 21cm power spectrum as a function of redshift. These are the results for GMRT (Paciga et al. 2013), PAPER32 (Parsons et al. 2014; Jacobs et al. 2015), MWA128 (Dillon et al. 2015; Beardsley et al. 2016) and LOFAR (Patil et al. 2017). A conservative 21cm model (solid blue curve) and a model with negligible heating (dashed orange curve) are shown for comparison. *Taken with permission from Bull et al. (2018).*

mid should be able to increase survey volumes by a factor of ~ 2 for a fully resolved galaxy survey and by a factor of $\gtrsim 3$ for statistical IM as compared to current galaxy surveys (Santos et al. 2014). Secondly, the direct relation between observed frequency of the cosmic 21cm line and the emission redshift readily provides the redshift information of the sources. Whereas photometric redshifts are necessary in conventional galaxy surveys due to the long time required to obtain spectroscopic redshift information, intensity mapping experiments will be able to circumvent this source of systematic errors.

As IM experiments will be probing the large scale structure of the Universe, this technique will primarily probe the baryon acoustic oscillation (BAO) scale and can thus serve to constrain cosmological parameters such as the dark energy equation of state parameter w and the spatial curvature Ω_K (Bull et al. 2014). These scales are also useful when measuring the effect of gravitational lensing on the 21cm signal and much higher signal to noise for lensing measurements can be established by non-resolved IM than can be expected from resolving individual galaxies (Pourtsidou & Metcalf 2014).

IM is thus a promising new technique, however, foregrounds, both galactic and extragalactic, can cause serious systematic errors and potentially bias the cosmological analysis (Wolz et al. 2014). Foreground removal techniques often rely on the assumption that foregrounds vary smoothly with frequency, while the cosmological signal is expected to

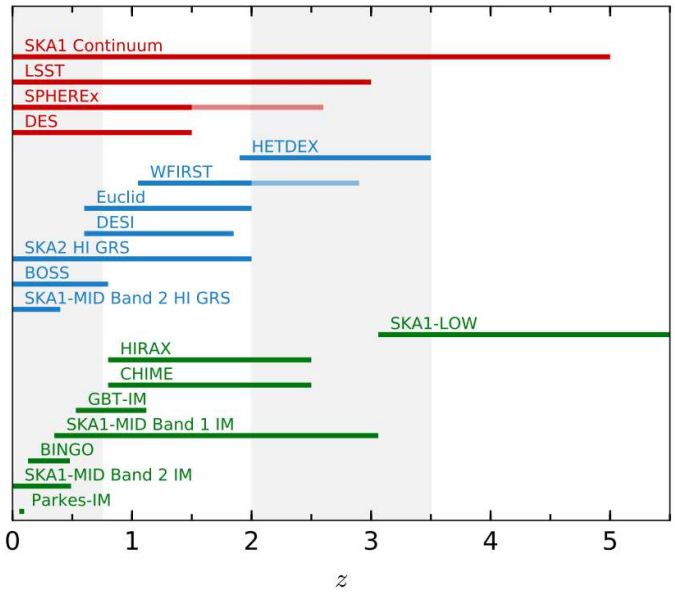

Figure 3.9: The approximate redshift ranges of verious current and future large-scale structure surveys including 21cm intensity mapping surveys. IM surveys are shown in green (bottom), spectroscopic galaxy redshift surveys in blue (middle) and photometric/-continuum surveys in red (top). From left to right, the grey and white shaded areas show an approximate division of the full redshift range into different eras, corresponding to the dark-energy-dominated regime, the onset of dark energy, the matter-dominated regime, and the fully matter-dominated regime. *Taken with permission from (Bull et al. 2018).*

be highly uncorrelated from the frequency at the MHz scale (Gnedin & Shaver 2004; Chapman et al. 2014b;a).

3.8 Summary

The hyperfine structure of the neutral hydrogen atom allows for a spin flip transition with energy difference equivalent to a photon emission at a rest-frame wavelength of 21cm. This chapter has introduced this 21cm transition in the context of cosmological observations to probe the most elusive epochs in the evolution of our Universe. We have studied the various contributions to the spin temperature and the ionization history which define the global signal of the cosmic 21cm radiation and have outlined a time-line for this sky-averaged signal. As the 21cm signal is tracing the hydrogen distribution the signal fluctuates on the sky and much can be learned from statistically observing and characterizing these fluctuations.

A claimed detection of the sky-averaged signal has recently made headlines and its unexpected findings have sparked a series of new and interesting studies to explain the observed phenomena. With a new generation of radio telescopes coming online in the near

future, discoveries akin to those made by the EDGES experiment will likely be a common occurrence. Though many observational challenges are yet to be overcome, these experiments should soon detect the statistical fluctuations of the 21cm signal, and peer deep into the Epoch of Reionization and the Cosmic Dawn. Projects such as the ambitious Square Kilometre Array will have the coming decade likely see a near overwhelming amount of data. We therefore turn our attention to the numerical and analytical analysis of future 21cm data in the remainder of the text.

Part III

Numerical Analysis of the Cosmic 21cm Signal

"Homo sapiens is an obsolete algorithm. After all,
what's the advantage of humans over chickens?"

– Yuval N. Harari.

Chapter 4

Machine learning and EoR emulation

In this chapter, we explore machine learning and its use in the field of 21cm cosmology. We focus on neural network emulation of semi-numerical EoR simulations as a way to increase the parameter inference efficiency for future 21cm analyses. With ever increasing data volumes, efficient analysis will be paramount to extracting maximal information from datasets and thus we investigate the use of an emulator as a means to bypass expensive simulation evaluations. We find strong predictive capabilities with our emulation technique when comparing with the direct evaluation of the model simulation. The most important feature to any machine learning technique is the selection of the training set which, in our case, informs the network and ultimately determines its behaviour. We therefore include an analysis of the training set used and the implications of lowering the training set size for our emulator. We find that the emulator retains its good predictive capabilities with training sets as small as 100 model evaluations, which may be a realistic number for fully numerical simulations.

This chapter explores the study of neural network emulation which was published in MNRAS (21 March 2018, Vol. 475, Issue 1, pp. 1213-1223): *C. J. Schmit and J. R. Pritchard; Emulation of reionization simulations for Bayesian inference of astrophysics parameters using neural networks*. The published version is available through: `https://doi.org/10.1093/mnras/stx3292`.

4.1 Big data in astronomy

Over the past two decades, the fields of astronomy and cosmology have experienced a surge in data availability that will continue for the foreseeable future. The Sloan Digital Sky Survey (SDSS) is one of the largest sky surveys conducted to date, which scans the sky daily to capture an astonishing 200 GB of data every night, observing millions of stars and galaxies. The great challenge for astronomical surveys in the near and distant

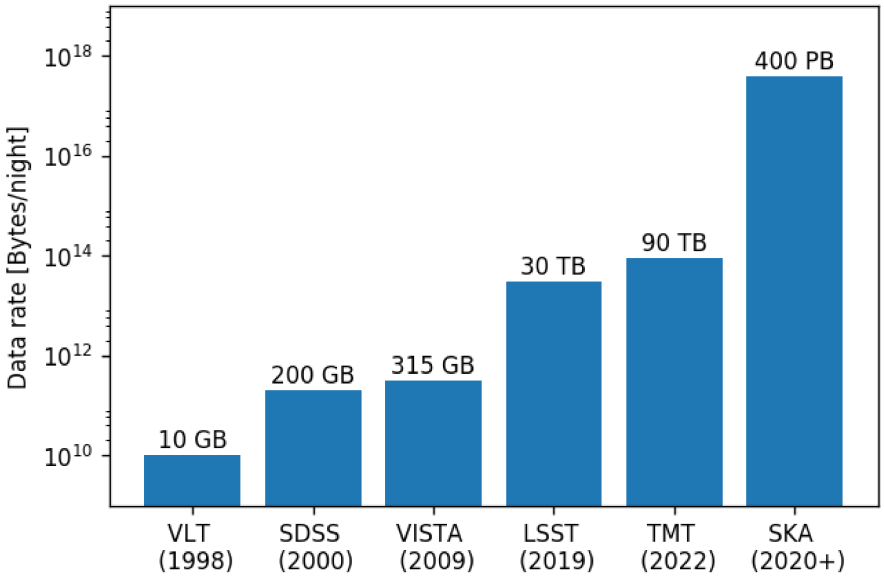

Figure 4.1: Data volumes of recent and upcoming telescopes. *Adapted with permission from Kremer et al. (2017).*

future is that the data rates of modern survey telescopes will dwarf those of SDSS by orders of magnitudes. Fig. 4.1 summarizes the average data acquisition rates of past and future surveys. It can clearly be seen that over the last few decades and especially with upcoming experiments, the explosion of data rates will pose an important problem for the astronomical community. How to best deal with these enormous datasets and what tools to use to extract the most information possible has become a focus for many research areas. With the advent of SKA, which is expected to produce many exabyte of raw data every night, this challenge lies at the heart of 21cm science. Luckily, this surge in data availability has been accompanied, as being partially caused by, an exponential rise in available computing power. These resources are applied in digitally scouring new and existing databases to extract information on scales that no human is able to do. This increased interest in data mining is again accompanied by the advancement of machine learning techniques, which seek to automate the knowledge discovery with minimal input from the scientist. Ball & Brunner (2010) rightfully pose the question of usefulness of data mining in a scientific context, when the techniques used do not have the physical insight a scientist looking at datasets might have. They proceed by listing 5 major advantages of machine learning and its application in data mining, which we summarize here.

- *Getting anything at all:* As seen in Fig. 4.1, future datasets will almost be overwhelmingly large, to the degree that a prudently automated approach to extract scientific information becomes necessary.

- *Simplicity:* Although a large number of machine learning algorithms exist, by selecting and applying a well suited and well tested method to the problem or dataset studied, productivity can be increased significantly.

- *Prior information:* The datasets could be flexibly allowed to either speak for themselves or readily incorporate results from prior analyses. As will be discussed in the next section, the distinction between supervised and unsupervised learning methods allows for this flexibility.

- *Pattern Recognition:* Although humans are notoriously good at recognising patterns, some patterns might even elude the shrewdest human observer due to being hidden in enormous data mounds or due to the high dimensional nature of the pattern. Machine learning techniques have been specifically designed to find these hidden structures.

- *Complementary approach:* Whether the machine learning techniques increase the scientific data output over traditional tools or not, they act as a complementary analysis which may alleviate systematic errors.

Although the interpretability of some techniques is questionable and Ball & Brunner (2010) detail that the limitations of the methodology are manifold, machine learning techniques have been successfully used in many areas of astronomy and science at large, and present a rich tool set to be applied in 21cm analyses as well.

Recent efforts (Shimabukuro & Semelin 2017; Kern et al. 2017), have now introduced machine learning techniques into the field of 21cm Cosmology. One of the first papers in this area, by Shimabukuro & Semelin (2017), studied the use of artificial neural networks to estimate model parameters for 21cm observations. The network they constructed used the power spectrum outputs from a training set of semi-numerical simulation runs, and was able to recover the input parameters of a mock power spectrum observation. This study set the precedent for the use of neural networks as a viable tool for parameter inference. Instead of computing a single value for the model parameters which the network thinks best describes an observation, we focussed on using a neural network in a Bayesian parameter inference study. A common approach is to evaluate the likelihood of a given parameter set to describe an observation, then vary the parameter set and re-evaluate said likelihood. Over many such iterations the model likelihood is sampled and we can estimate the best fit parameters of our model given the observation. In practice one such MCMC sampling of the likelihood requires many thousands of model evaluations and is impractical for models with considerable run-time, such as fully numerical reionization simulations. However, with a sufficiently well trained network, the

model can be approximated by the network and thus no further model evaluations, beyond those required to build a training set, are necessary.

This has been the primary motivation to study artificial neural networks and their potential in speeding up Bayesian parameter inference for 21cm observations and we will discuss the findings in the remainder of the chapter. In section 4.2 we define Machine learning and contrast the two major learning paradigms in the field. Section 4.3 looks at artificial neural networks in detail and we present the learning algorithm used. Then in section 4.4 we present a toy model analysis indicative of the analysis carried out in ***Paper I***. Section 4.5 introduces the reionization model used as well as the parameters varied in the analysis. Once the problem is sufficiently set-up, we present our training methodology together with our power spectrum emulation predictions in section 4.6. Finally, we discuss the Bayesian inference problem in section 4.7 and our results are presented in section 4.8. A final summary of this chapter is given in section 4.9.

4.2 Machine learning

A concise definition for machine learning can be found in Ivezic et al. (2014):

> *Machine learning is an umbrella term for a set of techniques for interpreting data by comparing them to models for data behaviour, such as various regression methods, supervised classification methods, maximum likelihood estimators, and the Bayesian method. [...]*

Machine learning can thus be regarded as applied statistics in the same way that engineering can be viewed as applied physics. The techniques used for machine learning tasks are deeply rooted in statistics and have well formulated theoretical foundations (Ivezic et al. 2014). This definition also alludes to the fact that a large range of machine learning techniques exist and their usage depends on the desired outcome of the analysis using them.

Techniques can generally be classified according to whether a feedback mechanism is present to penalise certain learning behaviours. If the system is given certain instructions or labels according to which the data is to be analysed, and is penalised if the system proposes a solution that violates these instructions, the learning technique used is referred to as a *supervised learning* technique. If, however, no labels are applied nor any feedback is given and the system is left on its own to explore the dataset, an *unsupervised learning* technique is used. We proceed by giving a short overview of both learning paradigms including a non-exhaustive list of example algorithms for each case.

4.2.1 Unsupervised learning

Unsupervised methods are *descriptive* of the dataset they are applied to and do not require training. Although unsupervised learning is applied to a broad class of problems, the three main areas in which unsupervised learning algorithms are used are density estimation, clustering, and dimensionality reduction.

Density estimation algorithms seek to determine the probability density function underlying a given dataset (Balogh et al. 2004; Ferdosi et al. 2011). A popular approach is that of Kernel density estimation (KDE) which can be thought of as the generalization of binning data into histograms. The method centres a kernel of arbitrary shape on the data points and accumulates the signal of each kernel function across the data range. For a set of N data points $\{x_i\}$ represented in D dimensions, KDE thus approximates the *pdf* from which the points were drawn via

$$\hat{f}_N(x) = \frac{1}{Nh^D} \sum_{i=1}^{N} K\left(\frac{d(x, x_i)}{h}\right), \tag{4.1}$$

where $K(u)$ denotes the kernel function with width h, and $d(x, x_i)$ is the distance between points x and x_i. Although Gaussian kernels are widely used, a large variety of kernel functions are useful. Another density estimation technique relies on the distance to the K-th nearest neighbour at each point. It simply estimates the density by determining the volume which incorporates the K nearest neighbours,

$$\hat{f}_K(x) = \frac{K}{V_D(d_K)}, \tag{4.2}$$

where $V_D(d_K)$ is the volume of a D-sphere with radius equal to the distance to the K-th nearest neighbour d_K.

Clustering algorithms try to group data points into a number of separate clusters. The K-means algorithm (eg. Ordovás-Pascual & Sánchez Almeida 2014), for example, takes a number K of clusters into which the data is to be separated, and minimizes the sum of square distances between data points and the centre of the clusters to which they have been assigned. This algorithm requires some initial guess as to how to separate the data, after which the method applies its minimization criterion. Whereas K-means relies on human input for the number of clusters, other algorithms, such as *mean-shift*, do not require this input and determine the number of clusters from the data. The *mean-shift* algorithm uses a KDE and allocates points into clusters associated with the peaks of the density estimate. Naturally, the number of clusters found by this algorithm depends on the number of peaks in the kernel density estimate of the data.

Dimensionality reduction is a way to directly address the curse of dimensionality incurred by the growing size of analysable properties of datasets. As the dimensionality of the data bases grow, one would like to select those properties that present the most information as possible. Principal component analysis (PCA) (Joliffe 1986) is probably one of the most widely used dimensionality reduction techniques and tries to transform the data according to the eigenvectors of the data matrix. Thus transforming the data into a basis given by the directions which contain the greatest amount of information, and finally cutting from the analysis those directions which contain the least amount of information. For a dataset X, with covariance matrix C_X, PCA transforms the data into a basis Y according to,

$$C_Y = R^T C_X R, \tag{4.3}$$

where the projection matrix R is made up by the principal components, which represent the directions of maximal variance of C_X. Once an ordered set of principal components is found, the ones containing the least information are truncated from the analysis, thus reducing the dimensionality of the dataset retaining the most information possible. The inherent limitation of PCA is that it relies on a linear transform of the dataset, which can be problematic when analysing very high dimensional or non-linear datasets. More complex non-linear dimensionality reduction techniques such as manifold learning can be applied in those cases (Vanderplas & Connolly 2009).

4.2.2 Supervised learning

Supervised learning techniques are *predictive* based on a well understood training dataset. The prediction - feedback cycles applied in supervised learning techniques are akin to that of a student - teacher relationship. The method (student) proposes a solution to a problem based on its current knowledge, and the feedback mechanism (teacher) penalises the method (student) if the solution is too far off the truth. Finally, the method (student) updates its knowledge accordingly. In a supervised learning algorithm, the penalisation is implemented via some loss function based on the training set, which the algorithm tries to minimize during its training. Classification and regression are the two main areas where supervised learning techniques are applied. Whereas classification can be simply understood as the act of classifying data into to a set of discrete classes, many machine learning algorithm use a similar framework for regression problems by assigning data to a set of classes which are themselves continuous numbers. This means that most of the algorithms presented here can be applied to both categories of problems.

Decision trees (DT) are made up of a hierarchical node structure, which recursively divides the input data at each node according to some criterion which minimizes the loss

function. Terminal nodes, called leaf nodes, are labelled with the desired classification label for data following the necessary path through the tree to arrive at the particular leaf. Some example applications to astronomical data can be found in Ball et al. (2006),Vasconcellos et al. (2011) and Franco-Arcega et al. (2013).

Support vector machines (SVM) are another supervised learning algorithm widely applied. SVM take a training set of data with associated class labels and try to find the hyperplane which separates the closest members of separate classes. This creates a set of decision boundaries according to which new data is to be classified. The training points falling on top of this decision boundary are called support vectors, and thus inform the loss function that is minimized through the boundary creation process.

Gaussian processes (GP) are used extensively in non-linear regression problems as they can lead to very flexible regression models (Rasmussen & Williams 2005). A GP is a collection of random variables which have a joint multivariate Gaussian distribution. In one dimension one can think of the GP as a mechanism to draw curves from a family which obey the statistics set by the GP. A set of training data will then limit the set of possible curves drawn to those passing through the training points. The more training points that are available, the more curves drawn from the GP will resemble that of the underlying model which produced the data.

Finally, arguably the most famous supervised learning algorithm is that of the *artificial neural network*. This algorithm tries to mimic the networks of neurons in animal brains, an their interconnected nature. Once again, supervised networks learn via minimization of a loss function which is informed by the training data. We will discuss neural networks and the learning algorithm used in our analysis in more detail in the following section.

4.3 Neural networks

Artificial neural networks (ANN) are a class of machine learning algorithms inspired by the signal processing through consecutive firing of connected neurons found in animal brains in Nature. The human visual cortex is a great example to illustrate naturally occurring neural networks. When looking at a picture, the incoming light stimulates neurons in the retina which send the information of the picture through the optical nerve into the primary visual cortex. This visual cortex consists of around 140 million neurons grouped into a number of functionally separate layers which feed forward the information received by the optical nerve. Consecutively information is send through to 5 different areas in the brain, again each consisting of millions of neurons which propagate signals through the brain according to the visual input. The amalgamation of neuron activity due to the signal

received finally ends up in what we perceive as the image observed. Our everyday experience of the world around us, allows us to progressively strengthen neuron connections and associate certain neuron activity with familiar concepts.

4.3.1 Architecture

In our analysis, we use a multilayer perceptron (MLP) as our artificial neural network design. A supervised MLP uses a set of training data $T \subset X \times Y$, where X denotes the input or parameter space and Y denotes the output space, upon which the neural network tries to fit a mapping $f : X \to Y$. This is to say that the neural network is finding a mapping between input and output data, which is sensitive to the key features of the training set. This mapping can then be used on unknown data where the neural network uses its acquired knowledge of the system to infer an output, either in form of a classification or a number.

An artificial neural network consists of three types of layers each consisting of a set of nodes or neurons, illustrated in Fig. 4.2. The input layer takes N_i data points into N_i input nodes from which we want to predict some output. Each node in the input layer is connected to all of N_j nodes in the first of L hidden layers via some weight $w_{ij}^{(1)}$. The input to the nodes in the hidden layer is a linear combination of the input data and the weights,

$$s_j^{(1)} = \sum_{i=1}^{N_i} x_i w_{ij}^{(1)}. \tag{4.4}$$

A neuron is then activated by some activation function $g : \mathbb{R} \to \mathbb{R}$. We use a sigmoid activation function, $g(s) = 1/(1 + e^{-s})$, as this non-linear function allows us to fit to any function in principle (Cybenko 1989). This activation step can be interpreted as each neuron having specialised on a certain feature in the system (Bishop 2006; Gal 2016) and when the data reflects this feature the neuron will be activated. The output from the neuron activation is then fed into the next hidden layer as input, such that the j^{th} neuron in the ℓ^{th} hidden layer computes,

$$t_j^{(\ell)} = g\left(s_j^{(\ell)}\right), \tag{4.5}$$

where, for $1 < \ell \leq L$,

$$s_j^{(\ell)} = \sum_{i=1}^{N_j} t_i^{(\ell-1)} w_{ij}^{(\ell)}. \tag{4.6}$$

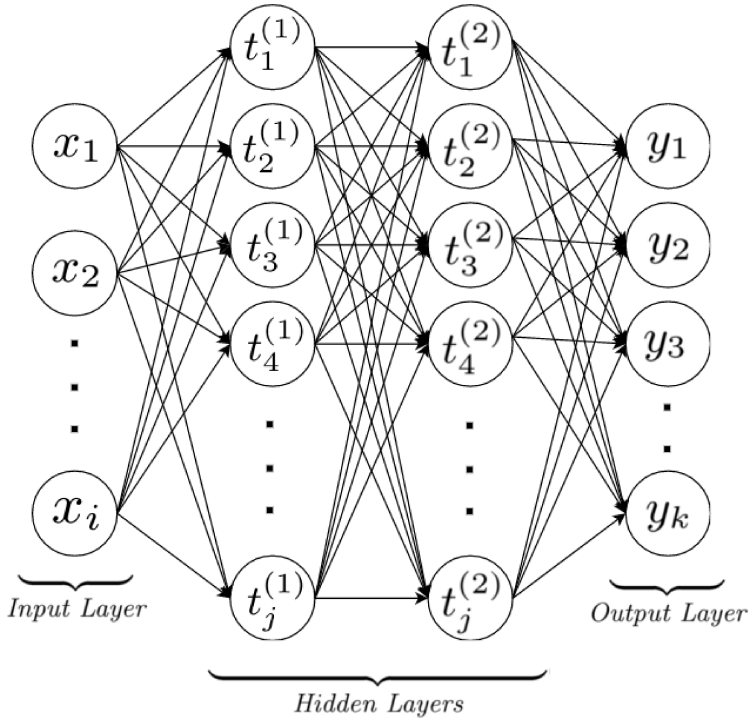

Figure 4.2: Multilayer perceptron layout.

Finally, the output layer combines the outputs from the final hidden layer into N_k desired output values,

$$y_k = \sum_{i=1}^{N_j} t_i^{(L)} w_{ik}^{(L+1)}. \tag{4.7}$$

The weights between neurons $w_{ij}^{(\ell)}$ are obtained during the training of the network, where we apply the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm (Ziegel et al. 1987) to minimize the mean-square error between the true value provided by the training data and the value predicted by the network. This training algorithm is ideal for sparse training sets and a low dimensional parameter space (Le et al. 2011), and will be discussed in the following section.

4.3.2 Learning algorithm

A popular training algorithm for machine learning problems is back-propagation via gradient descent (Rumelhart et al. 1986; Cheng & Titterington 1994; Abu-Mostafa 2012; Shimabukuro & Semelin 2017). However, back-propagation requires the user to manually set a learning rate which must fall within a finite range, too small and each training iteration produces vanishingly small changes, too large and the training steps overshoot.

An arbitrary learning rate does therefore not guarantee that the network will converge to a point with vanishing gradient and second order optimization methods can be used to guarantee convergence (Battiti 1992).

Suppose we have N_{train} training sets consisting of N_{i} input parameters and N_{out} output data. These training sets are fed into a neural network as described in the previous section. Training an ANN can then be viewed as an optimization problem where one seeks to minimize the total cost function $E(\boldsymbol{w})$, which is the sum-squared error over the training sets.

$$E(\boldsymbol{w}) = \sum_{n=1}^{N_{\text{train}}} E_n(\boldsymbol{w}) = \sum_{n=1}^{N_{\text{train}}} \left[\frac{1}{2} \sum_{i=1}^{N_{\text{out}}} \left(y_{i,n}(\boldsymbol{w}) - d_{i,n} \right)^2 \right], \tag{4.8}$$

where $y_{i,n}$ is the prediction made by the neural network in the i^{th} neuron of the output layer, using the n^{th} parameter set of all training inputs, $d_{i,n}$ is the true result for the i^{th} neuron in the output layer corresponding to the n^{th} parameter set, and thus E_n is the cost function associated with the n^{th} input parameter set.

We can expand the cost function around some particular set of weights $\boldsymbol{w}_0$ using a Taylor series,

$$\begin{aligned} E(\boldsymbol{w}) =& E(\boldsymbol{w}_0) + (\boldsymbol{w} - \boldsymbol{w}_0)^T \boldsymbol{g}_0 \\ &+ \frac{1}{2} (\boldsymbol{w} - \boldsymbol{w}_0)^T H_0 (\boldsymbol{w} - \boldsymbol{w}_0) + ..., \end{aligned} \tag{4.9}$$

where $\boldsymbol{g}_0$ is the vector of gradients and H_0 denotes the Hessian matrix with elements

$$h_{ij} = \frac{\partial^2 E}{\partial w_i \partial w_j}. \tag{4.10}$$

Whereas back-propagation is based on a linear approximation to the error surface, better performance can be expected when using a quadratic error model,

$$\begin{aligned} E(\boldsymbol{w}) \approx& E(\boldsymbol{w}_0) + (\boldsymbol{w} - \boldsymbol{w}_0)^T \boldsymbol{g}_0 \\ &+ \frac{1}{2} (\boldsymbol{w} - \boldsymbol{w}_0)^T H_0 (\boldsymbol{w} - \boldsymbol{w}_0). \end{aligned} \tag{4.11}$$

Provided H_0 is positive definite, this approximation to the error surface has a minimum, $\partial E / \partial \boldsymbol{w} = 0$, at

$$\boldsymbol{w} = \boldsymbol{w}_0 - H_0^{-1} \boldsymbol{g}_0. \tag{4.12}$$

Given that a quadratic approximation to the actual cost function is used, an iterative approach needs to be taken in order to find an estimate of the true minimum. Similar to back-propagation where $\boldsymbol{g}$ is used as the search direction, second order methods use $-H^{-1}\boldsymbol{g}$ as

the search direction. Thus the search direction during training iteration k is given by

$$\boldsymbol{\Delta}_k = -H_k^{-1}\boldsymbol{g}_k. \tag{4.13}$$

Solving this system of equations requires precise knowledge of the Hessian, as well as a well-conditioned Hessian, which is not always guaranteed. Instead of computing the Hessian and inverting it, the BFGS scheme seeks to estimate H_k^{-1} directly from the previous iteration. Mcloone et al. (2002) give the basic algorithmic structure as follows;

- Set the search direction $\boldsymbol{\Delta}_{k-1}$ equal to $-M_{k-1}\boldsymbol{g}_{k-1}$, where M_{k-1} is the approximation to H_{k-1}^{-1} at the $(k-1)^{\text{th}}$ iteration.
- Use a line search to find the weights which yield the minimum error along $\boldsymbol{\Delta}_{k-1}$,

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \eta_{\text{opt}}\boldsymbol{\Delta}_{k-1}, \tag{4.14}$$

$$\eta_{\text{opt}} = \min_{\eta}(E(\boldsymbol{w}_{k-1} + \eta\boldsymbol{\Delta}_{k-1})). \tag{4.15}$$

- Compute the new gradient $\boldsymbol{g}_k$.
- Update the approximation to M_k using the new weights and gradient information,

$$\boldsymbol{s}_k = \boldsymbol{w}_k - \boldsymbol{w}_{k-1} \text{ and } \boldsymbol{t}_k = \boldsymbol{g}_k - \boldsymbol{g}_{k-1}, \tag{4.16}$$

$$A_k = \left(1 + \frac{\boldsymbol{t}_k^T M_{k-1}\boldsymbol{t}_k}{\boldsymbol{s}_k^T\boldsymbol{t}_k}\right)\frac{\boldsymbol{s}_k\boldsymbol{s}_k^T}{\boldsymbol{s}_k^T\boldsymbol{t}_k}, \tag{4.17}$$

$$B_k = \frac{\boldsymbol{s}_k\boldsymbol{t}_k^T M_{k-1} + M_{k-1}\boldsymbol{t}_k\boldsymbol{s}_k}{\boldsymbol{s}_k^T\boldsymbol{t}_k}, \tag{4.18}$$

$$M_k = M_{k-1} + A_k - B_k. \tag{4.19}$$

The scheme initializes by taking a step in the direction of steepest descent by setting, $M_0 = \mathcal{I}$.

The limited-memory BFGS scheme we are using, recognizes the memory intensity of storing large matrix estimates of the inverse Hessian, and resets M_{k-1} to the identity matrix in equation (4.19) at each iteration and multiplies through by $-\boldsymbol{g}_k$ to obtain a matrix free expression for $\boldsymbol{\Delta}_k$.

- The LBFGS thus uses the following update formula (Asirvadam et al. 2004)

$$\boldsymbol{\Delta}_k = -\boldsymbol{g}_k + a_k\boldsymbol{s}_k + b_k\boldsymbol{t}_k, \tag{4.20}$$

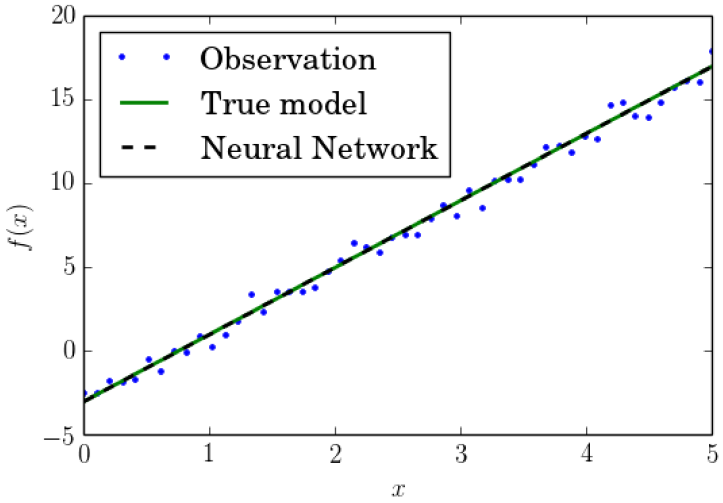

Figure 4.3: Observations (blue dots) from toy linear model (orange line) drawn in 50 equispaced x bins with a normal error. We also show the ANN prediction for the true parameter combination (black dashed line). Note that this is almost indistinguishable from the true model showing the excellent performance of the ANN.

with

$$a_k = -\left(1 + \frac{\boldsymbol{t}_k^T \boldsymbol{t}_k}{\boldsymbol{s}_k^T \boldsymbol{t}_k}\right) b_k + \frac{\boldsymbol{t}_k^T \boldsymbol{g}_k}{\boldsymbol{s}_k^T \boldsymbol{t}_k}, \quad \text{and} \quad b_k = \frac{\boldsymbol{s}_k^T \boldsymbol{g}_k}{\boldsymbol{s}_k^T \boldsymbol{t}_k}. \tag{4.21}$$

4.4 Emulation and toy model

In this section we apply our MLP to a toy model seeking to illustrate the emulation paradigm explored in our work.

Consider the following problem; We have a set of data to which we want to fit a certain model. Our data, shown in Fig. 4.3, is drawn from a linear model with errors drawn from a normal distribution with unit standard deviation. Our toy model has the following functional form,

$$f(x) = ax + b, \tag{4.22}$$

where we have chosen $(a, b) = (4, -3)$ as our set of fiducial parameters which thus defines the model from which the data was drawn. An observer who only has access to the data might want to know which coefficients of the model does the observed data prefer. An immediate approach to solving this problem is of course to use a Markov Chain Monte Carlo (MCMC) algorithm, to map out the likelihood of the parameter values given the data and associate the coefficients which present the highest likelihood values as the best fit parameters. What if, however, the evaluation of the model in question is much more cumbersome than simply evaluating a linear model at a point in parameter space, to the point where it becomes impractical to evaluate the model a large number of times. In this

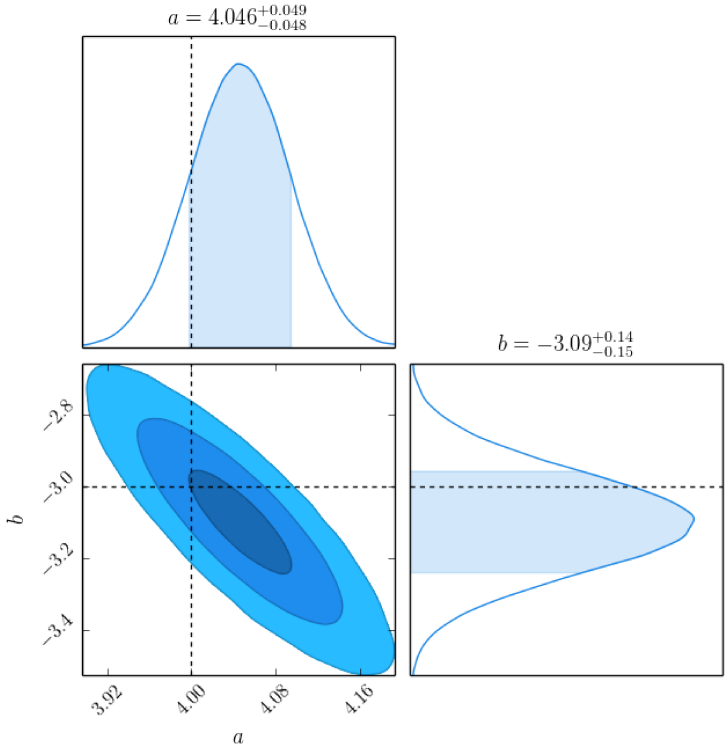

Figure 4.4: MCMC inference of the toy model via neural network emulation. From dark blue to light blue, we show the 1σ, 2σ and 3σ confidence intervals. Above the marginalized likelihood curves, we note the maximum likelihood value with 1σ error bounds.

case, one might want to approximate the behaviour of the model, $\tilde{f}(x) \sim f(x)$, in some way that is quick to evaluate at any point in parameter space. This approximating of the model evaluation step, through a faster evaluation mechanism is what we call *emulation* of the model. Here we approximate our toy model via an ANN, as described in the previous section, which trains on a set of model evaluations, evaluated on a grid of 20 parameter values in each direction. We assign the output of each output node of the ANN to a bin in the x range. Therefore, we have

$$y_i = \tilde{f}(x_i), \tag{4.23}$$

where i indexes the x bins of the observed data and the output nodes of the ANN. When computing the likelihood for an MCMC step, one now replaces the model evaluation with a prediction from the neural network. Such that the log-likelihood evaluated at each step of the MCMC becomes

$$\ln \mathcal{L} = -\sum_i \frac{\left[d(x_i) - \tilde{f}\,(x_i|\boldsymbol{\theta})\right]^2}{2\sigma^2}, \tag{4.24}$$

where $\tilde{f}$ denotes the neural network prediction for x-bin i at the parameters $\boldsymbol{\theta} = (a, b)$.

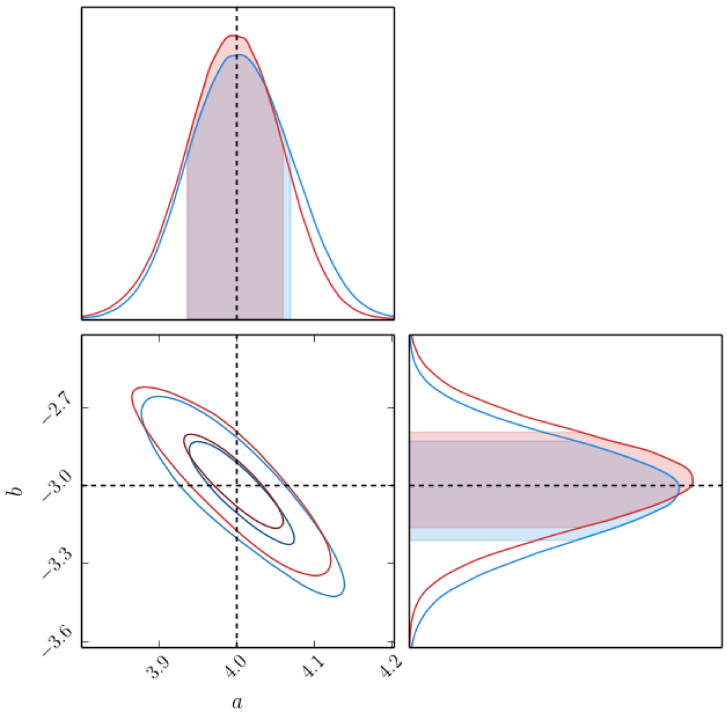

Figure 4.5: Bootstrapped MCMC inference of the toy model via neural network emulation (red) and the analytic model (blue). We show the 1σ and 2σ contours. Both methods recover the fiducial values within the 1σ confidence interval.

In Fig. 4.4 we present the results for this toy model for predicting the best fit values for the coefficients a and b, and find that the neural network indeed recovers the fiducial values within its 1σ error bounds.

The random nature of the Gaussian errors imposed on our mock observations means that we do not know whether any discrepancy between the fiducial values and the values recovered by the MCMC is due to some ineptitude of the network or due to the noise. We therefore bootstrap 10 observation-MCMC cycles in which any random fluctuations should average out. These predictions are shown in Fig. 4.5 and we compare the neural network emulation (red) to using the analytic expression for f in the likelihood (blue). Both methods manage to determine the parameters used to produce the mock observations to the same degree, implying that the emulation in this case was successful. We now continue by examining the usage of the neural network for emulation in the context of 21cm simulations.

4.5 Reionisation model

In order to produce the training sets upon which our neural network is ultimately trained, we need to model the EoR and the 21cm power spectrum as a function of some tangible model parameters.

The main observable of 21cm studies is the 21cm brightness temperature, defined by (Pritchard & Loeb 2012; Furlanetto et al. 2006; see equation (3.7))

$$\begin{aligned}\delta T_{\rm b}(\nu) \approx 27 x_{\rm HI}(1+\delta_{\rm b})\left(\frac{\Omega_{\rm b}h^2}{0.023}\right)\left(\frac{0.15}{\Omega_{\rm M}h^2}\frac{1+z}{10}\right)^{1/2} \\ \times\left(1-\frac{T_\gamma(z)}{T_{\rm S}}\right)\left[\frac{\partial_r v_r}{(1+z)H(z)}\right]^{-1}\,{\rm mK},\end{aligned} \tag{4.25}$$

where $x_{\rm HI}$ denotes the neutral fraction of hydrogen, $\delta_{\rm b}$ is the fractional overdensity of baryons, $\Omega_{\rm b}$ and $\Omega_{\rm M}$ are the baryon and total matter density in units of the critical density, $H(z)$ is the Hubble parameter and $T_\gamma(z)$ is the CMB temperature at redshift z, $T_{\rm S}$ is the spin temperature of neutral hydrogen, and $\partial_r v_r$ is the velocity gradient along the line of sight. One can define the 21cm power spectrum from the fluctuations in the brightness temperature relative to the mean,

$$\delta_{21}(\boldsymbol{x}, z) \equiv \frac{\delta T_{\rm b}(\boldsymbol{x}) - \langle \delta T_{\rm b}\rangle}{\langle \delta T_{\rm b}\rangle}, \tag{4.26}$$

where $\langle ...\rangle$ takes the ensemble average. The dimensionless 21cm power spectrum, $\Delta^2_{21}(k)$, is then defined as

$$\Delta^2_{21}(k) = \frac{k^3}{2\pi^2}P_{21}(k), \tag{4.27}$$

where $P_{21}(k)$ is given through

$$\left\langle \tilde{\delta}_{21}(\boldsymbol{k})\tilde{\delta}_{21}(\boldsymbol{k'})\right\rangle = (2\pi)^3\delta^{\rm D}(\boldsymbol{k}-\boldsymbol{k'})P_{21}(k). \tag{4.28}$$

Here, $\tilde{\delta}_{21}(\boldsymbol{k})$ denotes the Fourier transform of the fluctuations in the signal and $\delta^{\rm D}$ denotes the 3D Dirac delta function.

The 21cm power spectrum is the most promising observable for a first detection of the signal (Furlanetto et al. 2006), and encodes information about the state of reionization throughout cosmic history. For the evaluation of the 21cm power spectrum we utilize the streamlined version of `21cmFast`, which was used in the MCMC parameter study of Greig & Mesinger (2015). This version of `21cmFast` is optimized for astrophysical parameter searches.

The astrophysical parameters that we allow to vary in our model are three-fold.

Ionizing efficiency, ζ: The ionization efficiency combines a number of reionization parameters into one. We define $\zeta = A_{\rm He} f_* f_{\rm esc} N_{\rm ion}$, where $A_{\rm He} = 1.22$ is a correction factor to account for the presence of helium and converts the number of ionizing photons to the number of ionized hydrogen atoms, f_* is the star formation efficiency, $f_{\rm esc}$ is the escape fraction for UV radiation to escape the host galaxy, and $N_{\rm ion}$ is the number of ionizing photons per baryons produced. These parameters are poorly constrained at high redshifts. As $N_{\rm ion}$ depends on the metallicity and the initial mass function of the stellar population, we can approximate $N_{\rm ion} \approx 4000$ for Population II stars with present day initial mass function, and $N_{\rm ion} < 10^4$ for Population III stars. The value for the star formation efficiency f_* at high redshifts is extremely uncertain due to the lack of collapsed gas. Therefore, although $f_* \approx 0.1$ is reasonable for the local Universe it is uncertain how this relates to the value at high redshifts. Additionally a constant star formation rate has been disfavoured by recent studies (Mason et al. 2015; Mashian et al. 2016; Furlanetto et al. 2017). For our purpose however, a simplistic constant star formation model is sufficient. Similarly, the UV escape fraction $f_{\rm esc}$ observed for local galaxies only provides a loose constraint for the high redshift value. Although $f_{\rm esc} < 0.05$ is reasonable for local galaxies, large variations within the local galaxy population is observed for this parameter. We thus allow the ionization efficiency to vary significantly in our model to reflect the uncertainty on the limits of this parameter, and consider $5 \leq \zeta \leq 100$.

Maximal distance travelled by ionizing photons, $R_{\rm mfp}$: As structure formation progresses, dense pockets of neutral hydrogen gas emerge where the recombination rate for ionized proton - electron pairs is much higher than the average IGM. These regions of dense hydrogen gas are called Lyman limit systems, or damped Lyman-α systems, and effectively absorb all ionizing radiation at high redshifts. This effectively limits the bubble size of ionized bubbles during reionization. EoR models include the effect of these absorption systems as a mean free path of the ionizing photons. However, due to the limited resolution of `21cmFast`, this sub-grid physics is modelled as a hard cut-off for the distance travelled by ionizing photons. As our allowed range for this parameter we use, $2\ {\rm Mpc} \leq R_{\rm mfp} \leq 20\ {\rm Mpc}$.

Minimum virial temperature for halos to produce ionizing radiation, $T_{\rm vir}$: Star formation is ultimately regulated by balancing thermal pressure and gravitational infall of gas in virialized halos. Molecular hydrogen allows gas to cool rapidly, on timescales lower than the dynamical timescale of the system, such that an unbalance of the two opposing forces occurs and the gas collapses which triggers a star to form. Although initial bursts of population III stars are thought to be able to occur briefly in halos virialized at $T_{\rm vir} \sim 10^3$ K, these stars produce a strong Lyman-Werner background which leads to a higher dissociation of H_2 molecules. Star formation then moves to halos with $T_{\rm vir} > 10^4$ K, where HI

is ionized by virial shocks and atomic cooling is efficient. $T_{\rm vir}$ thus sets the threshold for star formation and we consider 10^4 K $< T_{\rm vir} < 2 \times 10^5$ K.

4.5.1 EoR simulation with 21cmFast

Simulations of the EoR are difficult in practice, due to the large range of scales which need to be simulated. In order to obtain statistically significant models of the ionization and 21cm radiation field during this epoch, GPc-sized simulation boxes are required. However, these fields are affected by small scale feedback mechanisms from sources much smaller than the size of the simulation boxes. As a resolution large enough to model both source behaviour and their large scale effects is impractical, approximations of the underlying physics is essential. Numerical models, such as those presented in Iliev et al. (2006a;b; 2014), combining full scale radiative transfer and N-body simulations of the EoR to achieve the highest fidelity simulations of the epoch are expensive and not suitable for parameter inference studies. In Mesinger et al. (2011) a compromise is struck between simulations of small-scale physics and using analytical models to predict some computationally intensive behaviour, resulting in the semi-numerical approach of `21cmFast`.

Simulating the 21cm radiation model, equation (3.7), is done in a number of steps. First, the density and velocity field are generated at high redshift as random Gaussian fields (Mesinger & Furlanetto 2007), and evolved to the EoR using the first order approximation for gravitational collapse via the Zel'dovich approximation (Zel'dovich 1970). This allows for a fast generation ($\sim$ 10 minutes) of the density field during the EoR. Second, the ionization field is produced using the *Fast Fourier Radiative Transfer* scheme of Zahn et al. (2011). This method computes $f_{\rm coll}$ directly on the evolved density field using a Press-Schechter type argument. In the simulation, ionized cells are thus identified as cells where

$$f_{\rm coll}(\boldsymbol{x}, z, R) \geq \zeta^{-1}, \tag{4.29}$$

with ζ being the ionizing efficiency previously introduced and R being a variable smoothing scale which is decreased down to the cell size. Third, the code introduces redshift space distortions due to peculiar velocities. Using the Zel'dovich approximation on the velocity field, the derivative of the line of sight velocity, v_r, needed in the expression for the brightness temperature fluctuations (equation 3.7), is written in k-space as

$$\begin{aligned} \frac{dv_r}{dr}(\boldsymbol{k}, z) &= ik_r v_r(\boldsymbol{k}, z) \\ &\approx \frac{k_r^2}{k^2}\dot{D}(z)\delta_{\rm nl}(\boldsymbol{k}). \end{aligned} \tag{4.30}$$

The code uses the final approximation to compute the effects of redshift space distortions. Finally, once the density field, ionization field and redshift space distortions are computed, they are straight forwardly combined to produce the brightness temperature field of equation (3.7).

4.6 Predicting the 21cm power spectrum

We use two different approaches to emulate the 21cm power spectrum. First, we use a simple two-layer MLP, as described in section 4.3, with 30 nodes in each layer, as we require the network to be sufficiently complex to map our set of 3 parameters to 21 power spectrum k-bins. This is done using the `MLPRegressor` class from the python package `SciKit-learn` (Pedregosa et al. 2011). The NN is then trained on a variety of training sets, see 4.6.1 and 4.6.2, obtained from `21cmFast` simulations. Then, for comparison we use trilinear interpolation of the training set, simply interpolating the power spectrum on a parameter grid.

4.6.1 Grid-based approch

In order to study the impact of the choice of training set on the predicting power of the ANN we prepared a variety of training sets. The most basic approach, in fact the one used for our toy model in section 4.4, is to distribute parameter values regularly in parameter space and obtaining the power spectrum for each point on a grid. We vary our parameters as per section 4.5, $5 \leq \zeta \leq 100$, 10^4 K $\leq T_{\rm vir} \leq 10^5 K$ and 2 Mpc $\leq R_{\rm mfp} \leq 20$ Mpc, as these reflect our prior on the likely parameter ranges. Each training set then consists of the power spectrum evaluated in 21 k-bins, set by the box size of 250 Mpc, upon which the ANN is trained. We compare 5 different training sets at 2 different redshift bins, $z = 8$ and $z = 9$. These training sets consist of 3, 5, 10, 15 and 30 points per parameter, which leads to training sets of total size 27, 125, 1000, 3375 and 27000 respectively.

This approach is the most basic and certainly the most straight forward to implement, however it comes with a number of drawbacks. Projected down, a gridded set of parameter values has multiple points which occupy the same parameter values. This implies that the simulation is evaluated multiple times at the same values for some parameter at each point in any given row in the grid, see Fig. 4.6. Furthermore, if the observable is varying slowly in some parameter, few points are needed to model its behaviour and thus valuable simulation time is wasted on producing points in the grid that add very little information.

Another important limitation is the exponential scaling of the total number of points with the number of parameters in the grid. In the simple three dimensional case, which

we are studying here, N evaluations per parameter lead to a total of N^3 points on the grid. Ultimately, it is desirable to allow the model cosmology to vary and include at least 6 cosmological parameters into the search as well as additional astrophysical parameters, such as the X-ray efficiency, f_X, obscuring frequency, $\nu_{\rm min}$, and the X-ray spectral slope, α_X. One is then looking at a total of 12 or more parameter dimensions for which evaluations on the grid are prohibitively expensive and other techniques are needed. A further problem is presented by the proportion of volume in the corners of a hyper-cubic parameter space[1]. High dimensional parameter spaces thus profit greatly by using hyperspherical priors which decrease the number of model evaluations in the low likelihood corner regions of parameter space drastically.

4.6.2 Latin hypercube approach

A second approch is to use the latin hypercube sampling (LHS) technique, shown in Fig. 4.6. Here, the parameter space is divided more finely, such that no two assigned samples share any parameter values. In two dimensions this method is equivalent to filling a chess board with rooks in such a way that no two of them threaten each other. Immediately, one of the shortcomings of the gridded parameter space is dealt with, in that the simulation need never be run at the same parameter value twice. The other main advantage of the LH is that its size does not increase exponentially with the dimension of parameter space. This property makes the LH the only feasible way of exploring high dimensional parameter spaces with ANNs (Urban & Fricker 2010).

We use a maximin distance design for our latin hypercube samples (Morris & Mitchell 1995). These designs try to simultaneously maximize the distance between all site pairs while minimizing the number of pairs which are separated by the same distance (Johnson et al. 1990). This maximin design for LHS prevents highly clustered sample regions and ensures homogeneous sampling. Prior knowledge of the behaviour of the power spectrum could also be used to identify the regions of parameter space where the power spectrum varies most rapidly and thus a higher concentration of samples should be imposed on such a region. Additionally, using a spherical prior region may help reducing the number of model evaluations used in the corners of parameter space where the likelihood is low (Kern et al. 2017).

For our training set comparisons we use 3 different LH training sets of size 100, 1000 and 10000 respectively.

[1] In 12 dimensions the proportion of the volume in the corners of a hypercube is $\sim 99.96\%$. That is the difference between the volume of the hypercube and that of an n-ball.

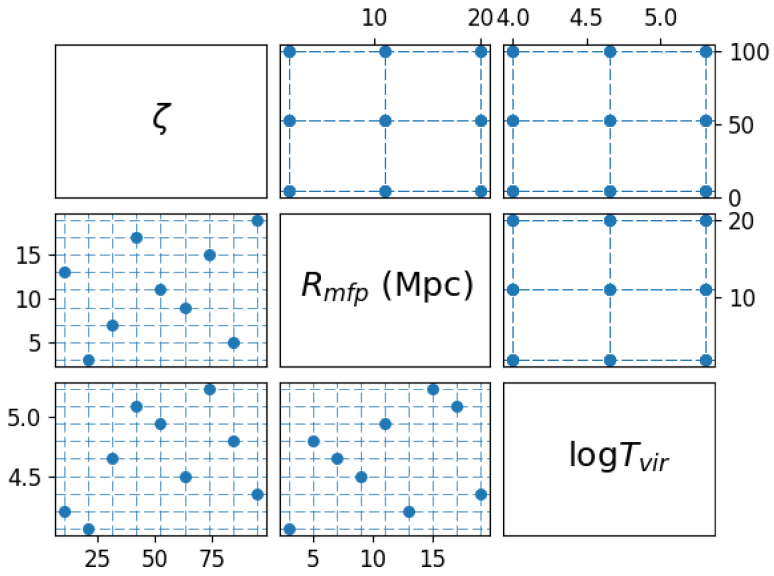

Figure 4.6: Visualization of the two training techniques. The parameter space is projected down to two dimensions in each plot. Top right: 27 regularly gridded parameters. Bottom left: 9 samples which are obtained using the latin hypercube sampling technique. Note that the number of samples are chosen such that the same number of projected samples are visible.

4.6.3 Power spectrum predictions

We now test the predictive power of our trained ANN. First, we define the mean square error between the true value of the power spectrum and an estimate given by the ANN,

$$\text{MSE} = \frac{1}{N_\text{p} N_\text{k}} \sum_{i=1}^{N_\text{p}} \sum_{j=1}^{N_\text{k}} \left(\frac{P_i^\text{true}(k_j) - P_i^\text{estimate}(k_j)}{P_i^\text{true}(k_j)} \right)^2 , \tag{4.31}$$

where N_p is the number of parameter combinations we estimate and compare, and N_k the number of k-bins used in the comparison. We produced a test set of 50 `21cmFast` power spectra at $z = 9$, sampled from a LH design to ensure a homogeneous spread in parameter space. This test set was then compared to a prediction from our ANN trained on three sizes of training sets, using 100, 1000 and 10000 samples distributed again using a LH design. We vary the training duration on each set and compare the predictions to the true values of the test set in Fig. 4.7. The error bars are obtained by selecting 75% of the total points in the training set at random for the network regression. The network is then trained on this subset and a value for the MSE is obtained. A new training sample is then selected at random and the process is repeated 10 times. The error bars thus signify the expected error from any given latin hypercube sampled training set of comparable size.

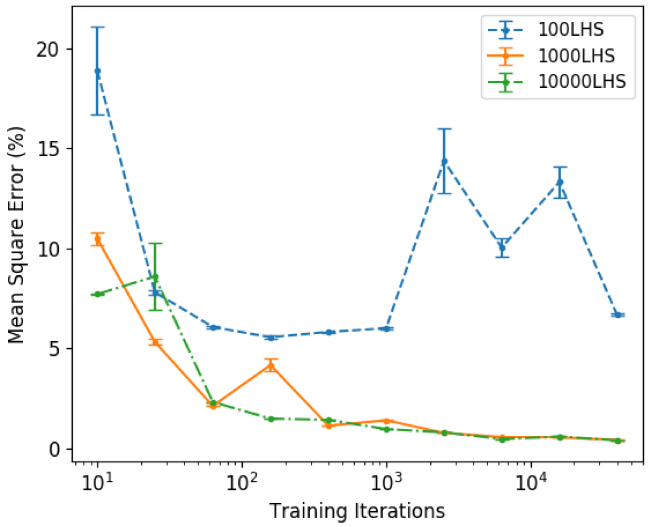

Figure 4.7: Mean square error of the neural network prediction compared to a fixed test set of 50 points at $z = 9$ as a function of the training iterations. At each number of training iterations, the training is repeated 10 times and we show the mean value of each resulting MSE and the variance on the mean as error bars. Shown are the behaviours for neural networks using 100 (blue), 1000 (orange) and 10000 (green) latin hypercube samples in the training set.

In the case of 10^3 and 10^4 samples in the training set, the neural network quickly approaches a relative mean square error of less than 1%. With more than 10^3 training iterations, both training sets show a clear reduction in the training efficiency. The 100 LHS curve is dominated at high training iterations by outlier parameter points which are particularly poorly constrained. We find that these outliers can affect the MSE heavily while having a relatively small effect on the final parameter inference. We define an outlier to be any k-bin whose square error is larger than 1, meaning a relative error of over 100%. For a training set of 100 points, one should then expect up to $\sim$ 2% of all k-bins to be outliers at any given training iteration. This unexpected behaviour may indicate an insufficient coverage of the training set, or that our neural network retains a high degree of flexibility even after regressing over 100 training samples. For our training set of 1000 points, the fraction of outliers produced reduces to less than $\sim$ 1%, when the training iterations are low, and we cease to find any outliers at more than 100 training iterations. This indicates a significant reduction in the freedom of the neural network and an increase of the confidence in our prediction. Of note is that some outliers have a greater impact than others and we find some whose square error $\sim$ 10, indicating a complete failure to predict the power in that particular k-bin. One should thus be cautious when using small training sets that may not sufficiently constrain the freedom of the neural network. Based

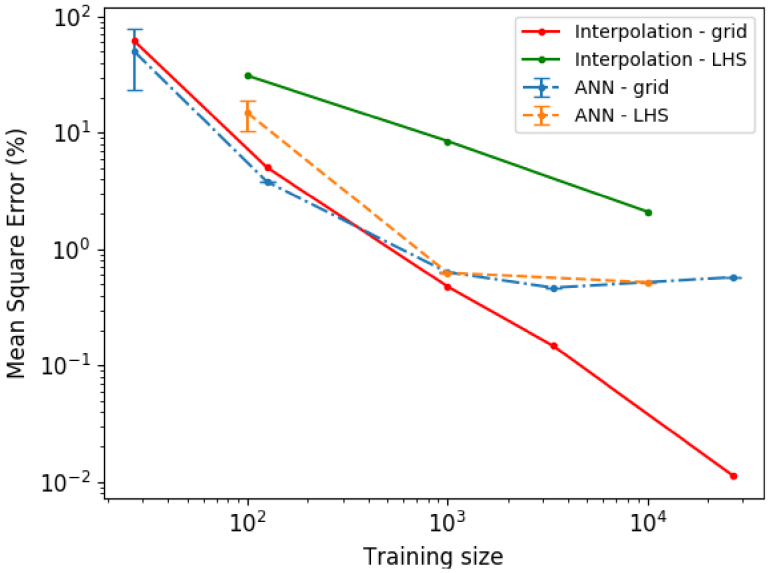

Figure 4.8: Comparison between the mean square error of interpolation on a grid (red solid line), the neural network using gridded training sets (blue dot-dashed line), interpolation (green solid line) and the neural network using LHC training sets (orange dashed line). Neural networks are trained using 10^4 training iterations. Plotted are the mean values after the NN is retrained 10 times, and the standard deviation to the mean is shown as error bars.

on the results for our two larger training sets, we proceed by using 10^4 training iterations in all our neural network training.

Further, we compare the mean square error between our training techniques against the training set size and sampling technique. In Fig. 4.8, we compare the mean square error in the prediction when the gridded parameter values are interpolated (red), or used to train our neural network (blue), with the predictions obtained when using a Latin hypercube sampled training set (green and orange). Similar to Fig. 4.7, we compute the mean and variance of the MSE over 10 separately trained networks by selecting 75% of the samples in the training set at random at a time.

As expected, when using a finer grid of parameters to interpolate the power spectrum, the accuracy of the prediction increases. Although the neural network predictions increase in accuracy for both the grid and the LHC, a clear plateauing in the addition of information by a larger training set can be observed. We thus observe a fundamental limit to the relative mean square error for the neural network design. This limit depends on the design parameters of the neural network and can be optimized via k-fold validation of the network's design or hyper-parameters. Varying the design parameters, such as the number of hidden layers or number of nodes per layer, and minimizing the mean square

error for a power spectrum prediction over k iterations can reduce networks error bound. Our network design limits errors at $\sim$ 1%, which is sufficiently below any confidence limit associated with our model, such that optimizing design parameters is of limited use. Optimisation via k-fold validation may be necessary when using fully numerical simulations which reflect a higher degree of physical accuracy than fast semi-numerical methods. No clear difference of the MSE can be seen comparing the latin hypercube sampled training sets and those produced on the grid in 3 dimensions and thus both methods do equally well in this case. We expect, although have not proven this explicitly, that a more significant discrepancy in higher dimensions of parameter space as discussed in section 4.6.2. In parameter inference studies it is common to vary large numbers of parameters. For example, in our application, one could add six cosmological parameters as well as three additional astrophysical parameters controlling for the epoch of heating (Greig & Mesinger 2017). The resulting 12 dimensional analysis would require training points in the same high dimensional space and a gridded training set would be impractically or even impossibly large, thus rendering the only feasible option a LH approach. It is also instructive to compare the performance of the interpolation on the grid to that on the LH, as one could conceivably just interpolate on the LHS in higher dimensions. The ANN manages to capture the information of the unstructured training data much better than simple interpolation does, whereas this is not necessarily the case for large gridded training sets. This result indicates a better performance for unstructured data by the ANN rather than interpolation and given that unstructured data are the only feasible way to access high dimensional model emulation, we argue that the ANN is a reliable way to perform this task. Similar conclusions were found in (Jennings et al. 2019), but we note that a future analysis including astrophysical parameters should check that this result holds for high dimensional analyses as the increased interpretability of interpolated models could be a boon over ANN emulation if the methods achieve similar levels of precision.

Fig. 4.9 to 4.11 show the predictions of a trained neural network (solid lines) and the true values of the power spectrum at the same point in parameter space (dashed lines). In order to determine the dependence of the accuracy of the predictions on the particular training set used, a subset of the training set is again randomly selected and used as the training set. Similar to before, the network is retrained 10 times while the predictions are averaged. The variance on the mean prediction in each k-bin is added as the expected error on the predicted mean value of the power spectrum. The power spectrum is dominated on small scales ($k > 1$ Mpc^{-1}) by shot noise and by foregrounds on large scales ($k < 0.15$ Mpc^{-1}). We therefore apply cuts at these scales in our analysis and indicate the noise dominated ranges by the grey shaded regions in Fig. 4.9, Fig. 4.10 and Fig. 4.11.

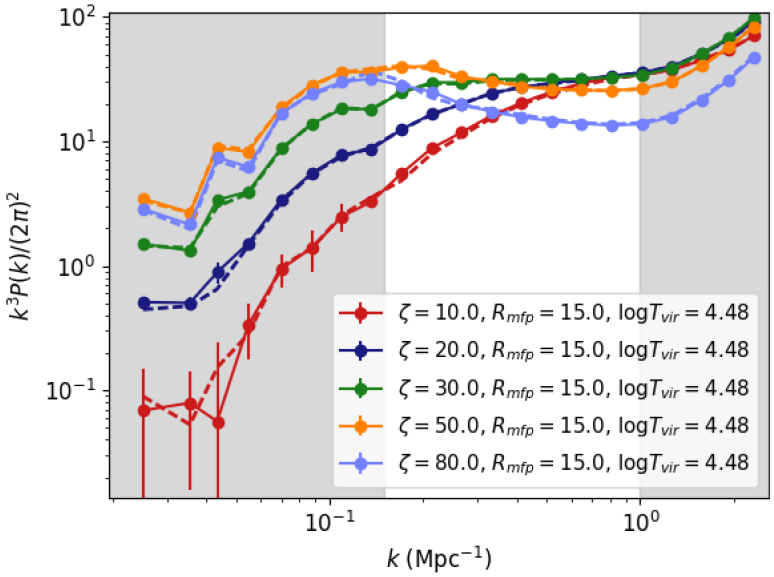

Figure 4.9: Comparison between neural network prediction of the 21cm power spectrum (solid line) and the `21cmFast` power spectrum (dashed line). We vary ζ at $z = 9$ from $\zeta = 10$ to $\zeta = 80$, and use 1000 training iterations on 75% of the 1000 LHS training set selected at random. This process is repeated 10 times and the mean values are shown with the variance on the mean as error bars.

We observe that the network produces a good fit to the true values within the region of interest. The size of the error bars indicates a very low dependence on the training subset used for training such that we conclude that the exact distribution of training sets in parameter space has little influence as long as it is homogeneously sampled. We also observe that the network manages to fit $T_{\rm vir}$ particularly well at large scales compared to the other two parameters whose error bars noticeably increase as k approaches the foreground cut-off. This shows that a sampling scheme that varies according to the dependence of the power spectrum on the input parameters may be advantageous to achieve some desired accuracy.

In the context of outliers, discussed earlier in this section, we see that the prediction for the power spectrum at $(\zeta, R_{\rm mfp}, \log T_{\rm vir}) = (30, 2, 4.48)$, in Fig. 4.11, overestimates the power at $k \approx 0.5$ Mpc^{-1} by a factor of ~ 2. This point would have a relatively large impact on the MSE as recorded in Fig. 4.7, even though the network is very well behaved for most regions in parameter space.

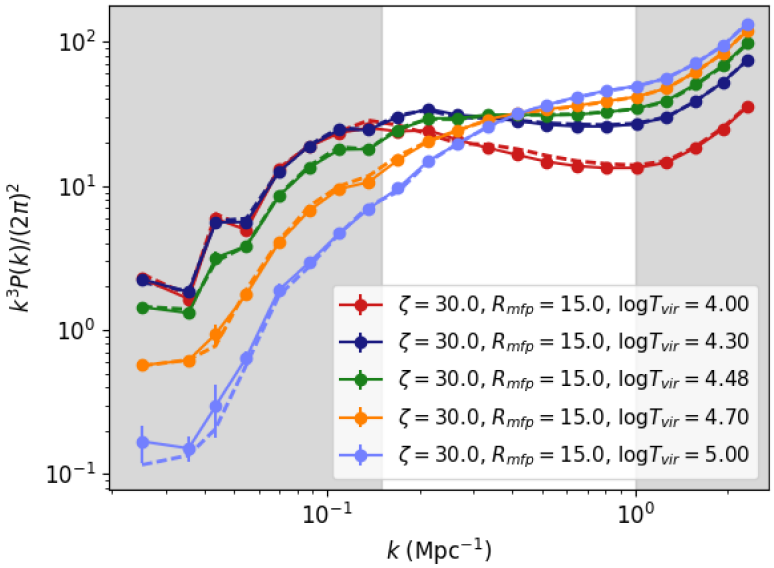

Figure 4.10: Comparison between neural network prediction of the 21cm power spectrum (solid line) and the `21cmFast` power spectrum (dashed line). We vary $T_{\rm vir}$ at $z = 9$ from $T_{\rm vir} = 10^4$ K to $T_{\rm vir} = 10^5$ K, similar to Fig. 4.9.

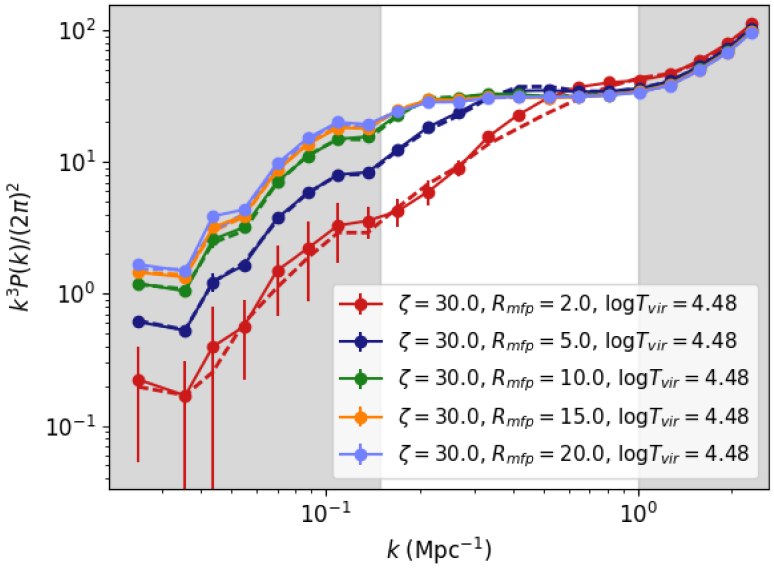

Figure 4.11: Comparison between neural network prediction of the 21cm power spectrum (solid line) and the `21cmFast` power spectrum (dashed line). We vary $R_{\rm mfp}$ at $z = 9$ from $R_{\rm mfp} = 2$ Mpc to $R_{\rm mfp} = 20$ Mpc, similar to Fig. 4.9.

4.7 Bayesian inference of astrophysical parameters.

In Bayesian parameter inference one is interested in the posterior distribution of the parameters $\boldsymbol{\theta}$ within some model $\mathcal{M}$. That is the probability distribution of the parameters given some dataset $\boldsymbol{x}$. We can then write Bayes' Theorem as

$$Pr(\boldsymbol{\theta}|\boldsymbol{x}, \mathcal{M}) = \frac{Pr(\boldsymbol{x}|\boldsymbol{\theta}, \mathcal{M})\pi(\boldsymbol{\theta}|\mathcal{M})}{Pr(\boldsymbol{x}|\mathcal{M})}, \tag{4.32}$$

to relate the posterior distribution $Pr(\boldsymbol{\theta}|\boldsymbol{x}, \mathcal{M})$ to the Likelihood, $\mathcal{L} \equiv Pr(\boldsymbol{x}|\boldsymbol{\theta}, \mathcal{M})$, the prior, $\pi(\boldsymbol{\theta}|\mathcal{M})$, and a normalisation factor called the evidence, $Pr(\boldsymbol{x}|\mathcal{M})$. This expression parametrises the probability distribution of the model parameters as a function of the likelihood, which, given a model and a dataset, can be readily evaluated under the assumption that the data points are independent and carry Gaussian errors,

$$\ln \mathcal{L} = -\frac{[\boldsymbol{x} - \boldsymbol{\mu}(\boldsymbol{\theta})]^2}{2\sigma_x^2} + C, \tag{4.33}$$

where C denotes a normalisation constant. In our case, the data will be a mock observation of the 21cm power spectrum, $\boldsymbol{x} = \{P_{\rm obs}(k_i)\}$, evaluated in 21 k-bins, the expectation value of the data will be the theoretical model prediction of the power spectrum, $\boldsymbol{\mu}(\boldsymbol{\theta}) = P(k, \boldsymbol{\theta})$, and for the variance on the data we assume that instrumental noise is the sole contributor characterised by a noise power spectrum, $\sigma_x^2 = P_{\rm Noise}(k)$.

4.7.1 Experimental design

We use `21cmSense`[2] (Pober et al. 2013; 2014) to compute the noise power spectrum for HERA331, with experimental details outlined in Beardsley et al. (2015) and summarized below. The noise power spectrum used is given by (Parsons et al. 2012)

$$P_{\rm Noise}(k) \approx X^2 Y \frac{k^3}{2\pi^2} \frac{\Omega'}{2t} T_{\rm sys}, \tag{4.34}$$

where X^2Y denotes a conversion factor for transforming from the angles on the sky and frequency to comoving distance, Ω' is the ratio of the square of the solid angle of the primary beam and the solid angle of the square of the primary beam, t is the integration time per mode, and $T_{\rm sys}$ is the system temperature of the antenna, which is given by the receiver temperature of 100 K plus the sky temperature $T_{\rm sky} = 60\,(\nu/300\ {\rm MHz})^{-2.55}$ K.

[2]Publically available at `https://github.com/jpober/21cmSense`.

As our experiment design, we assume a HERA design with 331 dishes distributed in a compact hexagonal array to maximize the number of redundant baselines, as HERA is optimized for 21cm power spectrum observations (DeBoer et al. 2017; Liu & Parsons 2016). Each dish has a diameter of 14 m, which translates into a total collecting area of ~ 50950 m^2. HERA antennas are not steered and thus use the rotation of the Earth to drift scan the sky. An operation time of 6 hours per night is assumed for a total of 1000 hours of integration time per redshift. We consider both single redshift and multiple redshift observations assuming a bandwidth of 8 MHz. Although experiments like HERA and the SKA will cover large frequency ranges $\sim 50 - 250$MHz, foregrounds can limit the bandpass to narrower instantaneous bandwidths.

4.7.2 Markov chain Monte Carlo

We aim to compare our parameter estimation runs to those of Greig & Mesinger (2015) by using the same mock and noise power spectrum for HERA331 as input for our neural network parameter search. Our fiducial parameter values are $\zeta = 30$, $R_{\rm mfp} = 15$ Mpc and $T_{vir} = 30000$ K.

First, we perform an independent parameter search in two redshift bins, $z = 8$ and $z = 9$, the latter comparing directly to Fig. 3 in Greig & Mesinger (2015). The fiducial values for the average neutral fraction at these redshifts are $\bar{x}_{\rm HI}(z = 8) = 0.48$ and $\bar{x}_{\rm HI}(z = 9) = 0.71$. For both the emulation and the 21CMMC runs we produce 2.1×10^5 points in the MCMC chain for a like-for-like comparison between the two techniques. This analysis uses the `emcee` python package for the Monte Carlo simulation (Foreman-Mackey et al. 2013).

Then, we analyse observations at redshifts $z = 8$, 9 and 10 by combining the information in these redshift bins. We take a linear combination of the χ^2 statistics in each redshift bin. Three separate ANNs are used for each redshift and are trained on the same training sets as for the individual redshift searches at $z = 8$ and 9. The fiducial neutral fraction for our final mock observation is $\bar{x}_{\rm HI}(z = 10) = 0.84$. A total of 2.1×10^5 are again obtained both in the neural network search and the equivalent 21CMMC run. The results following were then produced by combining `21cmFast`, the ANN and the MCMC sampler with the instructions above.

4.8 Discussion

Similar to Kern et al. (2017), we see a significant speed-up for the parameter estimation. For our fiducial chain size, we observe a speed up by 3 orders of magnitude for the sam-

Code - Training Set	z	ζ	$R_{\rm mfp}$	$\log T_{\rm vir}$
21CMMC	9	$41.28^{+24.85}_{-13.43}$	$13.38^{+4.28}_{-5.15}$	$4.59^{+0.37}_{-0.32}$
ANN - 100 LHS	9	$45.47^{+25.19}_{-17.18}$	$12.13^{+5.71}_{-5.05}$	$4.54^{+0.47}_{-0.28}$
ANN - 1000 LHS	9	$42.52^{+26.18}_{-13.74}$	$12.89^{+4.63}_{-5.29}$	$4.57^{+0.40}_{-0.31}$
ANN - 10000 LHS	9	$42.21^{+25.42}_{-14.12}$	$13.18^{+4.46}_{-5.14}$	$4.58^{+0.39}_{-0.31}$
21CMMC	8	$39.64^{+31.90}_{-16.11}$	$14.99^{+2.98}_{-3.64}$	$4.61^{+0.21}_{-0.23}$
ANN - 100 LHS	8	$43.06^{+26.16}_{-17.38}$	$14.58^{+3.47}_{-3.90}$	$4.64^{+0.19}_{-0.25}$
ANN - 1000 LHS	8	$42.71^{+31.30}_{-18.67}$	$14.67^{+3.19}_{-4.26}$	$4.62^{+0.21}_{-0.23}$
ANN - 10000 LHS	8	$39.78^{+31.68}_{-16.22}$	$14.61^{+3.15}_{-4.05}$	$4.60^{+0.22}_{-0.23}$
21CMMC	8,9,10	$31.08^{+8.70}_{-6.04}$	$15.15^{+2.86}_{-3.21}$	$4.51^{+0.17}_{-0.17}$
ANN - 100 LHS	8,9,10	$31.51^{+8.57}_{-6.32}$	$15.86^{+2.47}_{-3.62}$	$4.49^{+0.16}_{-0.19}$
ANN - 1000 LHS	8,9,10	$31.18^{+8.47}_{-6.08}$	$14.97^{+2.91}_{-3.78}$	$4.51^{+0.16}_{-0.17}$
ANN - 64 gridded	8,9,10	$32.46^{+13.90}_{-5.72}$	$12.52^{+3.47}_{-6.13}$	$4.61^{+0.11}_{-0.13}$
ANN - 125 gridded	8,9,10	$30.17^{+6.78}_{-5.04}$	$12.97^{+4.09}_{-3.69}$	$4.50^{+0.15}_{-0.16}$
ANN - 1000 gridded	8,9,10	$31.32^{+7.52}_{-5.20}$	$13.94^{+3.80}_{-4.68}$	$4.50^{+0.16}_{-0.16}$

Table 4.1: Median values and 68% confidence interval found in the parameter search via the brute-force method (21CMMC) and our ANN emulation at $z = 9$ and $z = 8$. The fiducial parameter values for both redshifts are given by $(\zeta, R_{\rm mfp}, \log T_{\rm vir}) = (30, 15, 4.48)$.

pling of the likelihood by emulation over the brute-force method. Our 21CMMC runtime of 2.5 days on 6 cores for a single redshift is reduced to 4 minutes using the emulator. In addition to the sampling, the neural network training requires on the order of $\sim$ 1 minute for 100 training samples to $\sim$ 1 hour for 10^4 training samples, which is not needed when evaluating the model at each point. Compared to the total runtime of 21CMMC the training time presents a minor factor.

4.8.1 Single redshift parameter constraints

Fig. 4.12 to 4.15 show the comparison between the brute-force parameter estimation as the red dashed contours and our ANN emulation using a variety of training set sizes at redshift $z = 9$ and $z = 8$ as the solid blue contours. For both redshifts, we show the one and two sigma contours obtained for 100 and 1000 LH samples as well as the marginalized posteriors convolved with a Gaussian smoothing kernel. As our posterior 1D marginalized parameter distributions are not found to be Gaussian, we compute the median and the 68% confidence interval defined by the region between the 16th and 84th percentile as our summary statistics in Table 5.3. We find excellent agreement between our method and 21CMMC for training sets of 10^3 and 10^4 samples at both redshifts, and good agreement with 100 samples.

We observe that errors retrieved by our network can be smaller than those obtained by 21CMMC, this is due to systematics. During the training period, our ANN constructs a model which approximates the `21cmFast` model and we proceed to sample the likelihood of the approximation. Therefore, assuming convergence of the chains, any difference between the recovered 68% confidence intervals are most likely due to systematic differences between the two models that are sampled. We estimate that we are subject to these systematic effects on the 1% - 10% level for large to small training sets, as per Fig. 4.7 and Fig. 4.8.

The $\zeta - \log T_{\rm vir}$ panels in Fig. 4.12 and Fig. 4.13 show that the neural network is sensitive to the same multi-modality found by 21CMMC, which is illustrated by the stripe feature at low $T_{\rm vir}$ and high ζ. This region represents less massive galaxies with a brighter stellar population, which can mimic our fiducial observation. Such a galaxy population would ionize the IGM earlier and thus by combining multiple redshifts and adding information about the evolution of the ionization process, this degeneracy ought to be lifted. Similarly, the $R_{\rm mfp} - \log T_{\rm vir}$ panel shows a clear bimodal feature for both 21CMMC and our neural network. Comparing to the results at $z = 8$ in Fig. 4.14 and Fig. 4.15, we see this multi-modal behaviour disappearing, which suggest that this degeneracy can be lifted by adding information in multiple redshift bins. Despite a clear downgrade of the fit to the brute-force method in the shape of both the 2D contours and the 1D marginalized posteriors, the training set using 100 samples still encloses the true parameter values of the observation in the 68% confidence interval as indicated in Table 4.1.

4.8.2 Multiple redshift parameter constraints

Fig. 4.16 and Fig. 4.17 show the contraints obtained when combining observations in three redshift bins at $z = 8$, 9 and 10 for training sets of 1000 and 100 samples per redshift respectively. As noted in the previous section, adding information about the evolution of the reionization process lifts some of the degeneracies in our recovered parameter constraints and both multi-modal features in the $\zeta - \log T_{\rm vir}$ and the $R_{\rm mfp} - \log T_{\rm vir}$ panels could be lifted. Of note is that combining multiple redshift bins highly improves the fit of the neural network trained on only 100 samples per redshift. We find that all our fiducial parameter values are well within the 68% confidence interval set out by the median and its 16th and 84th percentile for even this sparse training set.

Additionally, we compare the inference of a network trained on gridded training sets with similar sizes to our LH sampled training sets. Both 5^3 and 10^3 training sets recover similar constraints as the 100 LHS and 1000 LHS training sets, consistent with our findings in section 4.6. However, we observe a clear deterioration of the predictive

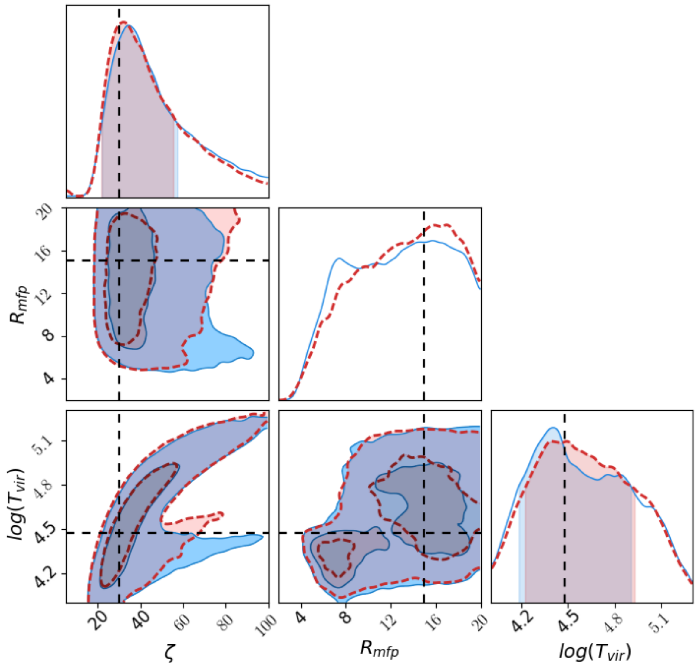

Figure 4.12: Comparison between the recovered 1σ and 2σ confidence regions of 21CMMC (red dashed lines) and the ANN emulator (blue solid lines) at $z = 9$. The ANN uses 1000 LHS for the training set and a 10^4 training iterations. The dotted lines indicate the true parameter values $(\zeta, R_{\rm mfp}, \log T_{\rm vir}) = (30, 15, 4.48)$.

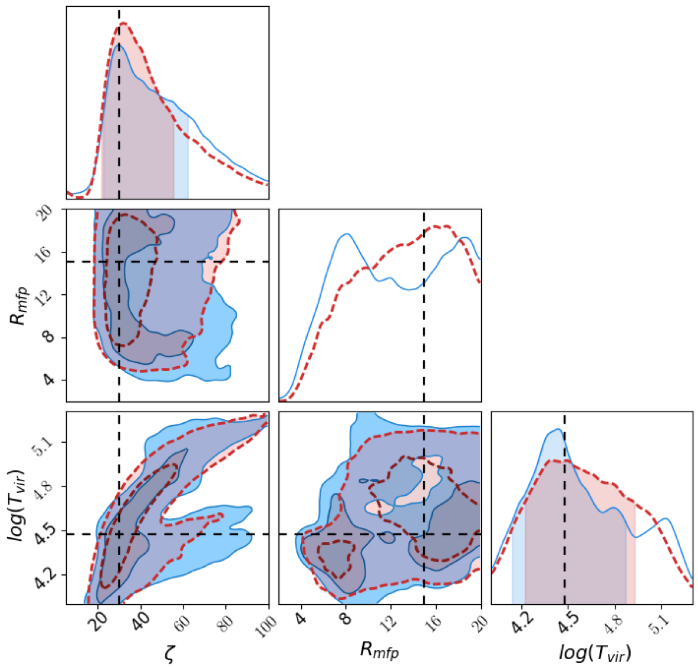

Figure 4.13: Comparison between the recovered 1σ and 2σ confidence regions of 21CMMC (red dashed lines) and the ANN emulator (blue solid lines) at $z = 9$. The ANN uses 100 LHS for the training set and a 10^4 training iterations. The dotted lines indicate the true parameter values $(\zeta, R_{\rm mfp}, \log T_{\rm vir}) = (30, 15, 4.48)$.

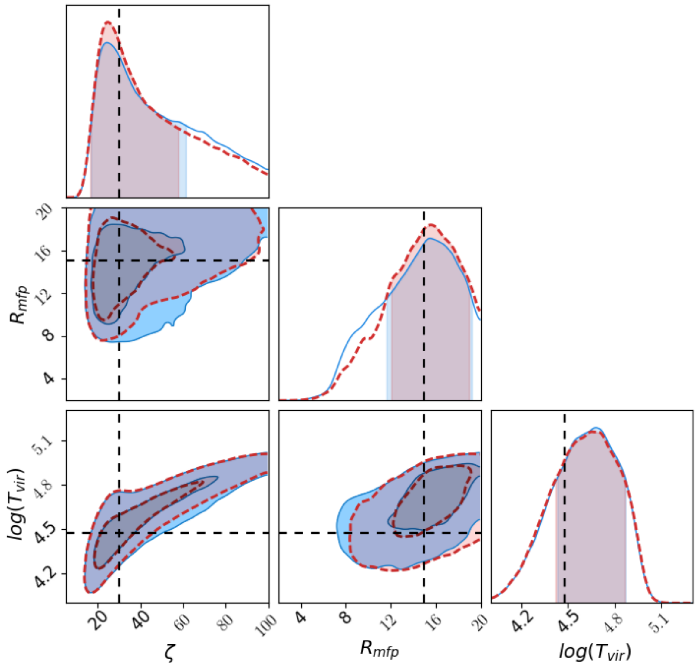

Figure 4.14: Comparison between the recovered 1σ and 2σ confidence regions of 21CMMC (red dashed lines) and the ANN emulator (blue solid lines) at $z = 8$. The ANN uses 1000 LHS for the training set and a 10^4 training iterations. The dotted lines indicate the true parameter values $(\zeta, R_{\rm mfp}, \log T_{\rm vir}) = (30, 15, 4.48)$.

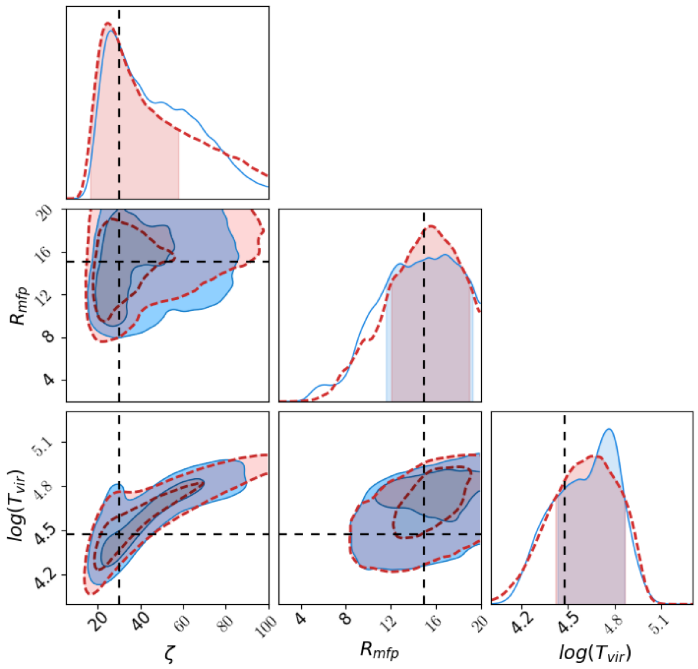

Figure 4.15: Comparison between the recovered 1σ and 2σ confidence regions of 21CMMC (red dashed lines) and the ANN emulator (blue solid lines) at $z = 8$. The ANN uses 100 LHS for the training set and a 10^4 training iterations. The dotted lines indicate the true parameter values $(\zeta, R_{\rm mfp}, \log T_{\rm vir}) = (30, 15, 4.48)$.

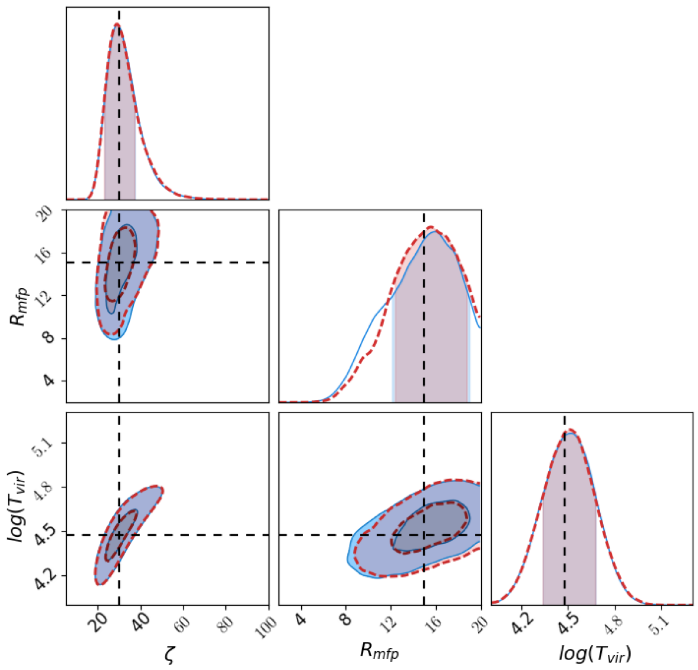

Figure 4.16: Comparison between the recovered 1σ and 2σ confidence regions of 21CMMC (red dashed lines) and the ANN emulator (blue solid lines) combining redshifts $z = 8$, $z = 9$, and $z = 10$. The ANN uses 1000 LHS for the training set at each redshift and a 10^4 training iterations. The dotted lines indicate the true parameter values $(\zeta, R_{\rm mfp}, \log T_{\rm vir}) = (30, 15, 4.48)$.

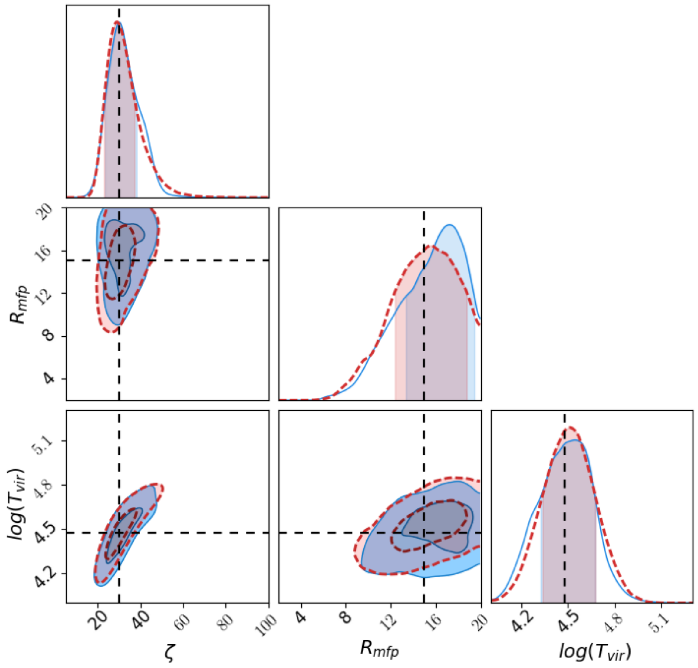

Figure 4.17: Comparison between the recovered 1σ and 2σ confidence regions of 21CMMC (red dashed lines) and the ANN emulator (blue solid lines) combining redshifts $z = 8$, $z = 9$, and $z = 10$. The ANN uses 100 LHS for the training set at each redshift and a 10^4 training iterations. The dotted lines indicate the true parameter values $(\zeta, R_{\rm mfp}, \log T_{\rm vir}) = (30, 15, 4.48)$.

power as we reduce the number of gridded training parameters to 4 points per parameter. Although the fiducial parameter values are recovered within the 16th to 84th percentile in Table 5.3, we fail to recover the fiducial values within the 2σ contours for 4^3 points.

4.8.3 Applications

With a speed-up of ~ 3 orders of magnitude, 21cm power spectrum emulation can be used for a variety of new or existing analyses, and we aim here to highlight some potential uses:

(*i*) 21cm experimental design studies (eg. Greig et al. 2015) use much the same principle as our model parameter inference outlined above. By varying the experimental layout or survey strategy, we effectively vary the noise power spectrum $P_{\rm Noise}(k)$ in equation (4.33), and can thus fit the optimal layout or survey strategy. These studies require fast model evaluations in order to be able to compare a multitude of survey strategies and experimental designs.

(*ii*) We find that using small training sets of 100 model evaluations, our emulation recovers parameter constraints to a similar degree of accuracy as those obtained when evaluating the model at each point in the chain. This may open up the possibility to move

away from semi-numerical models such as `21cmFast` and for the first time use radiative transfer codes (Ciardi et al. 2003; Iliev et al. 2006b; Baek et al. 2009; 2010) in EoR parameter searches. Semelin et al. (2017) have recently produced a first database of 45 evaluations of their radiative transfer code to provide 21cm brightness temperature lightcones evaluated on a 3D grid. The power spectra extracted from this database could be used as a training set for an ANN emulator. However, our analysis suggests that training sets with lower than 100 samples should be used with caution.

(*iii*) In addition to determining the best fit parameters of any given model, we would like to quantify the degree of belief in our model in the first place. Future data will be abundant, and as such we would like to be able to use it to inform us about the choice of model that best fits the data. Here too, the computational speed that emulation provides can be of use. Bayesian model comparison requires the computation of the evidence as the integral of the likelihood times the prior over all of parameter space. Nested sampling algorithms such as `MultiNest` (Feroz et al. 2009) provide an estimate for the evidence of a particular model together with the evaluation of the posterior, and thus benefits greatly from fast power spectrum computations.

(*iv*) The output nodes of the neural network treats each k-bin of the 21cm power spectrum separately. The weights of the trained network thus act to correlate the values in each k-bin according to the training set. There is therefore no restriction to predict other observables that are correlated to the 21cm power spectrum using the same emulator. The same network could thus encode the skewness or bispectrum of the 21cm fluctuations at the same time assuming the inclusion of these functions in the training sets.

4.9 Summary

With the advent of next generation telescopes such as MWA, HERA and the SKA, a first detection of the cosmic 21cm signal from the Epoch of Reionization is expected to be made within the next few years. One of the challenges accompanying this new generation of telescopes will be the vast amount of raw data generated. In this chapter we have introduced the machine learning paradigm as a crucial way to efficiently analyse these large amounts of incoming data. In particular, we have focussed on the cost reduction for expensive model evaluation in the context of EoR parameter inference from observations via model emulation. We show that emulating the models using artificial neural networks can speed up the model evaluations significantly, while maintaining a high degree of accuracy. We use an artificial neural network to train on a series of training sets which consist of 21cm power spectrum evaluations produced by the semi-numerical code `21cmFast`. As the limiting factor now becomes the creation of the training set, we study the evolution of

the error on the power spectrum predictions as a function of the training set size. We find that as few as 100 model evaluations may be sufficient to recover reasonable constraints on the parameters, especially when combining information across multiple redshift bins.

Machine learning techniques, such as the one presented in this chapter, will be crucial in extracting the most information possible from future 21cm observations. They are however by no means the only way to maximize our information gain and as such, in the following part we will introduce a number of analytical techniques which will similarly aid to fully characterize the information presented by the 21cm radiation.

Part IV

Analytic Analysis of the Cosmic 21cm Signal

"Scientific knowledge is a body of statements of varying degrees of certainty – some most unsure, some nearly sure, none absolutely certain."

– Richard P. Feynman.

Chapter 5

21cm Bispectrum

In this chapter, we utilize the inherent non-Gaussian nature of gravitational collapse to study the 21cm bispectrum at low redshifts ($z \sim 0.35 - 3$), targeted by upcoming neutral hydrogen intensity mapping experiments. We focus on an analytic derivation for the bispectrum due to gravitational collapse and a contribution by line of sight perturbations in the form of the lensing-ISW bispectrum and compare their amplitudes at the relevant redshifts. Within the next decade a significant number of next generation telescopes are expected to see their first light, capable of probing vast volumes of the low-redshift Universe. These telescopes are in a prime position for a first detection of the cosmic 21cm signal and should have the capabilities to detect both power spectrum and bispectrum of the 21cm brightness temperature fluctuations. The main result of this chapter is a Fisher forecast analysis of the bispectrum in the context of CHIME, MeerKAT and SKA. We find that the bispectrum proves to be a valuable source of cosmological information and has the potential to decrease errors on the cosmological parameters by an order of magnitude compared to *Planck*. Combining the information from both power spectrum and bispectrum of the observed sky yields the greatest constraining power. Finally, we compute the contribution of the lensing-ISW bispectrum, and find that, unlike for the cosmic microwave background analyses, it can safely be ignored for 21cm bispectrum observations.

The material for this chapter has been published in MNRAS (1 March 2019, Vol.483, Issue 3, pp. 4259-4275): *C. J. Schmit, A. F. Heavens and J. R. Pritchard; The gravitational and lensing-ISW bispectrum of 21cm radiation*. The published version is available through: `https://doi.org/10.1093/mnras/sty3400`.

5.1 Introduction

The cosmic microwave background (CMB) has truly propelled cosmology into an age of precision science. Over the last three decades, experiments such as COBE (Smoot et al. 1992), WMAP (Bennett et al. 2003; 2013), and Planck (Planck Collaboration et al. 2016b) have measured the CMB to an astonishing degree of accuracy, and its information is routinely combined with various other cosmological probes such as weak lensing, galaxy clustering and Type-1a supernovae. This great effort has allowed us to constrain many of the parameters of the geometrically flat, cold dark matter model with a cosmological constant (ΛCDM) to the percent level.

Although the *Planck* data favours a simple six parameter model over other models (Heavens et al. 2017), there remain tensions between the CMB measurements and local direct measurements of the Hubble parameter, h, (Bennett et al. 2014; Riess et al. 2016; 2018a;b), as well as low-redshift weak lensing measurements, which find slightly less matter clumping than expected from extrapolating the CMB findings (Heymans et al. 2013; MacCrann et al. 2015; Raveri 2016; Joudaki et al. 2017; Köhlinger et al. 2017). These tensions can arise if the assumed cosmological model is wrong since the CMB photons principally reveal the conditions of the Universe at the time of recombination at a relatively thin redshift slice at $z \simeq 1100$, when the Universe was matter dominated. Additional probes along the line of sight are required to give the full 3-dimensional context for the evolution of the Universe and study the evolution of low-redshift phenomena, such as dark energy. Galaxy surveys, such as the 2dF Galaxy Redshift survey (Colless et al. 2001), BOSS (Anderson et al. 2012), and SDSS (Ahn et al. 2014), are one such probe which determine the cosmological parameters by mapping the positions of galaxies in the sky and realizing that they are biased tracers of the underlying dark matter distribution. These surveys thus relate the galaxy power spectrum directly to the matter power spectrum from which the parameters can be determined. Weak lensing surveys, such as CFHTLenS (Heymans et al. 2012), KiDS (de Jong et al. 2013) and DES (Jarvis et al. 2016), present another low-redshift observation that complements the CMB observations, as the reconstructed lensing potential is directly related to the gravitational potential of the Universe. For both galaxy redshift surveys and weak lensing surveys it is crucial to obtain large galaxy samples by probing the largest observational volumes possible. One of the main difficulties for these surveys is to determine the redshift information of galaxies in their sample, as the largest volumes are attained by rapid photometry of the sources. Imprecise redshift information effectively blurs the radial information of the galaxies in the sample and propagates as a systematic error into the analysis.

Intensity mapping (IM) of the 21cm radiation at low redshifts possesses many advantageous properties which help to overcome the common challenges of conventional survey techniques and thus presents a promising new tool for observational cosmology. At late times most of the neutral hydrogen in the Universe has been ionized and the remaining atomic hydrogen resides mainly within self-shielded damped Lyman-α (DLA) systems inside galaxies and galaxy clusters. This means that the 21cm signal is a biased tracer of the dark matter density field and can therefore directly be connected to the cosmological parameters. When mapping the 21cm intensity, experiments integrate the emission of unresolved clouds of hydrogen gas within a given frequency bin. This provides CMB-like maps of the brightness temperature fluctuations in each frequency bin which each can be analysed for the statistical signal similar to the CMB. Due to the inherent relationship between the observed frequency of the 21cm signal and the redshift at which it was emitted, the signal readily provides spectroscopic redshifts and hence precise 3D information about the Universe. Although advances in optical surveys, such as DES, *Euclid*[1] and LSST[2], promise precise measurements of many aspects of dark energy and the galaxy distribution in the near future, IM surveys performed by experiments such as CHIME and SKA can be enormously advantageous. These surveys will best optical surveys in terms of survey volume and speed, and will be sensitive to the baryon acoustic oscillation (BAO) signature out to higher redshifts (Bull et al. 2018). Bull et al. (2015) have thoroughly examined the information gained from power spectrum observation of an extensive list of 21cm IM experiments and find competitive percent level forecasts on the cosmological parameters. However, due to the non-linear nature of structure formation, the power spectrum cannot fully probe the information content of the field, and an analysis of higher order statistics, such as the bispectrum, is warranted (Repp et al. 2015). Theoretical predictions of the 21cm bispectrum due to primordial non-Gaussianities and non-linear collapse from gravitational instability have only been made in the context of high redshift ($z > 50$) observations (Pillepich et al. 2007), giving promising predictions for the signal to noise for a bispectrum detection. Here we focus therefore on analysing the bispectrum as an additional observable in the context of upcoming IM experiments during the late stages of structure formation.

In addition, we examine another physical effect which can lead to a non-zero bispectrum and potentially contributes to the signal observed by IM experiments, the correlation between weak gravitational lensing and the integrated Sachs-Wolfe effect. As the 21cm emission travels towards our telescopes, it traverses the intergalactic medium and is subjected to the gravitational effects of the intervening matter. Matter fluctuations act as

[1] `https://www.euclid-ec.org/`
[2] `https://www.lsst.org/`

gravitational lenses on the 21cm photons, whose paths get distorted by their presence. This effect should be noticeable through the statistical distribution of the 21cm photons on the sky. In addition to this, as the Universe evolves into an acceleration-dominated era at low-redshifts ($z \lesssim 2$), the growth of structure lags behind the accelerated expansion of space. This effect causes the gravitational potentials of galaxy clusters to decrease in amplitude over time, resulting in a boost in energy for photons traversing those potentials. This late-time integrated Sachs-Wolfe effect (ISW) once again distorts the intensity distribution of photons in a survey volume. Cross-correlations between these two lines of sight effects improve cosmological parameter constraints from lensing surveys on the 10% level on large scales as shown by Zieser & Merkel (2016). Most importantly, however, ignoring the lensing-ISW (LISW) effect has been shown to bias CMB parameter inferences (Kim et al. 2013), and will at some level bias 21cm bispectrum observations. We compute both the LISW bispectrum and the bias resulting from neglecting it from upcoming IM experiments.

This chapter is organised as follows. Section 5.2 introduces the 21cm signal model used and we write down the angular power spectrum. In section 5.4, we revisit the 21cm bispectrum from Pillepich et al. (2007) and include a low-redshift 21cm signal model. The effects of lensing, the ISW effect, and the angular LISW bispectrum are discussed as well. We then compute both the 21cm bispectrum and the LISW bispectrum for all triangle configurations at $z = 1$. In section 5.5, we discuss upcoming intensity mapping experiments able to detect the 21cm bispectrum, and include foreground and noise models. Section 5.6 introduces the forecast model and determines the expected signal to noise for a LISW bispectrum detection as well as the bias introduced when neglecting it. Finally, we present and discuss the results of the parameter forecasts in section 5.8, before we summarize our findings in section 5.9. Throughout this chapter, we assume a six parameter ΛCDM cosmology with fiducial values $(\Omega_{\rm b}h^2, \Omega_{\rm CDM}h^2, \Omega_\Lambda, h, 10^9 \times A_{\rm s}, n_{\rm s}) = (0.022, 0.127, 0.684, 0.67, 1.562, 0.962)$.

5.2 The 21cm signal

The 21cm signal originates from the hyperfine ground state transition in the hydrogen atom. Its strength is governed by the relative abundance of HI atoms in the excited, triplet (1), state relative to the non-excited, singlet (0), state, parametrised through the spin temperature $T_{\rm S}$,

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} \exp\left(-\frac{T_*}{T_{\rm S}}\right), \tag{5.1}$$

where $T_* = h\nu_{21}/k_B \approx 68\text{mK}$, g_i is the statistical weight of the energy level i, $g_1/g_0 = 3$, and $T_S \gg T_*$. The intensity of the signal on the sky is then measured, and we model the signal in terms of its brightness temperature, which relates to the signal intensity via the Rayleigh-Jeans formula, $T_b(\nu) \approx I_\nu c^2/2k_B\nu^2$. Generally (see chapter 3), the 21cm signal is measured using the CMB as a background,

$$T_b(z) = \frac{T_S - T_\gamma(z)}{1+z}\tau, \tag{5.2}$$

where $T_\gamma(z)$ denotes the CMB temperature at redshift z and τ is the optical depth through a cloud of neutral hydrogen.

The spin temperature and thus T_b depend on the underlying HI density field as well as astrophysical effects, such that the brightness temperature is dependent on the position and can be split into a homogeneous and a fluctuating part,

$$T_b[\boldsymbol{r}(z), z] = \delta\bar{T}_b(z)\left\{1 + \delta_{HI}[\boldsymbol{r}(z), z]\right\}. \tag{5.3}$$

In the context of intensity mapping, we follow the model of Bull et al. (2015) and focus on the mean 21cm signal that is emitted by localised clumps of HI gas within galaxies and galaxy clusters for which the average brightness temperature over the sky can be approximated as (Santos et al. 2015)

$$\delta\bar{T}_b(z) \approx 566h\left[\frac{H_0}{H(z)}\right]\left[\frac{\tilde{\Omega}_{HI}(z)}{0.003}\right](1+z)^2\mu\text{K}. \tag{5.4}$$

Here, $\tilde{\Omega}_{HI}$ is the density of HI atoms in units of the current critical density (Camera et al. 2013; Villaescusa-Navarro et al. 2018),

$$\tilde{\Omega}_{HI}(z) \equiv \rho_{HI}(z)/\rho_{c,0}, \tag{5.5}$$

with a critical density today, $\rho_{c,0} = 3H_0^2/8\pi G$. The density of neutral hydrogen is related to the mass of the dark matter halos in the Universe,

$$\rho_{HI}(z) = \int_{M_{min}}^{M_{max}} dM \frac{dn}{dm} M_{HI}(M), \tag{5.6}$$

where dn/dm is the halo mass function, for which we use a simple Sheth-Tormen implementation, and $M_{HI}(M)$ is the HI mass in a halo of mass M. Following Bagla et al. (2010), we adopt a lower cutoff for the mass range containing HI gas, to correspond to a circular halo velocity of 30 km/s, meaning that halos with a lower circular velocity do not contain any HI gas. Typically, neutral hydrogen can be expected in star forming halos, and

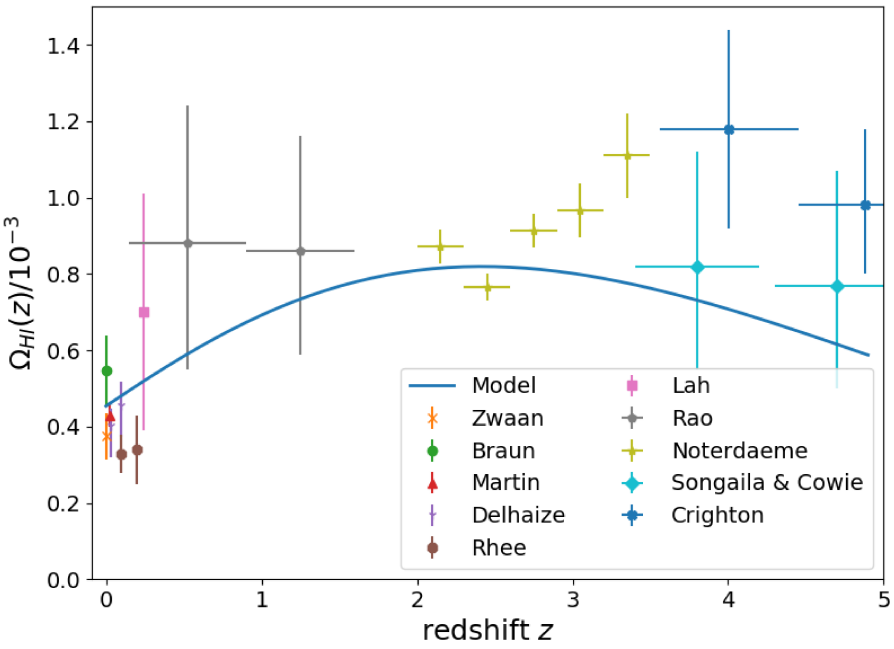

Figure 5.1: Comparison between our analytic model for the HI density, $\tilde{\Omega}_{\rm HI}$, as a function of redshift with current measurements. Included are the results from Zwaan et al. (2015); Braun (2012); Martin et al. (2010); Delhaize et al. (2013); Rhee et al. (2013); Lah et al. (2007); Rao et al. (2006); Noterdaeme et al. (2012); Songaila & Cowie (2010); Crighton et al. (2015). See Crighton et al. (2015) for full data list.

gas in halos with circular velocities of larger than 60 km/s can be expected to form stars. Additionally, self-shielding damped Lyman-α systems can be found in lower mass halos, justifying a somewhat lower velocity cut-off. The HI mass density is measured locally using 21cm emission (Zwaan et al. 2005; Martin et al. 2010) and at higher redshifts via damped Lyman-α systems, as they trace the HI distribution after the EoR (Prochaska et al. 2005). Crighton et al. (2015) summarize recent measurements of $\tilde{\Omega}_{\rm HI}$, and we compare the analytic model with these observations in Fig. 5.1. Villaescusa-Navarro et al. (2018) simulate the behaviour of $\tilde{\Omega}_{\rm HI}$ and find good agreement with the observations.

5.3 The 3D angular power spectrum

Similarly to CMB experiments, fluctuations of the 21cm brightness temperature on the sky allow us to construct an angular power spectrum. 21cm experiments have a smooth frequency response around a central observed frequency ν, which we model with a Gaussian window function $W_\nu(z)$, such that the observed brightness temperature fluctuation on

the sky can be written as

$$\delta T_{\rm b}^{\rm obs}(\hat{\boldsymbol{n}}, \nu) = \int dz W_\nu(z) \delta T_{\rm b}[r(z)\hat{\boldsymbol{n}}, z]. \tag{5.7}$$

The quantity $\delta T_{\rm b}^{\rm obs}$ thus denotes the observed temperature field projected onto the sky in a frequency bin labelled by ν. As seen before, the brightness temperature fluctuations depend on the underlying HI density field. At late times, most of the neutral hydrogen is located in self-shielded gas clouds inside galaxies, which means that the hydrogen density field is a biased tracer of the dark matter density field,

$$\delta T_{\rm b}[r(z)\hat{\boldsymbol{n}}, z] = \delta \bar{T}_{\rm b}(z)\{1 + b_{\rm HI}(z)\delta[r(z)\hat{\boldsymbol{n}}, z]\}. \tag{5.8}$$

In most of our analysis we are only concerned with the first order term as the monopole term is inaccessible through interferometry. Similarly to Battye et al. (2013) and Bull et al. (2015) we assume the bias to be a constant at low-redshifts. For our computations we fix $b_{\rm HI} = 2$, which is consistent with DLA observations (Font-Ribera et al. 2012; Hall et al. 2013). To first order in perturbation theory, the density fluctuations simply grow as a function of the growth factor,

$$\delta[r(z)\hat{\boldsymbol{n}}, z] = D_+(z)\delta(\boldsymbol{r}). \tag{5.9}$$

We then Fourier transform the density fluctuations, and suppress the explicit z dependence in our notation for simplicity,

$$\delta(\boldsymbol{r}) = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} \tilde{\delta}(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{r}}, \tag{5.10}$$

and subsequently expand the Fourier modes in spherical harmonics,

$$e^{i\boldsymbol{k}\cdot\boldsymbol{r}} = 4\pi \sum_{\ell m} i^\ell j_\ell(kr) Y_{\ell m}(\hat{\boldsymbol{k}}) Y^*_{\ell m}(\hat{\boldsymbol{n}}). \tag{5.11}$$

We find

$$\begin{aligned} \delta T_{\rm b}^{\rm obs}(\hat{\boldsymbol{n}}, \nu) =& 4\pi \sum_{\ell m} i^\ell \int dz W_\nu(z) \delta \bar{T}_{\rm b}(z) b_{\rm HI}(z) D_+(z) \\ & \times \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} \tilde{\delta}(\boldsymbol{k}) j_\ell[kr(z)] Y_{\ell m}(\hat{\boldsymbol{k}}) Y^*_{\ell m}(\hat{\boldsymbol{n}}). \end{aligned} \tag{5.12}$$

Using the definition of the harmonic transform of the signal on the sky in terms of multipole moments ℓ and m,

$$a^\nu_{\ell m} = \int d^2\hat{\boldsymbol{n}} \delta T_{\rm b}^{\rm obs}(\hat{\boldsymbol{n}}, \nu) Y_{\ell m}(\hat{\boldsymbol{n}}), \tag{5.13}$$

we can use the orthonormality condition for spherical harmonics (see Appendix C.1) to obtain

$$\begin{aligned} a^{\nu}_{\ell m} =& 4\pi i^{\ell} \int dz W_{\nu}(z) \delta\bar{T}_{\rm b}(z) b_{\rm HI}(z) D_{+}(z) \\ & \times \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} \tilde{\delta}(\boldsymbol{k}) j_{\ell}[kr(z)] Y_{\ell m}(\hat{\boldsymbol{k}}). \end{aligned} \tag{5.14}$$

Now, the angular 21cm power spectrum, C_{ℓ}, is defined in terms of ensemble average of two harmonic coefficients,

$$\left\langle a^{\nu_1}_{\ell m} a^{*\nu_2}_{\ell' m'} \right\rangle = \delta^{\rm K}_{\ell\ell'} \delta^{\rm K}_{mm'} C_{\ell}(\nu_1, \nu_2), \tag{5.15}$$

where $\delta^{\rm K}$ denotes the Kronecker delta function and we assume statistical isotropy. Combining equations (5.14) and (5.15), in conjunction with the Fourier space matter power spectrum relation,

$$\left\langle \tilde{\delta}(\boldsymbol{k}) \tilde{\delta}(\boldsymbol{k}') \right\rangle = (2\pi)^3 \delta^{\rm D}(\boldsymbol{k} + \boldsymbol{k}') P(k), \tag{5.16}$$

where δ^D is the Dirac delta function, we find the angular power spectrum to be

$$\begin{aligned} C_{\ell}(\nu_1, \nu_2) =& \frac{2}{\pi} \int dz W_{\nu_1}(z) \delta\bar{T}_{\rm b}(z) b_{\rm HI}(z) D_{+}(z) \\ & \int dz' W_{\nu_2}(z') \delta\bar{T}_{\rm b}(z') b_{\rm HI}(z') D_{+}(z') \\ & \times \int dk k^2 P(k) j_{\ell}[kr(z)] j_{\ell}[kr(z')]. \end{aligned} \tag{5.17}$$

For large ℓ we can use the Limber approximation (see Loverde & Afshordi 2008; equation (C.30)) such that the angular power spectrum becomes diagonal in frequency and reduces to

$$C_{\ell}(\nu) = b^2_{\rm HI} \int dz \left[\frac{W_{\nu}(z) \delta\bar{T}_{\rm b}(z) D_{+}(z)}{r(z)} \right]^2 \frac{P\left[\frac{\ell+1/2}{r(z)}\right]}{|r'(z)|}. \tag{5.18}$$

We compute the matter power spectrum, $P(k)$, using CAMB[3], and our results for the 21cm angular power spectrum are illustrated in Fig. 5.2, including our noise and foreground models described in section 5.5. These results were derived in Battye et al. (2013) and we rederived them here using consistent notation with the following analysis.

5.4 Angular 21cm bispectrum

At low-redshifts ($z \sim 1$), targeted by upcoming IM experiments, the dark matter density field has become non-Gaussian mainly due to the non-linear gravitational collapse of

[3]Publicly available at: `https://camb.info/`.

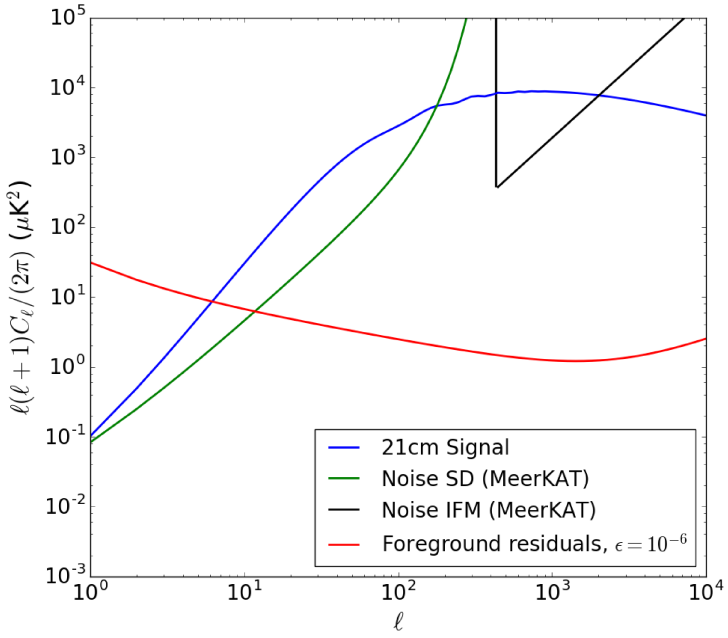

Figure 5.2: Angular 21cm power spectrum, noise and foreground residuals at $z = 1$. We show the noise curves for MeerKAT operated in single-dish (SD, green) mode as well as in interferometer (IFM, black) mode. Foreground residuals are plotted for a removal efficiency of $\epsilon = 10^{-6}$.

structure. As such, we expect the 21cm signal to contain a non-zero bispectrum. The nature of the bispectrum provides a radical increase of observable modes as compared to the power spectrum and thus presents a promising probe for cosmology. Non-Gaussianity can be added to the signal through multiple channels, and here we focus on two main effects that contribute to the 21cm bispectrum. In addition to the non-Gaussianities due to structure formation, denoted as $B^{\rm NLG}$, if the primordial density fluctuations are non-Gaussian, then that non-Gaussianity permeates through to late times as a contribution to the 21cm signal, which we write as $B^{\rm PNG}$. Furthermore, line of sight effects due to the gravitational distortion of light around massive objects and the accelerated expansion of the Universe, specifically via the integrated Sachs-Wolfe effect (Sachs & Wolfe 1967), induce a non-Gaussian contribution to the signal, denoted as $B^{\rm LISW}$. We can then write the total angular 21cm bispectrum as the sum of these contributing effects. Let $\alpha \equiv (\ell_1, \ell_2, \ell_3, m_1, m_2, m_3)$, then

$$B_\alpha^{\rm total} = B_\alpha^{\rm NLG} + B_\alpha^{\rm LISW} + B_\alpha^{\rm PNG}. \tag{5.19}$$

5.4.1 Non-linear gravity bispectrum

The bispectrum due to non-linear gravitational collapse of structure in the context of 21cm brightness temperature fluctuations can be calculated similarly to that in the context of galaxy surveys (See Fry (1984) and Appendix C.2 for details). The brightness temperature fluctuations are sourced by the fluctuations in the HI field, which is a biased tracer of the DM field (see equation (5.8)). The bispectrum is then defined by the Fourier transform of the 3-point function,

$$\begin{aligned} B^{21}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, z_1, z_2, z_3) &= \left\langle \delta\tilde{T}_\mathrm{b}(\boldsymbol{k}_1)\delta\tilde{T}_\mathrm{b}(\boldsymbol{k}_2)\delta\tilde{T}_\mathrm{b}(\boldsymbol{k}_3)\right\rangle \\ &= b_\mathrm{HI}^3\delta\bar{T}_\mathrm{b}(z_1)\delta\bar{T}_\mathrm{b}(z_2)\delta\bar{T}_\mathrm{b}(z_3)\left\langle\tilde{\delta}(\boldsymbol{k}_1, z)\tilde{\delta}(\boldsymbol{k}_2, z)\tilde{\delta}(\boldsymbol{k}_3, z)\right\rangle, \end{aligned} \tag{5.20}$$

where we assume a linear bias. Expanding the density perturbations to second order and applying Wick's theorem, the lowest order contribution to the bispectrum is (Pillepich et al. 2007; Appendix C.2)

$$\begin{aligned} B^{21}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, z_1, z_2, z_3) =& b_\mathrm{HI}^3 2\mathcal{K}(\boldsymbol{k}_1, \boldsymbol{k}_2)D_+^2(z_1)D_+(z_2)D_+(z_3) \\ &\times \delta\bar{T}_\mathrm{b}(z_1)\delta\bar{T}_\mathrm{b}(z_2)\delta\bar{T}_\mathrm{b}(z_3)P(\boldsymbol{k}_1)P(\boldsymbol{k}_2) + \text{cycl.}, \end{aligned} \tag{5.21}$$

where we define

$$\mathcal{K}(\boldsymbol{k}_1, \boldsymbol{k}_2) \equiv A_0 + A_1\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right)\cos\theta_{12} + A_2\cos^2\theta_{12}, \tag{5.22}$$

with $A_0 = 5/7$, $A_1 = 1/2$, $A_2 = 2/7$, and θ_{12} denotes the angle between $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$.

We can express the signal in harmonic space using equations (5.7) and (5.13). Taking the ensemble average of three harmonic coefficients yields the angular bispectrum. Using the methods developed in Verde et al. (2000) and Pillepich et al. (2007), as outlined in Appendix C.3, we compute the contribution to the angular 21cm bispectrum from the non-linear growth of structure to be

$$B_{\ell_1\ell_2\ell_3}^{\mathrm{NLG},m_1m_2m_3}(z) = B_{\ell_1\ell_2\ell_3}(z)\begin{pmatrix}\ell_1 & \ell_2 & \ell_3\\ m_1 & m_2 & m_3\end{pmatrix}, \tag{5.23}$$

where the parentheses denote the Wigner-3J symbol, which ensures that the triangle condition is met, expresses isotropy, and is akin to the Kronecker delta in 3D space. The bispectrum is non-zero if and only if,

(a) $-\ell_i \leq m_i \leq \ell_i$, for $i = 1, 2, 3$,

(b) $m_1 + m_2 = -m_3$,

(c) $|\ell_i - \ell_j| \leq \ell_k \leq \ell_i + \ell_j$, for all permutations of $(i, j, k) = (1, 2, 3)$,

(d) $\ell_1 + \ell_2 + \ell_3$ is a non-zero integer unless $m_1 = m_2 = m_3 = 0$.

Further, we can write the bispectrum as a sum of cyclic terms,

$$B_{\ell_1\ell_2\ell_3}(z) = B_{12}(z) + B_{13}(z) + B_{23}(z), \tag{5.24}$$

where

$$\begin{aligned} B_{12}(z) = \frac{16}{\pi} i^{\ell_1+\ell_2} \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{(4\pi)^3}} b_{\rm HI}^3 \int dz_1 dz_2 dz_3 dk_1 dk_2 k_1^2 k_2^2 \\ \times P(k_1)P(k_2)W_\nu(z_1)W_\nu(z_2)W_\nu(z_3)D_+^2(z_1)D_+(z_2)D_+(z_3)\delta\bar{T}_{\rm b}(z_1)\delta\bar{T}_{\rm b}(z_2)\delta\bar{T}_{\rm b}(z_3) \\ \times j_{\ell_1}[k_1 r(z_1)] j_{\ell_2}[k_2 r(z_2)] \sum_{\ell\ell'\ell''} i^{\ell'+\ell''}(-1)^\ell \beta_\ell(k_1,k_2)(2\ell'+1)(2\ell''+1) \\ \times j_{\ell'}[k_1 r(z_3)] j_{\ell''}[k_2 r(z_3)] \begin{Bmatrix} \ell_1 & \ell_2 & \ell_3 \\ \ell'' & \ell' & \ell \end{Bmatrix} \begin{pmatrix} \ell_1 & \ell' & \ell \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_2 & \ell'' & \ell \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_3 & \ell' & \ell'' \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned} \tag{5.25}$$

Here $\nu \equiv \nu(z)$, and we sum $\ell = 0, 1, 2$, $\ell' = \ell_1 - \ell, ..., \ell_1 + \ell$, and $\ell'' = \ell_2 - \ell, ..., \ell_2 + \ell$ and the braces denote the Wigner-6J symbol (eg. Sobelman 1979). The $\beta_\ell(k_1, k_2)$ functions connect to (5.22) such that

$$\beta_0 = 2A_0 + \frac{2}{3}A_2, \beta_1 = 2A_1\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right), \text{and } \beta_2 = \frac{4}{3}A_2. \tag{5.26}$$

This result has been found by Pillepich et al. (2007) and rederived here using consistent notation for our low-z temperature model in equation (5.4). In order to simplify this expression such that the implementation is practical and can be used in our forecasting model, we use the Limber approximation and compute the three contributing ℓ terms separately in Appendix C.4,

$$B_{12}^{\ell=0} = b_{\rm HI} A_{\ell_1\ell_2\ell_3}^{\ell=0} \int dz W_\nu(z)\delta\bar{T}_{\rm b}(z) D_+^2(z)\theta_{\ell_1}(z)\theta_{\ell_2}(z), \tag{5.27a}$$

$$B_{12}^{\ell=1} = b_{\rm HI} \sum_{\ell'\ell''} A_{\ell_1\ell_2\ell_3}^{\ell=1,\ell'\ell''} \int dz W_\nu(z)\delta\bar{T}_{\rm b}(z) D_+^2(z)\left[\theta_{\ell_1\ell'}^1(z)\theta_{\ell_2\ell''}^{-1}(z) + \theta_{\ell_1\ell'}^{-1}(z)\theta_{\ell_2\ell''}^1(z)\right], \tag{5.27b}$$

$$B_{12}^{\ell=2} = b_{\rm HI} \sum_{\ell'\ell''} A_{\ell_1\ell_2\ell_3}^{\ell=2,\ell'\ell''} \int dz W_\nu(z)\delta\bar{T}_{\rm b}(z) D_+^2(z)\theta_{\ell_1\ell'}(z)\theta_{\ell_2\ell''}(z), \tag{5.27c}$$

where the A^ℓ and the θ-functions are defined in equations (C.32), (C.36), (C.38), (C.42), (C.44), and (C.47). We have applied the Limber approximation here as the direct evaluation of equation (5.25) was found to be impractical. There is however evidence to suggest

that the Limber approximation does not remain accurate for narrow z-bins required for intensity mapping (Di Dio et al. 2018), and fast implementations for oscillatory function integration (Spurio Mancini et al. 2018) would be necessary to increase the accuracy of our calculation.

5.4.2 Lensing-ISW bispectrum

The presence and evolution of the gravitational potential along the line of sight affects the 21cm radiation and imprints statistical information about the state of the matter distribution on the signal. Firstly, the photon paths are disturbed by the presence of gravitational wells, resulting in a weak lensing contribution to the signal. The lensing potential, θ, for a source at distance r and at an angular position $\hat{\boldsymbol{n}}$ is a radial projection of the gravitational potential, Φ, (Bartelmann & Schneider 2001). In the Born approximation

$$\theta(r, \hat{\boldsymbol{n}}) = -\frac{2}{c^2} \int_0^r dr' \frac{S_k(r - r')}{S_k(r) S_k(r')} \Phi(r', \hat{\boldsymbol{n}}), \tag{5.28}$$

where S_k is determined by the curvature, and defined as

$$S_k(r) = \begin{cases} \sqrt{k}^{-1} \sin(r\sqrt{k}), & k > 0, \\ r, & k = 0, \\ \sqrt{|k|}^{-1} \sinh(r\sqrt{|k|}), & k < 0. \end{cases} \tag{5.29}$$

Observations of the weak lensing signal should be feasible by upcoming 21cm experiments and can help map the evolution of the growth function (Pourtsidou & Metcalf 2014).

A second line of sight effect, sourced by the gravitational potential, affects the 21cm photons. Due to the accelerated expansion of the Universe at late times, potential wells evolve on timescales shorter than the crossing time for photons. Therefore, photons that enter the gravitational well obtain a boost in energy, which is higher than the required energy to leave the well due to the decay of the potential while crossing. This results in an overall frequency gain which is additive along the photon's path. The frequency change due to this integrated Sachs-Wolfe (ISW) effect can be written as (Nishizawa 2014)

$$\frac{\Delta\nu}{\nu}(r, \hat{\boldsymbol{n}}) = \frac{2}{c^3} \int_0^r dr' \frac{\partial \Phi(r', \hat{\boldsymbol{n}})}{\partial t}, \tag{5.30}$$

where t denotes the conformal time.

These line of sight effects perturb the apparent radial and angular position of the brightness temperature signal on the sky, $\delta T_\mathrm{b} = \delta T_{\mathrm{b},0}(\hat{\boldsymbol{n}} + \nabla\theta, \nu + \Delta\nu)$, where $T_{\mathrm{b},0}$ is the

true, unperturbed signal, $\Delta\nu$ represents the frequency shift introduced by the ISW effect. Expanding this signal to first order in the gravitational potential gives

$$\delta T_{\rm b} = \delta T_{\rm b,0} + \nabla\delta T_{\rm b,0}\cdot\nabla\theta + \nu\frac{d\delta T_{\rm b,0}}{d\nu}\frac{\Delta\nu}{\nu}. \tag{5.31}$$

Considering a thin frequency shell, each term can be expanded in terms of multipole moments ℓ and m on the sky via equation (5.13). Thus the total coefficients separate into contributions from the signal when line of sight effects are ignored, indicated by the superscript "0", the lensing gradient and the ISW frequency shift,

$$a^{\nu}_{\ell m} = a^{0,\nu}_{\ell m} + a^{\rm L,\nu}_{\ell m} + a^{\rm ISW,\nu}_{\ell m}. \tag{5.32}$$

The lensing coefficient is given by (Appendix C.5)

$$a^{\rm L,\nu}_{\ell m} = \sum_{\ell' m'\ell'' m''} W^{mm'm''}_{\ell\ell'\ell''} a^{0,\nu *}_{\ell' m'}\theta^{\nu *}_{\ell'' m''}, \tag{5.33}$$

where $W^{mm'm''}_{\ell\ell'\ell''}$ relates to the gaunt integral, $\mathcal{H}$ (cf. Verde & Spergel 2002), via

$$W^{mm'm''}_{\ell\ell'\ell''} \equiv \frac{1}{2}(-1)^{m+m'+m''} L_{\ell\ell'\ell''}\mathcal{H}^{mm'm''}_{\ell\ell'\ell''}, \tag{5.34}$$

with

$$L_{\ell\ell'\ell''} \equiv -\ell(\ell+1) + \ell'(\ell'+1) + \ell''(\ell''+1). \tag{5.35}$$

Taking the ensemble average of three harmonic coefficients, we note that the line of sight terms in equation (5.32) are linear in the potential along the line of sight. The potential along the line of sight can be taken to be uncorrelated to the density field at source, such that linear terms vanish in the bispectrum and only second-order terms remain,

$$\langle a^{\nu}_{\ell_1 m_1} a^{\nu}_{\ell_2 m_2} a^{\nu}_{\ell_3 m_3}\rangle = \langle a^{0,\nu}_{\ell_1 m_1} a^{0,\nu}_{\ell_2 m_2} a^{0,\nu}_{\ell_3 m_3}\rangle + \sum_{\substack{\ell' m' \\ \ell'' m''}} W^{m_1 m' m''}_{\ell_1\ell'\ell''} \left\langle a^{*0,\nu}_{\ell' m'}\theta^{*,\nu}_{\ell'' m''} a^{0,\nu}_{\ell_2 m_2} a^{\rm ISW,\nu}_{\ell_3 m_3}\right\rangle + \text{perms.} \tag{5.36}$$

The first term is akin to the bispectrum due to non-linear gravitational collapse as discussed in the previous section and the remaining terms make up the contribution due to the line of sight effects. The ISW and lensing effects are uncorrelated to the undisturbed signal, as the initial photon distribution from a distant source is not affected by any effects that distort this signal on the line of sight. This allows us to separate the LISW contributions from the 21cm angular power spectrum in our expression for the bispectrum. We find

$$\langle a^{\nu}_{\ell_1 m_1} a^{\nu}_{\ell_2 m_2} a^{\nu}_{\ell_3 m_3}\rangle_{\rm LISW} = W^{m_1 m_2 m_3}_{\ell_1\ell_2\ell_3} C_{\ell_2}(\nu) Q_{\ell_3}(\nu) + 5 \text{ perms.}, \tag{5.37}$$

where we have applied statistical isotropy to relate the 2-point statistics to the power spectra,

$$\left\langle a^{*0,\nu}_{\ell m} a^{0,\nu}_{\ell' m'} \right\rangle = C_\ell(\nu)\delta^K_{\ell\ell'}\delta^K_{mm'}, \tag{5.38}$$

$$\left\langle \theta^{*\nu}_{\ell m} a^{\mathrm{ISW},\nu}_{\ell' m'} \right\rangle = Q_\ell(\nu)\delta^K_{\ell\ell'}\delta^K_{mm'}. \tag{5.39}$$

The LISW power spectrum is given by (cf. Verde & Spergel 2002) (see Appendix C.6)

$$Q_\ell(\nu) = \frac{2\eta(z)}{c^4}\int_0^z dz' \frac{S_k[r(z)-r(z')]}{S_k[r(z)]S_k[r(z')]r^2(z')}\left.\frac{\partial P_\Phi}{\partial z'}(k,z')\right|_{k=\ell/r(z')}, \tag{5.40}$$

where

$$\eta(z) = -(1+z)\frac{d\delta\bar{T}_\mathrm{b}}{dz}(z). \tag{5.41}$$

Although the derivation of the LISW in the context of the 21cm signal is new, the LISW bispectrum is a contamination to the primordial non-Gaussianity bispectrum observed on the CMB in both temperature and E-mode polarization (Goldberg & Spergel 1999; Giovi et al. 2003; Lewis et al. 2011), and, if ignored, introduces a significant bias in the non-Gaussianity parameter f_NL as measured by the *Planck* mission using the skew-C_ℓ statistic (Munshi & Heavens 2010). Planck Collaboration et al. (2016a) report a 2.8σ detection of the LISW bispectrum from temperature maps alone, which increases to a 3σ detection including their polarization data. We therefore compute the amplitude of the 21cm LISW bispectrum and its effect as a contamination on the signal due to non-linear gravitational collapse.

5.4.3 Primordial bispectrum

Primordial non-Gaussianity (PNG) in the density fluctuations are the most direct way to probe inflationary physics. Depending on the functional form of the inflaton field, PNG can be generated during inflation (see Bartolo et al. (2004) and Liguori et al. (2010) for extensive reviews). The most accurate measurements of PNG to date (Planck Collaboration et al. 2016a), are consistent with perfectly Gaussian initial fluctuations. However, with errors of order $\sigma_{f_\mathrm{NL}} \sim 5-40$, depending on the triangle shape, the CMB cannot constrain the non-Gaussianity parameter on the $f_\mathrm{NL} \lesssim 1$ level, crucial for eliminating a variety of inflationary models such as models that include an early contraction phase (Komatsu et al. 2009). As any PNG would affect the distribution of dark matter in the Universe, and thus that of baryons, a contribution to the 21cm bispectrum is expected. Pillepich et al. (2007) compute the angular bispectrum from PNG during the dark ages and compare it to the bispectrum from non-linear collapse. They find that the primordial bispectrum is ~ 50 times weaker than the gravitational bispectrum at large scales, but a

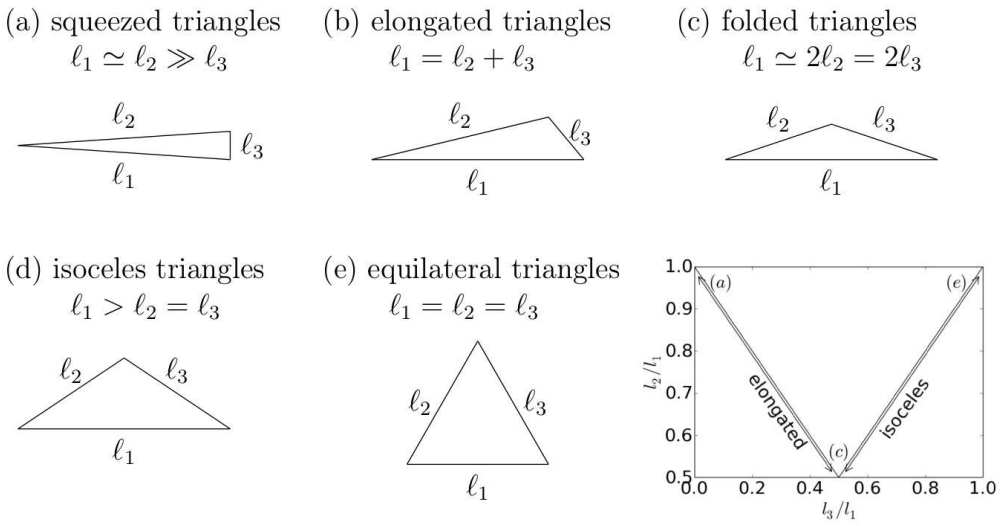

Figure 5.3: We fix the order of the bispectrum modes to be $\ell_1 \geq \ell_2 \geq \ell_3$. Then, (a) - (e) show the relation between triangle configurations and the bispectrum modes they represent. When plotting the bispectrum as a function of the ratios of ℓ_2/ℓ_1 vs. ℓ_3/ℓ_1, different triangle configurations separate into different areas of the plot as shown in the lower right. The bispectrum occupies a triangular shaped region which is due to the triangle condition obeyed by the bispectrum.

cosmic variance limited experiment could produce competitive, $\sigma_{f_{\rm NL}} \sim 1$, results. Moreover, Muñoz et al. (2015) study the 21cm bispectrum from PNG during the dark ages and find that 21cm observations can improve CMB constraints for PNG significantly due to the the high number of observable modes. They predict that a cosmic variance limited experiment would be able to measure $f_{\rm NL}$ down to $\sigma_{f_{\rm NL}} \sim 0.03$, and thus begin to be able to constrain single-field slow-roll inflation (Maldacena 2003; Acquaviva et al. 2003), which predicts $f_{\rm NL} = 10^{-2}$, thus ruling out a large number of exotic inflation scenarios. Observations of PNG at lower redshifts rely on the scale dependence of the halo bias and can achieve competitive constraints for the primordial non-Gaussianity parameter (Mao et al. 2013; D'Aloisio et al. 2013; Li & Ma 2017; Raccanelli et al. 2017; Karagiannis et al. 2018). The prospects of constraining PNG with the cosmic 21cm signal are thus promising.

In this analysis, we focus on the information gain toward the cosmological parameters from the late-time 21cm bispectrum. As the gravitational bispectrum dominates the bispectrum during the dark ages (Pillepich et al. 2007), the primordial bispectrum will remain sub-dominant at late times due to the progression of structure formation. We will therefore ignore the PNG contribution to the bispectrum here.

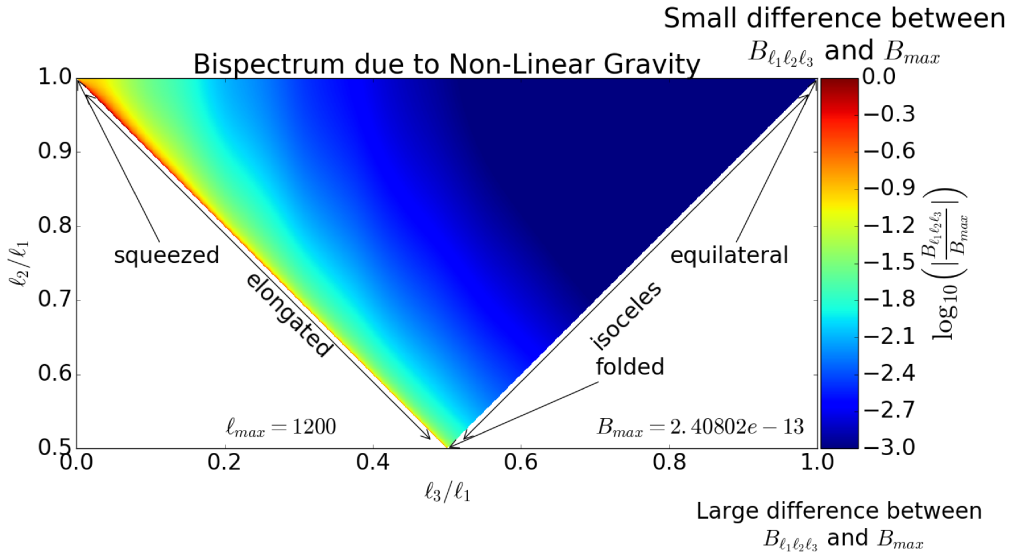

Figure 5.4: We plot the amplitude of the angular bispectrum due to non-linear gravity collapse for $\ell_{\rm max}$ at $z = 1$. The colour scale shows the order of magnitude difference in amplitude of the bispectrum relative to the triangle configuration with the largest bispectrum amplitude. An elongated squeezed triangle configuration shows the largest amplitude with $B_{\rm max} = 2.40802 \times 10^{-13}{\rm mK}^3$.

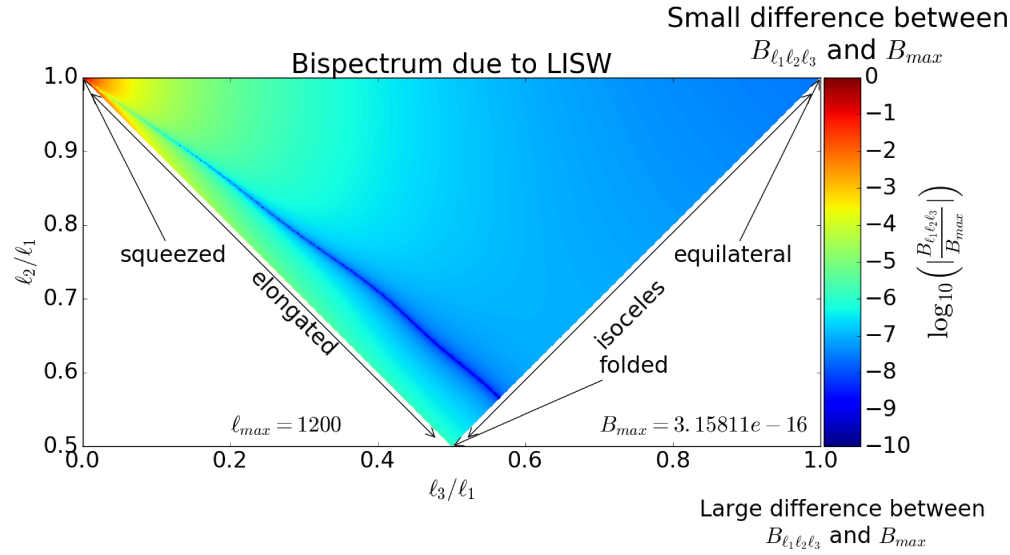

Figure 5.5: We plot the amplitude of the angular LISW bispectrum for $\ell_{\rm max}$ at $z = 1$. The colour scale shows the order of magnitude difference in amplitude of the bispectrum relative to the triangle configuration with the largest bispectrum amplitude. The squeezed triangle configuration show the largest amplitude with $B_{\rm max} = 3.15811 \times 10^{-16}{\rm mK}^3$.

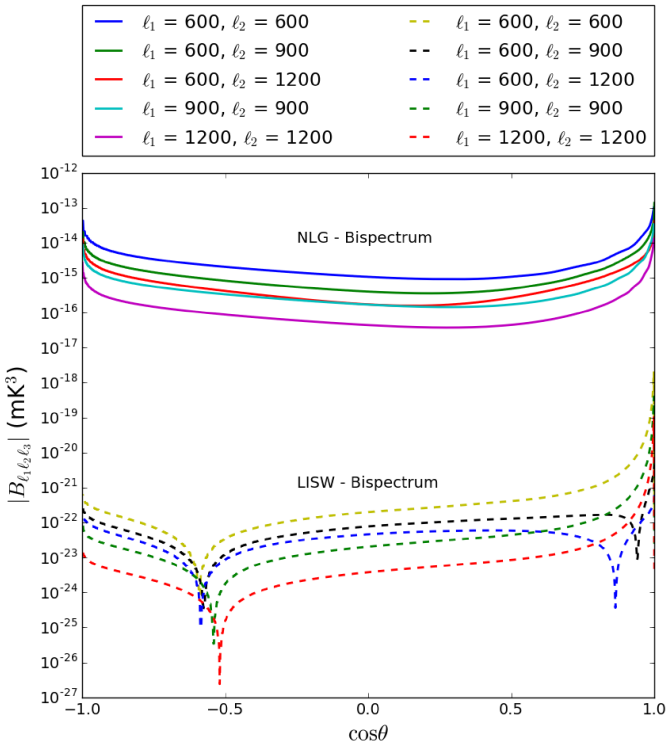

Figure 5.6: Plot of the NLG (solid lines) and LISW (dashed lines) bispectrum as a function of the opening angle θ between two fixed triangle sides ℓ_1 and ℓ_2 at $z = 1$.

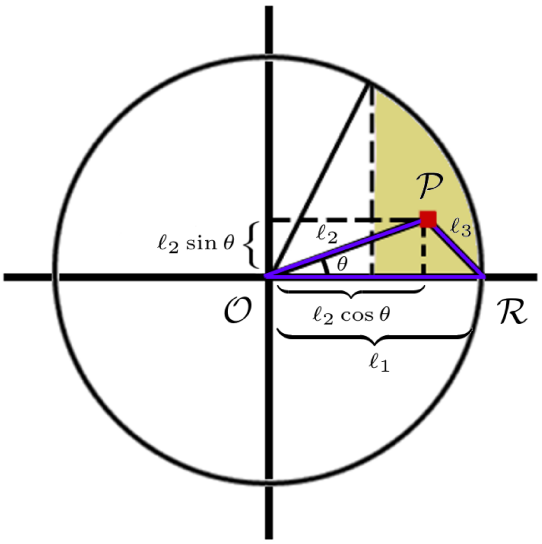

Figure 5.7: This figure illustrates the interpretation of Fig. 5.8. All unique triangles with $\ell_1 \geq \ell_2 \geq \ell_3$ can be constructed when $\mathcal{P}$ lies within the shaded region, enclosed by the circle of radius ℓ_1, the horizontal diameter of the circle, and the vertical line intersecting the circle at $\theta = \pi/3$. Each pixel value in Fig. 5.8 corresponds to the bispectrum of the triangle configuration with $\ell_2 - \ell_3$ corner in the same location.

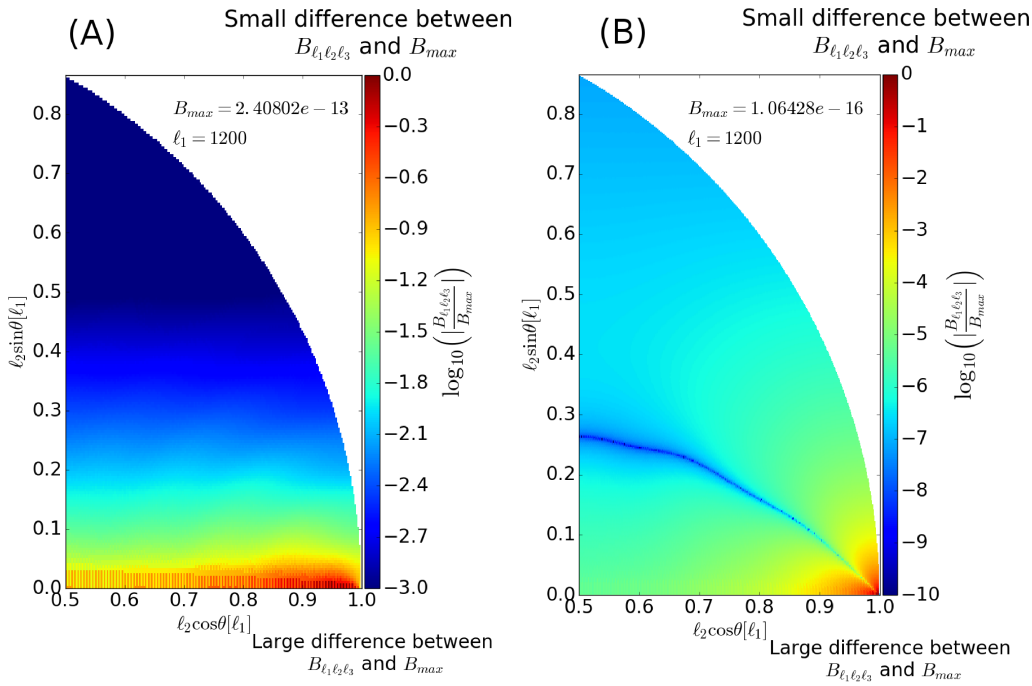

Figure 5.8: We plot the amplitude of the angular (A) non-linear gravity and (B) LISW bispectrum for $\ell_{\rm max} = 1200$ at $z = 1$. The x and y axes are in units of ℓ_1. The colour scale shows the order of magnitude difference in amplitude of the bispectrum relative to the triangle configuration with the largest bispectrum amplitude. For (A), the non-linear gravity bispectrum, the elongated and squeezed triangle configurations show the largest amplitude with $B_{\rm max} = 2.40802\times10^{-13}{\rm mK}^3$. For (B), the LISW bispectrum, the squeezed triangle configuration shows the largest amplitude with $B_{\rm max} = 1.06428 \times 10^{-16}{\rm mK}^3$.

5.4.4 Bispectrum representation

The bispectrum can be represented geometrically as a correlation of the signal from the corners of a triangle where the length of the sides is related to the wavenumber of the bispectrum. For our computation of the bispectrum we use the triangular representation of Jeong & Komatsu (2009) shown in Fig. 5.3. We set $\ell_1 \geq \ell_2 \geq \ell_3$, and fix ℓ_1, while varying ℓ_2 and ℓ_3. Plotting the ratios to the largest ℓ-mode against each other results in a triangular plot where squeezed bispectrum configurations occupy the upper left corner, equilateral configurations occupy the upper right corner, folded triangles are in the triangle peak, and elongated and isoceles triangles occupy the sides of the triangle. We show the relative amplitudes of our bispectrum calculations for $\ell = 1200$ at $z = 1$ for the NLG and the LISW bispectrum in Fig. 5.4 and 5.5 respectively. In order to visualize the overall trend of the bispectrum as a function of triangle configuration, we interpolate the bispectrum between neighbouring pixels as the statistical isotropy of the signal requires the sum of modes to be even and thus renders every other pixel zero.

We see for both cases that most of the signal is coming from squeezed or quasi-squeezed triangle configurations. The non-linear gravity bispectrum shows a large contribution from elongated triangles and three orders of magnitude lower contributions from equilateral triangle configurations. Similarly, the LISW bispectrum experiences almost no contributions from equilateral triangles. The dark blue stripe in Fig. 5.5 is due to a sign flip of the bispectrum and the bispectrum approaches zero for triangle configurations close to the feature.

An alternative representation with a single degree of freedom is presented in Majumdar et al. (2018), where the bispectrum is plotted for two fixed side lengths as a function of the opening angle of the triangle. Fig. 5.6 shows our results of NLG and LISW bispectra as a function of the opening angle for two fixed side lengths and illustrates the large amplitude difference of ~ 7 orders of magnitude between the non-linear gravity and LISW bispectrum for these modes.

Here, we also propose a new representation for the bispectrum which gives a direct visual connection to the triangle configuration at each point. The sketch in Fig. 5.7 shows the interpretation of this representation. Fig. 5.8 shows our 'sail' plots of the bispectrum for NLG and LISW which contain the bispectrum values for all unique triangle shapes. We fix the longest side of a triangle to be the horizontal radius of a circle of length ℓ_1, $\overline{\mathcal{OR}}$. When labelling the second largest side ℓ_2, we require $\ell_2 \cos\theta \in [\frac{1}{2}\ell_1, \ell_1]$ and $\ell_2 \sin\theta \in [0, \frac{\sqrt{3}}{2}\ell_1]$ to construct all possible unique triangles, as any others are obtained through rotation and relabelling of the sides. Now, for each point $\mathcal{P}$ in the shaded region, we compute the bispectrum of the corresponding triangle configuration and show the result as a colour scale at that point. We thus produce a colour map, where the $x - y$

coordinates are identical to the coordinates of the point $\mathcal{P}$ of the corresponding triangle, allowing for a direct and natural interpretation of the map.

Fig. 5.8 panel (A) shows the same behaviour as Fig. 5.4, where the largest bispectrum is obtained by squeezed triangles, close to the x-axis, and the lowest in the equilateral limit. Further, Fig. 5.8 panel (B) can be interpreted in the same way and compared to Fig. 5.5. Whereas the triangle plots allow the bispectrum for a given triplet of modes (ℓ_1, ℓ_2, ℓ_3) to be read directly, connecting regions of the plot with particular triangle shapes can be cumbersome. On the one hand, the triangle shape corresponding to any given pixel value in our 'sail' plots can directly be read off by constructing a triangle according to Fig. 5.7. On the other hand, reading the corresponding (ℓ_1, ℓ_2, ℓ_3) triplet may not be straightforward, as $\ell_2 = \sqrt{x^2 + y^2}$, and $\ell_3 = \sqrt{\ell_1^2 + x^2 + y^2 - 2\ell_1 x}$.

For both triangle plots and 'sail' plots the full bispectrum information is only obtained when stacking the plots for all different values of ℓ_1. We have included figures at $\ell_1 = 1200$ as an example of the value of the bispectrum.

5.5 Instrument and foreground assumptions

In this section we explore instruments and foregrounds which will both limit the detectability and sensitivity of the quantities derived up to this point.

5.5.1 Instruments

We examine three different experiments in this analysis: CHIME, MeerKAT and SKA.

CHIME[4]*:* The Canadian Hydrogen Intensity Mapping Experiment (CHIME), based in British Columbia, is an interferometer consisting of four 100 x 20 metre semi-cylinders equipped with radio receivers sensitive to 400MHz - 800MHz ($z \sim 0.8 - 2.5$). This experiment is a dedicated low-redshift 21cm intensity mapping experiment targeting BAO scales, with applications in FRB detection and pulsar monitoring. We select this telescope as a currently operational intensity mapping experiment, with the potential for late-time 21cm signal detection.

MeerKAT[5]*:* MeerKAT is an array of sixty-four 13.5 metre dishes located in the Karoo desert in South Africa. The dishes are equipped with three separate receivers, with the low-frequency band going from 580 MHz to 1015 MHz ($z \sim 0.4 - 1.4$). This SKA precursor will eventually be fully integrated into SKA-MID. We select MeerKAT as it

[4]see `https://chime-experiment.ca/`

[5]see `http://www.ska.ac.za/science-engineering/meerkat/`

is a near-future 21cm experiment with the potential to do low-redshift intensity mapping and is a precursor of SKA-mid.

SKA-MID[6]*:* The SKA Mid-Frequency Aperture Array is the South African part of the multi-purpose Square Kilometre Array and will consist of 190 15 metre dishes, at the MeerKAT location in the Karoo desert. SKA-Mid will be able to perform intensity mapping in both single-dish (autocorrelation) and interferometer mode. We focus on the 350 MHz - 1050 MHz ($z \sim 0.35 - 3$) range which SKA-MID will be operating at. SKA-mid is selected to illustrate the degree to which observations may constrain cosmology over the next decade.

5.5.2 Instrumental noise

CHIME, MeerKAT and SKA-MID can all be operated in an interferometric mode, where the noise power spectrum can be modelled as (Zaldarriaga et al. 2004; Pourtsidou & Metcalf 2014)

$$C_\ell^{\rm N} = \frac{T_{\rm sys}^2 (2\pi)^2}{\Delta\nu t_{\rm o} f_{\rm cover}^2 \ell_{\rm max}^2}, \tag{5.42}$$

where $T_{\rm sys}$ is the system temperature of the dishes in the array, t_o is the total observing time and $\ell_{\rm max} = 2\pi D_{\rm tel}/\lambda$ determines the largest multipole moment accessible by an array with diameter $D_{\rm tel}$ at an observed wavelength λ. The covering fraction, $f_{\rm cover}$, is the ratio of the collecting area, $A_{\rm coll}$, to the physical area covered by the array, st. $f_{\rm cover} = A_{\rm coll}/[\pi(D_{\rm tel}/2)^2]$. The system temperature is given as the sum of the angle-averaged sky temperature and the temperature of the antenna, $T_{\rm sys} = T_{\rm ant} + T_{\rm sky}$. At the frequencies considered in our analysis, the system temperature is dominated by $T_{\rm ant}$ at $\sim 30 - 50$K, such that we assume $T_{\rm sys} \approx T_{\rm ant}$. Interferometers cannot resolve scales larger than those set via the minimal baseline, which we model as a sharp noise increase at $\ell < \ell_{\rm min} = 2\pi D_{\rm min}/\lambda$.

MeerKAT and SKA-MID can also be operated in single-dish mode, for which the thermal noise per beam is given via (Olivari et al. 2018)

$$\sigma_{\rm t} = \frac{T_{\rm sys}}{\sqrt{t_{\rm pix}\Delta\nu}}, \tag{5.43}$$

where $T_{\rm sys}$ is the system temperature of the dishes, $\Delta\nu$ is the frequency binwidth and $t_{\rm pix}$ is the integration time per beam. The integration time per beam is obtained by distributing

[6] see `https://www.skatelescope.org/mfaa/`

the total integration time t_o, across N_d dishes,

$$t_{pix} = N_d t_o \frac{\theta^2_{FWHM}}{S_{area}}, \tag{5.44}$$

where $\theta^2_{FWHM} = \pi^2/\ell^2_{max}$ is the beam area, which sets the smallest scale that can be observed by each dish as $\ell_{max} = 2\pi D_{dish}/\lambda$, where we use the diameter of the dish, D_{dish}, as opposed to the diameter of the array in the interferometer case, and S_{area} denotes the survey area. Further, the signal is suppressed by the beam when angular wavenumbers exceed the resolution of the instrument. We model this beam suppression as an exponential increase in the noise as a function of ℓ,

$$C^N_\ell = e^{\sigma^2\ell^2} C^{N,thermal}_\ell, \tag{5.45}$$

with $\sigma^2 = \theta^2_{FWHM}/8\ln 2$, setting the scale of signal suppression. The full noise power spectrum for a single-dish array is then given by (eg. Dodelson 2003)

$$C^N_\ell = \sigma^2_t \theta^2_{FWHM} e^{\frac{\theta^2_{FWHM}\ell^2}{8\ln 2}}. \tag{5.46}$$

The noise parameters for each experiment are listed in table 5.1. Experiments will likely bin their observations into bins with $\Delta\nu \leq 1$MHz (Pourtsidou et al. 2016), however decreasing the window width for bispectrum observations increases the computation run-time to levels of impracticality, such that we take a conservative bin width of $\Delta\nu = 10$MHz. As the target emission is sourced by the discrete galaxy population, a shot noise contribution is expected. Chang et al. (2008) find the shot noise contribution in the case of post-EoR intensity mapping observations to be negligible, and we ignore this term here.

5.5.3 Foregrounds

Cosmological 21cm observations suffer from large foreground contaminations from both galactic and extragalactic sources. Successful detections of the signal hinge strongly on the accurate modelling and removal of these contaminations which can be 4 to 6 orders of magnitude larger than the signal (Liu et al. 2009; Alonso et al. 2014). Here, we model four contaminating foreground sources (Santos et al. 2005): extragalactic point sources, extragalactic free-free emission, galactic synchrotron emission and galactic free-free emission. Due to the smooth frequency variation of these foregrounds, a variety of foreground removal strategies have been proposed (Oh & Mack 2003; Barkana & Loeb 2005; Wolz et al. 2014; Alonso et al. 2015). All of these however leave some degree of residual am-

Parameter	CHIME	MeerKAT	SKA-Mid
$\nu_{\rm min}$	400 MHz	580 MHz	350 MHz
$\nu_{\rm max}$	800 MHz	1020 MHz	1050 MHz
$T_{\rm ant}$	50 K	29 K	28 K
$D_{\rm tel}$	100 m	800 m	1 km
$D_{\rm min}$	20 m	29 m	34 m
$A_{\rm coll}$	8000 m^2	9000 m^2	33000 m^2
$N_{\rm d}$	4	64	190
$D_{\rm dish}$	20 m	13.5 m	15 m
$S_{\rm area}$	25000 deg^2	25000 deg^2	25000 deg^2
$t_{\rm o}$	10^4 hours	10^4 hours	10^4 hours
$\Delta\nu$	10 MHz	10 MHz	10 MHz

Table 5.1: Experimental noise parameters for CHIME, MeerKAT and SKA-Mid. We include parameters for both interferometry and single-dish noise models.

plitude on the signal. These residuals are often modelled as a power law (Bull et al. 2015), and we will adopt this model here. Our foreground residuals are modelled as

$$C_\ell^{\rm FG}(\nu) = \epsilon^2 \sum_X A_X \left(\frac{\ell_{\rm f}}{\ell}\right)^{n_X} \left(\frac{\nu_{\rm f}}{\nu}\right)^{m_X}, \tag{5.47}$$

where the sum is taken over all contributing sources X. The power law coefficients n_X and m_X, as well as the amplitudes associated with each foreground are listed in table 5.2. Similar to Bull et al. (2015), we multiply our foreground model with a removal efficiency coefficient ϵ. Then, $\epsilon = 1$, if no foreground removal has been applied. In their analysis they found that foregrounds would have to be removed to a level of $\epsilon \lesssim 10^{-5}$ to access cosmological information. Foregrounds at redshift $z \gtrsim 1$ can be large and complex (Smoot & Debono 2017) and although subtraction to this level should be straight forward theoretically, as one just needs to model the smooth frequency component and subtract it from the observation, the large dynamic range between foregrounds and signal makes this process difficult (Shaw et al. 2014). Additional complications due to the shape of the beam varying as a function of frequency can result in mode-mixing and make this simple model of smooth frequency-dependent foregrounds inadequate in practice Liu et al. (2009). Foreground removal is a key challenge for observations of this kind and strides towards a more robust understanding are being made. Here, we focus on the capabilities of a noise-limited detection of the bispectrum and we therefore adopt an optimistic value of $\epsilon = 10^{-6}$. When this value is raised to higher than $\sim 10^{-5}$, we observe a large deterioration of the constraints in agreement with Bull et al. (2015). In such a case, our forecasts deteriorate to the 10% level at $\epsilon = 10^{-4}$, where the foregrounds swamp out most of the

Foreground source	A_X[mK]^2	n_X	m_X
Extragalactic point sources	57	1.1	2.07
Extragalactic free-free	0.014	1.0	2.1
Galactic synchrotron	700	2.4	2.8
Galactic free-free	0.088	3	2.15

Table 5.2: Foreground model parameters (Santos et al. 2005).

cosmological information. We would like to stress however that for a real observation, the level to which the foregrounds can be controlled, will largely determine the success of the detection.

Fig. 5.2 thus shows an example comparison of the foreground residuals removed with $\epsilon = 10^{-6}$ to the noise models of MeerKAT in single-dish and interferometer mode and the power spectrum model at $z = 1$.

5.6 The Fisher matrix

The Fisher information matrix (Fisher 1935; Tegmark et al. 1997; Hobson et al. 2010) is a powerful tool which allows us to estimate the minimum error one can expect from an upcoming experiment by assuming that the likelihood assumes a multivariate Gaussian form in the model parameters. By Taylor expanding the log-likelihood around its maximum-likelihood value, one can define the Fisher matrix as

$$\boldsymbol{F}_{ij} \equiv \left\langle \frac{\partial^2 \mathcal{L}}{\partial\theta_i \partial\theta_j} \right\rangle, \tag{5.48}$$

where $\mathcal{L} \equiv -\ln L$, the negative log-likelihood. The Cramer-Rao inequality (e.g. Heavens 2009; Hobson et al. 2010) then gives a lower bound for the errors one is expected to attain. When marginalizing over all other parameters in the analysis, the expected error on parameter i is given by

$$\sigma_i \geq \sqrt{(\boldsymbol{F}^{-1})_{ii}}, \tag{5.49}$$

which reduces the problem of predicting the minimum errors for an experiment to computing the Fisher matrix and inverting it. Tegmark et al. (1997) report the Fisher matrix for Gaussian data as

$$\boldsymbol{F}_{ij} = \frac{1}{2}\mathrm{Tr}(\boldsymbol{A}_i \boldsymbol{A}_j + \boldsymbol{C}^{-1}\boldsymbol{M}_{ij}), \tag{5.50}$$

where $\boldsymbol{C}$ denotes the covariance matrix, $\boldsymbol{A}_i \equiv \boldsymbol{C}^{-1}\boldsymbol{C}_{,i}$, $\boldsymbol{M}_{ij} \equiv \boldsymbol{\mu}_{,i}\boldsymbol{\mu}_{,j}^T + \boldsymbol{\mu}_{,j}\boldsymbol{\mu}_{,i}^T$, and $\boldsymbol{\mu} \equiv \langle \boldsymbol{x} \rangle$, where $\boldsymbol{x}$ denotes the data vector. We use the standard comma notation to signify derivatives with respect to the parameter, $\boldsymbol{C}_{,i} \equiv \partial \boldsymbol{C}/\partial\theta_i$.

For power spectrum forecasts, the data vector is taken to be the angular coefficient observed at some frequency ν,

$$\boldsymbol{x}^{\nu}_{\ell m} = a^{\nu}_{\ell m}. \tag{5.51}$$

As $\mu = \left\langle a^{\nu}_{\ell m} \right\rangle = 0$, the second term in the trace vanishes and

$$\begin{aligned} \boldsymbol{F}_{ij} &= \frac{1}{2}\mathrm{Tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_{,i}\boldsymbol{C}^{-1}\boldsymbol{C}_{,j}) \\ &= f_{\mathrm{sky}} \sum_{\nu} \sum_{\ell} (2\ell+1) \frac{C^{\nu}_{\ell,i} C^{\nu}_{\ell,j}}{\left(C^{\nu,\mathrm{tot}}_{\ell}\right)^2}, \end{aligned} \tag{5.52}$$

where we have summed over all m indices, and introduced a sky covering fraction $f_{\mathrm{sky}} = 0.5$ which effectively decreases the information gain by half and accounts for the correlation of nearby modes by the sky mask. We assume that the signal, noise and foreground residuals are all uncorrelated to each other, thus we find that

$$C^{\nu,\mathrm{tot}}_{\ell} = C^{\nu,\mathrm{S}}_{\ell} + C^{\nu,\mathrm{N}}_{\ell} + C^{\nu,\mathrm{FG}}_{\ell}. \tag{5.53}$$

For the bispectrum analysis, to ensure that the data vector for this analysis is Gaussian distributed, we use a weighted average of the angular bispectrum as the data (see Appendix C.7)

$$\boldsymbol{x}^{\nu}_{\ell_1\ell_2\ell_3} = \sum_{m_1 m_2 m_3} a^{\nu}_{\ell_1 m_1} a^{\nu}_{\ell_2 m_2} a^{\nu}_{\ell_3 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \tag{5.54}$$

where the matrix represents the Wigner 3-J symbol. We then assume that the Fisher matrix is dominated by the dependence of the mean bispectrum and thus the second term of equation 5.50 as it is when the power spectrum is taken to be the data vector and a large number of modes are observed. For computational ease, we assume that our bispectrum is uncorrelated between different frequency bins, such that we observe the bispectrum from a single frequency bin centred at ν only. The total Fisher matrix is thus the sum of the contributions from all frequency bins and all contributing modes $\lambda \equiv (\ell_1, \ell_2, \ell_3)$, which obey the triangle conditions,

$$\boldsymbol{F}_{ij} = \sum_{\nu} \sum_{\lambda} \boldsymbol{F}^{\nu,\lambda}_{ij}. \tag{5.55}$$

Then, applying Wick's theorem to evaluate the covariance matrix (ie. equation (C.77)),

$$\boldsymbol{C} = \left\langle (\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^t \right\rangle, \tag{5.56}$$

after applying the sum (C.80) in computing $\langle x \rangle_{,i}$, we finally find

$$\boldsymbol{F}_{ij} = \sum_{\nu} \sum_{\ell_1 \ell_2 \ell_3} \frac{\mu_{,i}^{\nu,\ell_1\ell_2\ell_3} \mu_{,j}^{\nu,\ell_1\ell_2\ell_3}}{\Delta_{\ell_1\ell_2\ell_3} C_{\ell_1}^{\nu} C_{\ell_2}^{\nu} C_{\ell_3}^{\nu}}, \tag{5.57}$$

where $\Delta_{\ell_1\ell_2\ell_3}$ is 6, 2, or 1 when all ℓ's, two ℓ's or no ℓ's are the same respectively, and

$$\mu_{,i}^{\nu,\ell_1\ell_2\ell_3} = \frac{\partial B_{\ell_1\ell_2\ell_3}^{\mathrm{NLG}}(\nu)}{\partial \theta_i} + \frac{\partial B_{\ell_1\ell_2\ell_3}^{\mathrm{LISW}}(\nu)}{\partial \theta_i}. \tag{5.58}$$

This expression is similar to those used in Komatsu & Spergel (2001); Komatsu et al. (2005); Mangilli et al. (2013); Hill (2018) when studying a Fisher forecast model for the bispectrum.

We use equations (5.52) and (5.57) to compute the Fisher matrix for power spectrum and bispectrum observations respectively. The information from both modes of analysis can be combined simply by adding the Fisher matrices if we assume both statistical measures to be uncorrelated. The cross correlations of the bispectrum and power spectrum result in a 5-point function which, to first order in perturbation theory, vanishes. A full second order analysis of the five-point function is therefore necessary to correctly combine the two statistical measures. Here we follow the approach of Takada & Jain (2004), assume that both statistical measures are uncorrelated or the correlations are small, and defer the computation of the 5 point function to a future analysis.

5.7 LISW detection signal to noise

Although we have seen in 5.4.4 that the LISW bispectrum signal can be significantly lower than that of the NLG bispectrum, considering shape differences between the contribution may allow a significant signal increase. In order to assess whether the LISW bispectrum signal is detectable by future experiments, we assume that, for $\alpha \equiv (\ell_1, \ell_2, \ell_3, m_1, m_2, m_3)$ obeying the triangle conditions, we observe a bispectrum $B_{\alpha}^{\mathrm{obs}}$. Suppose the shape of $B_{\alpha}^{\mathrm{LISW}}$ is fixed and it can be distinguished from other contributions to the bispectrum. Then, we create and minimize

$$\chi^2 = \sum_{\alpha} \frac{\left(B_{\alpha}^{\mathrm{obs}} - B_{\alpha}^{\mathrm{th}}\right)^2}{\sigma_{\alpha}^2}, \tag{5.59}$$

where

$$B_{\alpha}^{\mathrm{th}} \equiv \mathcal{A} B_{\alpha}^{\mathrm{LISW}} + \mathcal{B} B_{\alpha}^{\mathrm{NLG}}, \tag{5.60}$$

for some amplitudes $\mathcal{A}$ and $\mathcal{B}$. For simplicity, it is assumed that amplitude for non-linear gravity is known exactly, so it can be subtracted from the observations, such that

$$\chi^2 = \sum_\alpha \frac{\left(\tilde{B}_\alpha^{\rm obs} - \mathcal{A} B_\alpha^{\rm LISW}\right)^2}{\sigma_\alpha^2}, \tag{5.61}$$

where $\tilde{B}_\alpha^{\rm obs} \equiv B_\alpha^{\rm obs} - \mathcal{B} B_\alpha^{\rm NLG}$. Minimising this function with respect to the LISW amplitude $\mathcal{A}$, we obtain an estimator

$$\hat{\mathcal{A}} = \frac{\sum_\alpha \tilde{B}_\alpha^{\rm obs} B_\alpha^{\rm LISW}/\sigma_\alpha^2}{\sum_\alpha \left(B_\alpha^{\rm LISW}\right)^2/\sigma_\alpha^2}, \tag{5.62}$$

with a variance on the estimator given by

$$\sigma^2_{\hat{\mathcal{A}}} = \frac{1}{\sum_\alpha \left(B_\alpha^{\rm LISW}\right)^2/\sigma_\alpha^2}. \tag{5.63}$$

The variance on the bispectrum is computed in Spergel & Goldberg (1999) as

$$\sigma_\alpha^2 = \left\langle B_\alpha^2 \right\rangle - \langle B_\alpha \rangle^2 \simeq \Delta_\alpha C_{\ell_1}^{\nu,{\rm tot}} C_{\ell_2}^{\nu,{\rm tot}} C_{\ell_3}^{\nu,{\rm tot}}, \tag{5.64}$$

where Δ_α is 6, 2, or 1 when all ℓ's, two ℓ's or no ℓ's are the same respectively. The C_ℓ here denote the angular 21cm power spectrum including detector noise.

Assuming now that a fiducial value for our estimator is $\mathcal{A} = 1$, and that our estimator is unbiased, $\langle \hat{\mathcal{A}} \rangle = \mathcal{A}$, we compute the signal to noise ratio for an IM experiment probing the LISW bispectrum,

$$\frac{S}{N} = \sqrt{\sum_{\text{all } \alpha} \frac{(B_\alpha^{\rm LISW})^2}{\sigma_\alpha^2}}. \tag{5.65}$$

Importantly, we can sum out all m indices by applying (C.80), such that

$$\frac{S}{N} = \sqrt{\sum_{\ell_1\ell_2\ell_3} \frac{(B_{\ell_1\ell_2\ell_3}^{\rm LISW})^2}{\sigma^2_{\ell_1\ell_2\ell_3}}}. \tag{5.66}$$

Fig. 5.9 shows the detection signal to noise ratio as a function of the largest ℓ mode included in the sum in the optimal case for which the shape of other contributing bispectra is known exactly. Despite the large number of modes added, a direct detection of the LISW contribution to the bispectrum is impossible as even including information from small-scales does not increase the signal to noise ratio significantly above 10^{-3}. In comparison to the CMB, where the LISW contribution represents a major contaminant for primordial non-Gaussianity observations (Kim et al. 2013; Planck Collaboration et al. 2016a), the

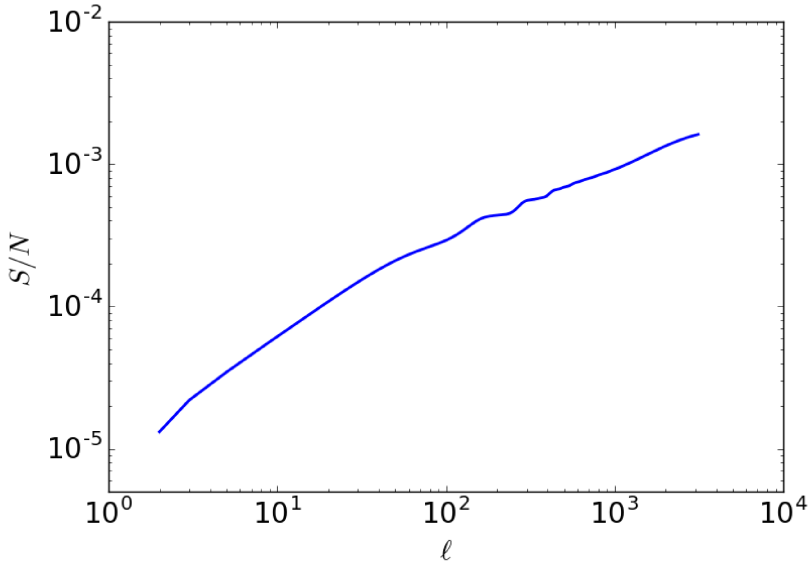

Figure 5.9: Signal to noise for a LISW bispectrum signal detection vs ℓ, the highest multipole moment observed, using MeerKAT in interferometry mode.

principal reason for the small S/N ratio found here, is that the power spectrum contribution to the noise term (see equation (5.64)) is significantly larger due to the late stage of the gravitational growth of structure.

Although a direct detection of the LISW bispectrum is impossible, ignoring it potentially biases the measurement of cosmological parameters from bispectrum observations (Planck Collaboration et al. 2016a). We compute this bias term here (Kim et al. 2004; Taylor et al. 2007), but find values $\sim 8-10$ orders of magnitude lower than the expected errors, and thus the LISW effect insignificantly affects the 21cm bispectrum.

5.8 Fisher predictions

Computing equations (5.52) and especially (5.57) is expensive due to the large number of modes, frequency bins and parameter combinations necessary to compute. For the bispectrum, the number of modes scales as $\sim \ell_{\rm max}^3$, meaning that bispectra containing small angular scales require an extraordinarily large number of modes to be computed. The limiting scale at a given frequency is set by maximal baseline in the interferometer or the dish size of the single-dish observation. In order for the bispectrum computation to be practical, we ignore any modes dominated by noise ($\ell_{\rm max} \gtrsim 1600$), and thus impose an upper bound of $\ell_{\rm max} = \min(2\pi D_{\rm tel/dish}/\lambda, 1600)$. Additionally, the minimal baseline for an interferometer, $D_{\rm min}$, sets the largest observable mode and thus we impose an

		Interferometer			**Single-Dish**		**Combined**	
		CHIME	MeerKAT	SKA	MeerKAT	SKA	MeerKAT	SKA
Parameter	Fid. Value	*Marginalized error for bispectrum analysis*						
$\Omega_{\rm CDM}h^2$	0.127	3.1×10^{-4}	3.7×10^{-4}	2.3×10^{-4}	1.1×10^{-3}	1.0×10^{-3}	2.4×10^{-4}	1.7×10^{-4}
$\Omega_{\rm b}h^2$	0.022	8.2×10^{-5}	1.0×10^{-4}	6.0×10^{-5}	3.0×10^{-4}	2.9×10^{-4}	8.4×10^{-5}	5.4×10^{-5}
Ω_Λ	0.684	2.1×10^{-4}	2.4×10^{-4}	1.7×10^{-4}	1.7×10^{-3}	1.6×10^{-3}	2.3×10^{-4}	1.7×10^{-4}
$n_{\rm s}$	0.962	5.7×10^{-4}	5.7×10^{-4}	3.9×10^{-4}	2.5×10^{-3}	2.3×10^{-3}	3.8×10^{-4}	3.0×10^{-4}
$A_{\rm s}\times10^9$	1.562	5.9×10^{-3}	7.2×10^{-3}	4.5×10^{-3}	2.1×10^{-2}	1.9×10^{-2}	4.2×10^{-3}	3.2×10^{-3}
H_0	67	8.0×10^{-2}	9.7×10^{-2}	5.6×10^{-2}	3.0×10^{-1}	2.8×10^{-1}	7.0×10^{-2}	4.6×10^{-2}
Parameter	Fid. Value	*Marginalized error for power spectrum analysis*						
$\Omega_{\rm CDM}h^2$	0.127	9.3×10^{-4}	4.9×10^{-4}	3.8×10^{-4}	2.5×10^{-3}	1.4×10^{-3}	4.6×10^{-4}	3.4×10^{-4}
$\Omega_{\rm b}h^2$	0.022	4.2×10^{-4}	2.9×10^{-4}	2.4×10^{-4}	1.0×10^{-3}	6.5×10^{-4}	2.5×10^{-4}	2.0×10^{-4}
Ω_Λ	0.684	2.9×10^{-3}	1.6×10^{-3}	1.2×10^{-3}	1.2×10^{-2}	6.4×10^{-3}	1.5×10^{-3}	1.1×10^{-3}
$n_{\rm s}$	0.962	1.5×10^{-3}	9.0×10^{-4}	7.2×10^{-4}	8.0×10^{-3}	4.5×10^{-3}	8.4×10^{-4}	6.5×10^{-4}
$A_{\rm s}\times10^9$	1.562	8.4×10^{-3}	8.4×10^{-3}	6.3×10^{-3}	4.9×10^{-2}	2.5×10^{-2}	7.6×10^{-3}	5.4×10^{-3}
H_0	67	3.0×10^{-1}	2.3×10^{-1}	1.9×10^{-1}	8.9×10^{-1}	5.1×10^{-1}	2.1×10^{-1}	1.6×10^{-1}
Parameter	Fid. Value	*Marginalized error for combined power spectrum + bispectrum analysis*						
$\Omega_{\rm CDM}h^2$	0.127	9.2×10^{-5}	1.2×10^{-4}	5.5×10^{-5}	6.5×10^{-4}	4.9×10^{-4}	1.1×10^{-4}	5.3×10^{-5}
$\Omega_{\rm b}h^2$	0.022	4.0×10^{-5}	5.8×10^{-5}	3.1×10^{-5}	2.1×10^{-4}	1.7×10^{-4}	5.3×10^{-5}	2.9×10^{-5}
Ω_Λ	0.684	1.7×10^{-4}	2.2×10^{-4}	1.4×10^{-4}	1.7×10^{-3}	1.4×10^{-3}	2.2×10^{-4}	1.4×10^{-4}
$n_{\rm s}$	0.962	1.4×10^{-4}	1.1×10^{-4}	8.2×10^{-5}	1.2×10^{-3}	8.1×10^{-4}	1.0×10^{-4}	7.3×10^{-5}
$A_{\rm s}\times10^9$	1.562	1.7×10^{-3}	2.2×10^{-3}	1.2×10^{-3}	1.3×10^{-2}	9.8×10^{-3}	1.9×10^{-3}	1.1×10^{-3}
H_0	67	3.2×10^{-2}	4.5×10^{-2}	2.2×10^{-2}	1.9×10^{-1}	1.6×10^{-1}	4.1×10^{-2}	2.1×10^{-2}

Table 5.3: Marginal errors recovered from the Fisher forecasts. We show the results for 10^4 hours of integration time for bispectrum-only observations, power spectrum-only observations and the combined analysis. Each of these analyses is performed in interferometry and single-dish mode, and we finally combine the analysis to obtain the information gained over all scales.

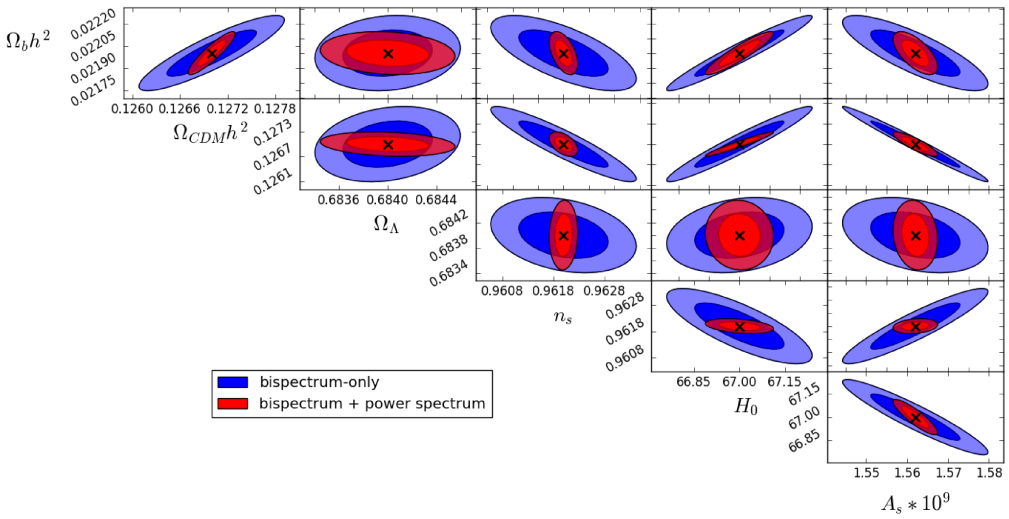

Figure 5.10: Fisher forecasts for bispectrum-only (blue ellipses) and power spectrum + bispectrum (red ellipses) observations for MeerKAT in interferometry mode. We show the 68% and 95% credibility intervals for the cosmological parameters used in our analysis for a total integration time of 10^4 hours. The cross shows the fiducial value of the parameters, and uniform priors are assumed.

$\ell_{\rm min} = 2\pi D_{\rm min}/\lambda$ for observations performed in interferometry mode. As a consequence, interferometric studies of the 21cm bispectrum will not be sensitive to squeezed triangle configurations, and thus will not contain information from triangles which maximize its amplitude. Further, the Fisher information varies smoothly as a function of $\ell_{\rm max}$ and we therefore compute the statistical quantities with a step-size of $\Delta\ell_{\rm max} = 80$ and linearly interpolate when summing over ℓ in equation (5.57). In order to compute the Fisher matrix, we pre-compute bispectrum derivatives relative to each parameter varied for each ℓ-triplet in the analysis, and subsequently read the results from file when computing different Fisher matrix elements. This process is done once before running the Fisher matrix analysis and takes roughly 3 days using a 40 core workstation. The long runtime is due to the slow evaluation of the NLG bispectrum. Whereas producing the data for figure 5.5 requires a few minutes, the data for figure 5.4 is produced with a runtime of $\sim$ 1 hour. Once derivatives are pre-computed, a full analysis run can be done in $\sim$ 3 hours on 40 cores.

The results for the Fisher analysis are shown in table 5.3. The table is subdivided into three sections, comparing the bispectrum-only observations for each of our fiducial experiments in the first seven rows. The next seven rows show the results for our power spectrum-only observations, before we combine both results in the final section of the table. Furthermore, we compare results from observations made in interferometry and

single-dish mode, and combine their Fisher matrices in the final column as they are uncorrelated.

We find that bispectrum observations have the potential to improve the parameter constraints from power spectrum observations significantly due to the large number of accessible modes, $N_{\rm modes} \sim \ell^3_{\rm max}$, but the bispectrum does not contain all the information and a combination of both statistics is required to obtain the best constraints. All experiments exhibit errors from the bispectrum forecasts which are a factor of $\sim 1.1 - 7$ better than compared to the power spectrum. Combinations of both show an order of magnitude reduction in our error forecasts for most parameters. We find the strongest constraints across our analysis for Ω_Λ, $n_{\rm s}$ and H_0. This is in line with the expectation that IM experiments should improve the constraints of H_0 and Ω_Λ the most, as well as parameters such as $n_{\rm s}$ which are correlated to these (Bull et al. 2015). Comparing our results for MeerKAT in interferometry mode to Planck Collaboration et al. (2016b), the 21cm power spectrum forecasts show a factor of $\sim 2 - 5$ decrease in marginalized errors for most parameters, with the exception of $\Omega_{\rm b}h^2$ where we do not achieve the same level of sensitivity. The bispectrum promises to tighten constraints on all cosmological parameters by up to a factor of 10, thus having the potential to bridge the gap between current low-redshift and CMB observations of the cosmological parameters. Even in single-dish mode, we find that our power spectrum forecasts result in similar errors as those observed with *Planck*, and the bispectrum again improving these findings typically by a factor of 3. The best possible constraints are achieved by combining interferometric and single-dish observations of both the power spectrum and the bispectrum, these combinations marginally improve the constraints obtained from interferometric power spectrum and bispectrum combinations.

Of note is that we find that CHIME achieves better constraints from bispectrum observations than MeerKAT, even though power spectrum observations find error constraints of a factor of ~ 2 worse than MeerKAT. We find the CHIME noise power spectrum to be an order of magnitude lower than that of MeerKAT and would thus naively expect the CHIME power spectrum observations to result in stronger constraints. This is not the case as the 21cm power spectrum peaks on scales $\ell \sim 800$ which are on the edge of resolvability for CHIME. Thus despite higher instrumental noise, compared to CHIME, MeerKAT is able to resolve smaller scales, due to its larger baselines, and thus is sensitive to the peak in the signal power spectrum. For bispectrum observations, neither telescope is sensitive to the largest amplitude triangles and despite being sensitive to a larger number of modes, they are noisier for MeerKAT observations, such that CHIME is able to use the bispectrum to a higher potential.

Fig. 5.10 shows the 1σ and 2σ error ellipses from our analysis for both bispectrum-only and power spectrum plus bispectrum combined observations by MeerKAT in inter-

ferometry mode. The combination of the information gain from both bispectrum and power spectrum is thus not only useful to decrease errors, but can be a helpful tool to break degeneracies between parameters.

5.9 Summary

The most precise observations of the CMB to date (Planck Collaboration et al. 2016b) have confirmed the simple picture of a six parameter cosmological model with a cosmological constant and a flat curvature. Although there is not sufficient evidence to strongly favour any other model at this time (Heavens et al. 2017), tensions between CMB and low-redshift observations of weak lensing and local measurements of the Hubble rate still persist. New low-redshift probes may help to rectify these short-comings of the model and give new insights into the cosmological evolution since the time of recombination. The cosmological 21cm signal is an ideal probe as HI and thus the 21cm signal is present at all epochs after the CMB is released. 21cm intensity mapping experiments will soon supplement galaxy surveys for mapping the large scale structure of the universe by observing the diffuse 21cm emission from hydrogen gas inside low-redshift galaxies. These experiments will probe unprecedented cosmological volumes and provide precise redshift information for their observations, due to the direct relation between the observed frequency of the signal and the redshift of the source.

We have studied the 21cm bispectrum and power spectrum in the context of IM observations by CHIME, MeerKAT, and SKA-mid and derived the expression for the 21cm bispectrum due to the non-linear collapse of structure post reionization. For the first time, we derived the expected contribution to the 21cm bispectrum from the lensing-ISW bispectrum which is due to the evolution of the density field along the line of sight. In contrast to CMB observations, we find, as expected, that the lensing-ISW bispectrum as only introduces a negligible bias to the parameter constraints and we predict a cumulative signal to noise ratio of 10^{-3}, making a detection impossible. We introduce a new way of visualizing the bispectrum which allows for a direct relation between the triangle shape and the resulting amplitude. Finally, we analysed the predictive capabilities of these bispectrum contributions in the context of a Fisher forecast model and found that the bispectrum from IM experiments has the potential to greatly improve cosmological parameter contrains. Although not sensitive to the largest amplitude triangles, the large number of observable modes should allow interferometric IM experiments to extract enough information to decrease parameter errors by an order of magnitude compared to the *Planck* measurements. For the best case scenario, the combined analysis of interferometry and single-dish observations of both power spectrum and bispectrum with

SKA-mid, an impressive level of precision can be achieved. We find a relative marginalized error of $< 0.1\%$ for all cosmological parameters, except for $\Omega_{\rm b}h^2$ for which we find a relative error of $\sim 0.13\%$. The bispectrum is especially sensitive to $n_{\rm s}$ where we find a relative marginalized error of $< 0.01\%$. It is important to reiterate that these results are heavily subject to the level of foreground removal that can be achieved, and we have used an optimistic foreground removal efficiency of $\epsilon = 10^{-6}$ to explore the full, noise limited potential of bispectrum observations.

Chapter 6

3D Power Spectrum Analysis

In this short chapter, we introduce the 3D analysis of the 21cm signal in the spherical harmonic spherical Bessel function basis as an approach to deal with wide angle 21cm survey data. We first express the 21cm power spectrum in this basis. Then we determine the perturbation of this result by the presence of redshift space distortions from peculiar velocities in the hydrogen gas emitting the signal. Finally, we introduce the window function of a real 21cm survey into our calculation as well as the effects of an angular mask.

6.1 Introduction

With the advent of 21cm observatories able to survey large sections of the sky, many of the analysis techniques based on flat-sky assumptions will be inadequate to analyse the incoming data. It is therefore essential to extend these techniques to deal with large area observations. Additionally, one of the strengths of the 21cm observations highlighted in the previous chapters is the direct relation between observed frequency and redshift of the source. To take full advantage of this property, we need to introduce an alternative to the conventional binning of the signal and analysing each bin separately.

Similar challenges were overcome in the analysis of weak lensing surveys at the start of the millennium. For lensing studies, precise redshift information of the source galaxies is expensive to obtain and for large volume surveys is therefore obtained photometrically. Although errors on the photometric redshifts can be appreciable, their inclusion in the analysis can remove systematic errors which makes obtaining the redshifts worth the additional integration time. Presented with the redshift information of the sources, the statistical analysis ought to take advantage of this additional information. Heavens (2003) has demonstrated the proper way to deal with large angle surveys that contain redshift information of the sources in the context of weak lensing studies. Here, the idea of a natural

3D basis for the source population on the spherical sky was exploited to decrease statistical errors in the analysis. The natural basis considered is that of a spherical harmonic spherical Bessel function basis, as these represent the eigenfunctions of the Laplace operator in spherical coordinates. Whereas the source population for weak lensing studies consists of individually resolved galaxies, leading to a discrete formulation of the basis transforms, in the context of 21cm intensity mapping surveys, the signal can be considered continuous as a result of the unresolved mapping of 21cm brightness temperatures across the sky and the presence of precise radial information of the source plane. For a continuous field $f(r, \theta, \phi)$ expressed in spherical polar coordinates, we define the spherical harmonic spherical Bessel function transform as

$$f_{\ell m}(k) = \sqrt{\frac{2}{\pi}} \int d^3\boldsymbol{r} f(r, \theta, \phi) j_\ell(kr) Y^*_{\ell m}(\theta, \phi), \tag{6.1}$$

where $Y^*_{\ell m}$ denotes the complex conjugate of the spherical harmonics at angular multipoles ℓ and m, and j_ℓ represents the spherical Bessel function. Applying the orthonormality relations for spherical harmonics, an inverse transform defining the field f in terms of its coefficients in the 3D basis can be derived as

$$f(r, \theta, \phi) = \sqrt{\frac{2}{\pi}} \int d\kappa \kappa^2 \sum_{\ell m} f_{\ell m}(\kappa) j_\ell(\kappa r) Y_{\ell m}(\theta, \phi). \tag{6.2}$$

The separation of modes according to their wavelengths and the freedom to analyse linear modes only is another advantage of the radial decomposition through the spherical Bessel function.

In this chapter we thus apply the spherical harmonic spherical Bessel function basis to the 21cm brightness temperature field and compute the 3D 21cm power spectrum in section 6.2. Two modifications to the resulting power spectrum are then introduced. In section 6.3 we find the 3D 21cm power spectrum when peculiar velocities of the sources lead to redshift space distortions. Then, in section 6.4 we discuss two effects introduced by the limitations of observations, and thus a radial window function as well as an angular mask are included into our calculation here. Finally we summarize the chapter and allude to possible extensions of the calculation in section 6.5.

6.2 21cm cosmology in three dimensions

The 21cm signal, similar to weak lensing, is an inherently three dimensional signal and information is bound to be lost due to discontinuous sampling of the signal along the line of sight. At late times the signal is directly related to the underlying dark matter

density field, as neutral hydrogen gas in the post EoR Universe is primarily located in self-shielded gas clouds inside galaxies. We can thus write the brightness temperature as a biased tracer of the density field, similar to chapter 5,

$$\delta T_{\rm b}[r(z)\hat{\boldsymbol{n}}, z] = \delta\bar{T}_{\rm b}(z)\left\{1 + b_{\rm HI}(z)\delta[r(z)\hat{\boldsymbol{n}}, z]\right\}, \tag{6.3}$$

where we have split the signal into a homogeneous and fluctuating part. To first order in perturbation theory, the density fluctuations simply grow as a function of the growth factor,

$$\delta[r(z)\hat{\boldsymbol{n}}, z] = D_+(z)\delta[\boldsymbol{r}(z)]. \tag{6.4}$$

In order to avoid the information loss from the discrete binning of the observations, we transform the observed brightness temperature fluctuations into the spherical harmonic spherical Bessel basis, using equation (6.1),

$$a_{\ell m}(k) = \sqrt{\frac{2}{\pi}}\int d^3\boldsymbol{r}_{\rm f}\, \delta T_{\rm b}[r_{\rm t}(z)\,\hat{\boldsymbol{n}}, z]\; j_\ell[kr_{\rm f}(z)]\; Y^*_{\ell m}(\hat{\boldsymbol{n}}). \tag{6.5}$$

In the above expression we have noted explicitly that the transformation of the signal is performed according to some fiducial cosmology, as indicated by the subscript 'f'. Similarly, the signal will be observed in the context of some true cosmology, or model cosmology when the signal is simulated, and thus we note this by the index 't'. In the following, we will omit the explicit expression of fiducial functions, such that $r_{\rm f} = r$. We can further see that this expression splits into two terms after applying equation (6.3), where the first only contributes at $\ell = m = 0$, due to the isotropic distribution of the signal on the sky. For angular multipoles ℓ and m different from 0, the signal transform can then be expressed as

$$a_{\ell m}(k) = \sqrt{\frac{2}{\pi}}\int dz\, d^2\hat{\boldsymbol{n}}\, \delta\bar{T}_{\rm b}(z)\, b_{\rm HI}(z)\, D_+(z)\, \delta[\boldsymbol{r}_{\rm t}(z)]\; j_\ell[kr(z)]\; Y^*_{\ell m}(\hat{\boldsymbol{n}})\, \frac{c}{H(z)}\, r^2(z). \tag{6.6}$$

Next, we use (6.2) to inverse transform the density field in order to relate the coefficients of the brightness temperature signal to the coefficients of the density field. This transformation is done with respect to the underlying true cosmology of the data, therefore

$$\delta[\boldsymbol{\rho}(z)] = \sqrt{\frac{2}{\pi}}\int d\kappa\kappa^2 \sum_{\ell m}\delta_{\ell m}(\kappa)\, j_\ell[\kappa\rho(z)]Y_{\ell m}(\hat{\boldsymbol{n}}), \tag{6.7}$$

where, for notational ease, we write $r_{\rm t} = \rho$. Only the model cosmology guarantees isotropy, which is a required feature of the matter power spectrum, which we will show this expression to be related to.

Then, combining expressions (6.6) and (6.7), the angular dependence is restricted to the spherical harmonics. For an all-sky survey the angular dependence can thus be integrated out using the fact that spherical harmonics are normalized and orthogonal,

$$\int d^2\hat{\boldsymbol{n}} Y_{\ell m}(\hat{\boldsymbol{n}}) Y^*_{\ell' m'}(\hat{\boldsymbol{n}}) = \delta^{\rm K}_{\ell\ell'}\delta^{\rm K}_{mm'}, \tag{6.8}$$

where $\delta^{\rm K}$ denotes the Kronecker delta function. Using this relation and performing the sum over all angular multipoles, we obtain

$$a_{\ell m}(k) = \int d\kappa \kappa^2 M_\ell(k,\kappa)\delta_{\ell m}(\kappa), \tag{6.9}$$

and we have defined

$$M_\ell(k,\kappa) \equiv \frac{2c}{\pi}\int dz \frac{r^2(z)}{H(z)}\delta\bar{T}_{\rm b}(z) b_{\rm HI}(z) D_+(z) j_\ell[kr(z)]\; j_\ell[\kappa\rho(z)]. \tag{6.10}$$

The three dimensional 21cm power spectrum is then obtained as the expectation value of two spherical harmonic spherical bessel coefficients,

$$\left\langle a_{\ell m}(k) a^*_{\ell' m'}(k')\right\rangle = \int d\kappa d\kappa' \kappa^2 \kappa'^2 M_\ell(k,\kappa) M_{\ell'}(k',\kappa') \left\langle \delta_{\ell m}(\kappa)\delta^*_{\ell' m'}(\kappa')\right\rangle. \tag{6.11}$$

Heavens (2003) defines the matter power spectrum in terms of the density field coefficients through

$$\left\langle \delta_{\ell m}(\kappa)\delta^*_{\ell' m'}(\kappa')\right\rangle = \frac{P(\kappa)}{\kappa^2}\delta^{\rm D}(\kappa-\kappa')\delta^{\rm K}_{\ell\ell'}\delta^{\rm K}_{mm'}, \tag{6.12}$$

where $\delta^{\rm D}$ is the Dirac delta function. We thus arrive at an expression which connects the three dimensional 21cm power spectrum explicitly to the matter power spectrum, as

$$C_\ell(k,k') \equiv \left\langle a_{\ell m}(k) a^*_{\ell m}(k')\right\rangle = \int d\kappa \kappa^2 P(\kappa) M_\ell(k,\kappa) M_\ell(k',\kappa). \tag{6.13}$$

Computing this power spectrum is numerically difficult due to the rapid oscillating nature of the spherical Bessel functions being integrated when computing $M_\ell(k,\kappa)$. The widely used Limber approximation (Loverde & Afshordi 2008; and appendix C.4) can be applied to reduce the number of integrals that need to be computed, and we arrive at the

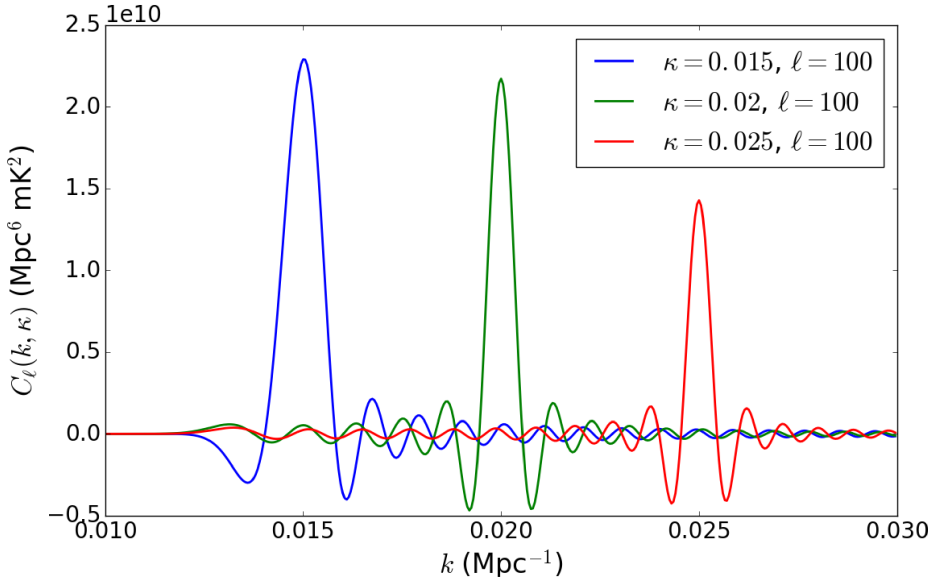

Figure 6.1: 3D 21cm power spectrum at angular mode $\ell = 100$ for $\kappa = 0.015\text{Mpc}^{-1}$ (blue), $\kappa = 0.02\text{Mpc}^{-1}$ (green), and $\kappa = 0.025\text{Mpc}^{-1}$ (red) as a function of radial mode k. Here we have applied the Limber approximation hence plotting equation (6.14).

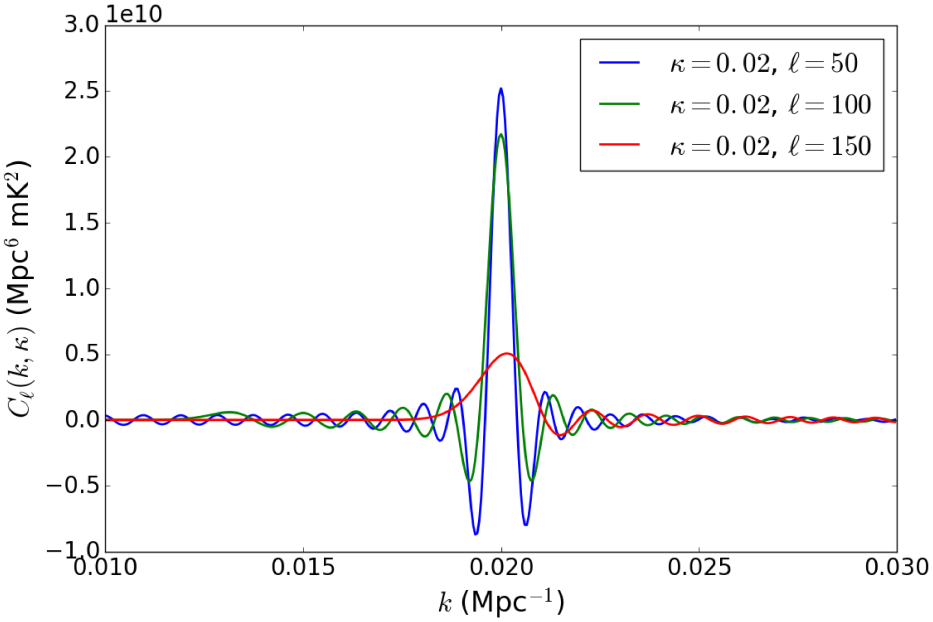

Figure 6.2: 3D 21cm power spectrum at radial mode $\kappa = 0.02\text{Mpc}^{-1}$ for $\ell = 50$ (blue), $\ell = 100$ (green), and $\ell = 150$ (red) as a function of radial mode k. Here we have applied the Limber approximation hence plotting equation (6.14).

simplified form

$$C_\ell^{\text{Limber}}(k,k') = \frac{2c^2}{\pi}\int dz \frac{r^2(z) b_{\text{HI}}^2(z)}{|r'(z)|H^2(z)} \delta\bar{T}_{\text{b}}^2(z) D_+^2(z) j_\ell[kr(z)] j_\ell[k'r(z)] P\left(\frac{\ell}{r(z)}, z\right). \quad (6.14)$$

Although being widely used, the Limber approximations fails in a number of situations (**??**Di Dio et al. 2018) and a more rigorous fast computation of the Bessel functions is desirable (see eg. Levin 1982; 1996; 1997; Assassi et al. 2017). Recently, Spurio Mancini et al. (2018) have compared implementations of fast numerical methods for the evaluation of multiple integrals involving Bessel functions as an alternative to applying the Limber approximation. In order for our formalism to be usable, a detailed study into the validity of the Limber approximation in this context and an adaptation of such fast codes would be necessary.

In Figs 6.1 and 6.2 we show the approximate computation of the 3D power spectrum, using the Limber approximation, for a number of angular and radial modes.

6.3 The effects of redshift-space distortions

In this section we study the peculiar velocities of the hydrogen gas in the IGM and their effect on the 3D power spectrum. Peculiar velocities give rise to distortions in redshift-space, so the observed redshift does not correspond to the true cosmological redshift of the gas observed. In a first step we will see what these distortions are for high redshift objects and then include the effect into our 3D 21cm power spectrum.

6.3.1 Peculiar velocities

When making cosmological observations, we are inherently limited to the the observed redshift as a means to infer the distance towards the object in question. In a completely homogeneous universe, the mapping from the redshift to the comoving distance would be directly given through the FRW metric, however inhomogeneities in the density field introduce peculiar velocities which affect this relation (Kaiser 1987). The observed redshift can be related to the cosmological redshift if the peculiar velocity is know. Consider a photon travelling on a null geodesic, emitted at some time t_{E} and detected later at t_{O}. A short time Δt_{E} later, the source, travelling with peculiar velocity $u \equiv dr/dt$, emits another photon which the observer sees Δt_{O} after the first photon was caught. Between emitting both pulses of light, the source has thus changed its relative comoving radial separation to the observer by $u\Delta t_{\text{E}}$. The expressions for the distance to the source at emission of both

photons is then

$$r = \int_{t_E}^{t_O} \frac{c}{a(t)} dt, \tag{6.15}$$

for the first, and

$$r + u\Delta t_E = \int_{t_E+\Delta t_E}^{t_O+\Delta t_O} \frac{c}{a(t)} dt \tag{6.16}$$

for the second. Using (6.15) in (6.16) and assuming that both time intervals are short, such that $a(t)$ remains constant between the emission of both light pulses, we find the observed redshift as a function of the peculiar velocity of the source as

$$1 + z_{obs} \equiv \frac{\Delta t_O}{\Delta t_E} = \frac{a(t_O)}{a(t_E)} + \frac{ua(t_O)}{c}. \tag{6.17}$$

Using the definition of the cosmological redshift from chapter 2, the observed redshift for a source with peculiar velocity u is thus

$$z_{obs} = z_{true} + \frac{u}{c}. \tag{6.18}$$

For cosmological observations, the distance to the source recovered from the observed redshift is thus not identical to the radial distance to the source. Let s be the distance derived from z_{obs}, whereas r is the true distance to a cloud of hydrogen gas, then

$$s = \int_0^{z_{obs}} \frac{cdz'}{H(z')}, \tag{6.19}$$

and

$$r = \int_0^{z_{true}} \frac{cdz'}{H(z')}, \tag{6.20}$$

which can be combined with equation (6.18) to find the relation between redshift-space and real space distances,

$$s = r + \frac{u}{H(z)}. \tag{6.21}$$

6.3.2 Redshift-space analysis

Whereas at the beginning of this chapter we have assumed that the redshift we observe can be directly related to a distance coordinate of the source, we will now expand on the idea of a peculiar velocity of the source and see how the 3D power spectrum changes as sources are no longer stationary. Observing hydrogen gas at a given redshift, the associated distance coordinate is distorted by the peculiar velocity of the gas. Therefore, we have to write the brigthness temperature fluctuation in terms of the observed redshift and

the distance measure related to it, s,

$$\delta T_{\mathrm{b}}^{(s)}(\boldsymbol{s}) = \left[1 + b_{\mathrm{HI}}(s)\delta^{(s)}(\boldsymbol{s})\right]\delta\bar{T}_{\mathrm{b}}(s). \tag{6.22}$$

Note that this relation is different from equation (6.3), which is expressed in terms of the true distance to the source gas, r,

$$\delta T_{\mathrm{b}}^{(r)}(\boldsymbol{r}) = \left[1 + b_{\mathrm{HI}}(r)\delta^{(r)}(\boldsymbol{r})\right]\delta\bar{T}_{\mathrm{b}}(r). \tag{6.23}$$

As before, we transform the redshift space signal

$$a_{\ell m}^{(s)}(k) = \sqrt{\frac{2}{\pi}}\int d^3\boldsymbol{s}\left[1 + b_{\mathrm{HI}}(s)\delta^{(s)}(\boldsymbol{s})\right]\delta\bar{T}_{\mathrm{b}}(s)j_\ell(ks)Y^*_{\ell m}(\hat{\boldsymbol{n}}). \tag{6.24}$$

An application of the continuity condition for the density of neutral hydrogen reveals that we may transform from redshift space to real space using $d^3\boldsymbol{s}\,\rho_{\mathrm{HI}}(s) = d^3\boldsymbol{r}\,\rho_{\mathrm{HI}}(r)$ and thus obtain, for ℓ and m different from zero,

$$a_{\ell m}^{(s)}(k) = \sqrt{\frac{2}{\pi}}\int d^3\boldsymbol{r}\left[1 + b_{\mathrm{HI}}(r)\delta^{(r)}(\boldsymbol{r})\right]\delta\bar{T}_{\mathrm{b}}(s)j_\ell(ks)Y^*_{\ell m}(\hat{\boldsymbol{n}}), \tag{6.25}$$

where we have conserved the hydrogen distribution across the change of frame of reference. We will now proceed by showing that the redshift space quantities are in fact first order quantities, which is why we have explicitly noted the zero order term in the above expression. The redshift space distortions are thus restricted to the argument of the brightness temperature signal and the spherical Bessel function. We have seen in section 6.3.1 that the radial coordinate in redshift space is related to the real distance to the source r via a small perturbation, and thus we can Taylor-expand the redshift space quantities,

$$\delta\bar{T}_{\mathrm{b}}(s)j_\ell(ks) \approx \delta\bar{T}_{\mathrm{b}}(r)j_\ell(kr) + \frac{u(\boldsymbol{r})}{H(r)}\frac{d}{dr}\left[\delta\bar{T}_{\mathrm{b}}(r)j_\ell(kr)\right]. \tag{6.26}$$

The peculiar velocity u of a cloud of gas can be related to the underlying density field as the baryonic matter is coupled to the bulk flow of the dark matter. In appendix D.1 we find the relation to be

$$u(\boldsymbol{r}) = aH(a)f(\Omega_{\mathrm{m}})\sqrt{\frac{2}{\pi}}\int dk\sum_{\ell m}\delta_{\ell m}^{(r)}(\kappa)Y_{\ell m}(\hat{\boldsymbol{n}})\frac{d}{d\rho}j_\ell(\kappa\rho), \tag{6.27}$$

where

$$f\left(\Omega_{\mathrm{m}}\right) \equiv \frac{a}{\delta}\frac{d\delta}{da} \simeq \Omega_{\mathrm{m}}^{0.6}. \tag{6.28}$$

We can thus see that the second term in the Taylor expansion (6.26) is $O(\delta)$. Therefore, the 3D coefficient retains two terms contributing at first order. As before, our assumption of an isotropic signal guarantees that the 0th order term only contributes for angular moments $\ell = m = 0$, such that the coefficient is $O(\delta)$ as long as we stay away from the monopole,

$$\begin{aligned} a_{\ell m}^{(s)}(k) = &\frac{2}{\pi}\int dr d\kappa \frac{r^2 f(\Omega_{\rm m})}{1+z(r)}\frac{d}{d\rho}j_\ell(\kappa\rho)\frac{d}{dr}\left[\delta\bar{T}_{\rm b}(r)j_\ell(kr)\right]\delta_{\ell m}(\kappa) \\ &+\frac{2}{\pi}\int dr d\kappa r^2\kappa^2\delta\bar{T}_{\rm b}(r)j_\ell(\kappa\rho)j_\ell(kr)\delta_{\ell m}(\kappa). \end{aligned} \tag{6.29}$$

The second term in this expression can be easily identified with equation (6.9). In a similar fashion, we invert the order of integration in the first term and define

$$\mathcal{M}_\ell(k,\kappa) = \frac{2}{\pi}\int dr \frac{r^2}{1+z(r)}\frac{d}{d\rho}j_\ell(\kappa\rho)\frac{d}{dr}\left[\delta\bar{T}_{\rm b}(r)j_\ell(kr)\right]. \tag{6.30}$$

Here we note that at low redshifts the brightness temperature signal varies only slowly and we assume that $d\delta\bar{T}_{\rm b}(r)/dr \approx 0$, which allows us to eliminate derivatives from the above expression. The only derivatives remaining are those of spherical bessel functions where we can use the following identity

$$\frac{d}{dx}j_\ell(x) = j_{\ell-1}(x) - \frac{l+1}{x}j_\ell(x), \tag{6.31}$$

remembering that for this calculation we have $\ell > 0$. Next, we introduce the explicit redshift evolution of the density field. It is easy to see that the growth function carries through to this point, and we can change the variable of integration to obtain

$$\begin{aligned} \mathcal{M}_\ell(k,\kappa) = &\frac{2c}{\pi}\int dz \frac{kr^2(z)}{(1+z)H(z)}\delta\bar{T}_{\rm b}(z)D_+(z) \\ &\times\left\{ j_{\ell-1}[\kappa\rho(z)]j_{\ell-1}[kr(z)] - \frac{\ell+1}{\kappa\rho(z)}j_\ell[\kappa\rho(z)]j_{\ell-1}[kr(z)] \right. \\ &\left. -\frac{\ell+1}{kr(z)}j_{\ell-1}[\kappa\rho(z)]j_\ell[kr(z)] + \frac{(\ell+1)^2}{k\kappa r(z)\rho(z)}j_\ell[\kappa\rho(z)]j_\ell[kr(z)]\right\}, \end{aligned} \tag{6.32}$$

where we have factored out one factor of κ for computational ease when introducing the power spectrum. Therefore, the spherical harmonic spherical bessel coefficient can be written as

$$a_{\ell m}^{(s)}(k) = f(\Omega_{\rm m})\int d\kappa\kappa\mathcal{M}_\ell(k,\kappa)\delta_{\ell m}(\kappa) + a_{\ell m}^{(r)}(k), \tag{6.33}$$

and we thus find that the effects of redshift space distortions introduce an additional term in the 3D coefficient. Finally, a similar calculation as in section 6.2 shows that the power

spectrum results in

$$C_\ell^{(s)}(k,k') = C_\ell^{(r)}(k,k') + f(\Omega_\mathrm{m}) \int d\kappa \kappa P(\kappa) \left[M_\ell(k,\kappa)\mathcal{M}_\ell(k',\kappa) + \mathcal{M}_\ell(k,\kappa)M_\ell(k',\kappa) \right] + f^2(\Omega_\mathrm{m}) \int d\kappa P(\kappa)\mathcal{M}_\ell(k,\kappa)\mathcal{M}_\ell(k',\kappa). \tag{6.34}$$

We finally see that the effects of redshift space distortions result in a perturbation to the real space power spectrum, probing the function $f(\Omega_\mathrm{m})$ directly.

6.4 Observational effects

In the context of 21cm intensity mapping experiments, real observations will be split into a series of tomographic slices. The instrumental response within each observed frequency bin can be modelled as a Gaussian window function, thus changing the power spectrum from equations (6.13) and (6.34). These equations are a good approximation for the power spectrum for observations with very closely separated frequency bins, or for modes larger than the bin separation, $k \lesssim \pi/\Delta r$. In general, however, a continuous radial projection through the spherical Bessel functions will be inappropriate for tomographic observations. Let the observed signal from a frequency slice labelled by an index i be

$$\delta T_{\mathrm{b},i}^{\mathrm{obs}}(\hat{\boldsymbol{n}}) = \int dz W_{\nu_i}(z) \delta T_\mathrm{b}[\boldsymbol{\rho}(z)], \tag{6.35}$$

where $W_{\nu_i}(z)$ denotes the Gaussian window function centred on ν_i, the central frequency of the ith tomographic slice. The radial dependence of the observed signal is then set by the frequency bin and we transform the observed signal into spherical harmonic space,

$$a^i_{\ell m} = \int d^2\hat{\boldsymbol{n}} dz W_{\nu_i}(z) \delta T_\mathrm{b}[\boldsymbol{\rho}(z)] Y^*_{\ell m}(\hat{\boldsymbol{n}}). \tag{6.36}$$

As in the previous sections, when expressing the signal in terms of fluctuations in the density field, the 0th order term vanishes for $\ell \neq 0 \neq m$. For all modes except the monopole, we introduce the inverse transform of the density fluctuations and resolve the angular integral to get

$$a^i_{\ell m} = \sqrt{\frac{2}{\pi}} \int dz d\kappa \delta_{\ell m}(\kappa) j_\ell[\kappa\rho(z)] W_{\nu_i}(z) \delta\bar{T}_\mathrm{b}(z) b_\mathrm{HI}(z) D_+(z). \tag{6.37}$$

Comparing this result to equation (6.9), we see that we find a similar result as in the continuous case, except that the mixing matrix M is projected onto the ith tomographic

bin as

$$a^i_{\ell m} = \int d\kappa M^i_\ell(\kappa)\delta_{\ell m}(\kappa), \tag{6.38}$$

with

$$M^i_\ell(\kappa) = \sqrt{\frac{2}{\pi}} \int dz j_\ell[\kappa\rho(z)] W_{v_i}(z)\delta\bar{T}_{\rm b}(z) b_{\rm HI}(z) D_+(z). \tag{6.39}$$

The cross power between two frequency slices i and j is then

$$C^{i,j}_\ell = \int d\kappa \kappa^2 P(\kappa) M^i_\ell(\kappa) M^j_\ell(\kappa). \tag{6.40}$$

which has a similar form to the result found in section 6.2.

Additionally, future 21cm observations, although probing large areas of the sky, may be limited to a subset of the full-sky information. This can occur either through observational constraints which limit experiments to observe less than the full-sky, or by some regions in the observed field being inadequate for the desired analysis and one wanting to mask out those regions from the observation. Both cases can be modelled by the application of an angular mask to the signal model, such that

$$\delta T^{\rm obs}_{\rm b}(\boldsymbol{r}) = \delta T_{\rm b}(\boldsymbol{r}) X(\hat{\boldsymbol{n}}), \tag{6.41}$$

where $X(\hat{\boldsymbol{n}})$ gives the sensitivity to the signal over the sky. As Heavens & Taylor (1995) note, an analysis in spherical coordinates can naturally incorporate this angular mask, and we find that the coefficients assume a similar form for the 21cm analysis as presented in their work,

$$a^{\rm obs}_{\ell m}(k) = \int d\kappa \kappa^2 \sum_{\ell' m'} M_{\ell\ell'}(k,\kappa)\delta_{\ell' m'}(\kappa) X^{mm'}_{\ell\ell'}, \tag{6.42}$$

where we define $M_{\ell\ell'}$ similarly to equation (6.10) as

$$M_{\ell\ell'}(k,\kappa) \equiv \frac{2c}{\pi} \int dz \frac{r^2(z)}{H(z)} \delta\bar{T}_{\rm b}(z) b_{\rm HI}(z) D_+(z) j_\ell[kr(z)] j_{\ell'}[\kappa\rho(z)], \tag{6.43}$$

and

$$X^{mm'}_{\ell\ell'} = \int d^2\hat{\boldsymbol{n}} Y^*_{\ell m}(\hat{\boldsymbol{n}}) Y_{\ell' m'}(\hat{\boldsymbol{n}}) X(\hat{\boldsymbol{n}}). \tag{6.44}$$

When no mask is applied or the observations achieve full-sky coverage, then the mask $X(\hat{\boldsymbol{n}}) = 1$ and $X^{mm'}_{\ell\ell'} = \delta^{\rm K}_{\ell\ell'}\delta^{\rm K}_{mm'}$, reducing the coefficient to that found in equation (6.9).

6.5 Summary

The cosmic 21cm signal is an inherently three-dimensional signal due to the ubiquitous nature of atomic hydrogen in the Universe. Radiation of this kind will come to us from all distances across the entire sky. When next generation telescopes observe large regions of the sky, the flat-sky approximation that many analysis techniques are based on will eventually become inadequate in characterizing the signal, potentially leading to systematic errors. We have introduced the concept of a 3D analysis in the natural spherical basis that is the spherical harmonic spherical Bessel function basis in the context of 21cm observations as a means to analyse 21cm data from such large area observations. The resulting 3D power spectrum is an integral over mixing matrices, characteristic for the 3D analysis. Redsift space distortions due to the peculiar motion of the gas emitting the 21cm radiation leads to the pertubation of the real space spectrum.

In this chapter we have derived expressions of the 3D power spectrum, which provides a framework for analysis of 21cm surveys for which the flat-sky approximation is inadequate. In this formalism, radial and transverse effects naturally separate and we therefore propose this formalism for use in large angle 21cm studies. The computation of the power spectrum is however numerically difficult and requires a fast evaluation of highly oscillatory integrals. Further work is thus required to evaluate the power spectrum in this formalism and determine the level at which statistical and systematic errors of a 21cm study may be affected when using this basis.

Part V

Conclusions

"You're not Dirac, Claude."

– Jonathan R. Pritchard.

Chapter 7

Conclusion

7.1 Thesis conclusion

The standard cosmological model, ΛCDM, has been carefully created from a series of theoretical advances throughout the 20th century. The details of the model have been confirmed in the recent decades with the growing number of precise measurements of the anisotropic nature of the CMB and the statistical distribution of structure. As a result, we have a model of the origin, present and fate of the Universe closer to the truth than ever before. There remain however a large number of unsolved questions, both on the theoretical side as well as observationally. We know from supernovae observations that a strange pressure has taken over the energy budget of the Universe causing an accelerated expansion of space-time. Though theoretically modelled well by a cosmological constant, we do not know of a mechanism creating this dark energy which represents around 70% of all the energy in the Universe. Further, not even the majority of gravitational matter is understood well. Only about 5% of the energy content can is in the form of baryonic matter, leaving about 25% as dark matter, currently unexplained by the standard model of particle physics. There exist also a number of statistically significant observational discrepancies between different cosmological probes, which all cause tension for the model. As an example, the locally measured expansion rate from the distance ladder is consistently faster than that inferred from CMB observations. Furthermore, epochs such as the Dark Ages and the Epoch of Reionization remain unobserved, yet present crucial developmental stages of the Universe.

Many of these current shortcomings in our understanding will be addressed in the near future by a new generation of land- and space-based observatories. Whether it is projects like the James-Webb Space Telescope, trying to expand on the success of the Hubble Space Telescope, the Large Synoptic Survey Telescope and Euclid raising the bar for dark energy and dark matter surveys, or the Square Kilometre Array peering into

unprecedented depths for galaxy surveys and able to observe the Epoch of Reionization, the coming decade will experience an explosion of high quality cosmological data. As a consequence of these technological advances, the astronomical community is tasked with the refinement of their analysis techniques which take full advantage of the incoming data. This thesis has identified two avenues to engage with the growing quantity and quality of the observed data in the context of 21cm observations. The two main parts of the thesis thus discuss numerical and analytic analysis of the 21cm signal.

In part III we highlight the potential of machine learning techniques applied in the data analysis of cosmic 21cm observations. Inferring the signatures of the first stars and galaxies from EoR observations requires model evaluations which are potentially numerically expensive. With the large influx of data upcoming, efficient numerical emulation of the models can help to increase the rate at which the data can be analysed. Chapter 4 therefore presents the use of an artificial neural network to complete this task. We train our network on a training sample of 21cm power spectra obtained by fast semi-numerical simulations of the EoR. Using the latin-hypercube distribution technique for the training parameters, an optimal learning strategy can be achieved. Not only do we show that a speed-up of 3 orders of magnitude with respect to conventional model evaluation can be achieved during the parameter inference, but we also show that with low size training sets, the network still performs well enough to be used in an application with observational data. This opens a wide range of applications for neural networks to be used in data analysis of upcoming 21cm observations. Whether it would be the first-time use of fully numerical simulations in a parameter inference study or the inclusion of the network in a model comparison study, neural networks, and more broadly machine learning, will play a crucial role in tackling the large cosmological data sets of the 21st century.

In part IV we investigate two analytical techniques to enhance 21cm data analysis. With the prospect of obtaining 21cm maps from intensity mapping surveys, the power spectrum of the observed brightness temperature fluctuations can be readily measured. Due to the non-linear collapse of structure, these maps should contain a significant non-Gaussian component which will be able to aid in constraining the cosmological parameters. This non-Gaussian component will be manifest in higher order statistics of the field, and we therefore analyse the expected bispectrum due to this non-linear collapse. Furthermore, we compute the bispectrum due to the cross-correlation of lensing and the ISW effect. These line of sight effects will distort the photon path and a bispectrum contribution is expected. We find that the bispectrum can indeed further constrain cosmology, beyond the power spectrum capabilities. SKA-mid intensity mapping surveys examining the bispectrum will be able to obtain marginalized errors of order $< 0.1\%$. The lensing-ISW contribution to this measurement will be negligible.

Following this analysis, we introduce an analytic framework within which future full-sky or near full-sky 21cm surveys could be analysed. For a three dimensional signal, the natural basis in which to analyse the data is the spherical harmonic spherical Bessel function basis. We derive expressions for the 3D 21cm power spectrum as well as consider the effects of redshift space distortions due to peculiar velocities of hydrogen gas in this basis. Finally, we consider a radial window function as well as an angular mask and adapt the 3D power spectrum for when data is radially binned and parts of the sky are masked out.

7.2 Scientific outlook

The work described in this thesis thus highlights the possible gain that improving our analysis techniques will bring in a new era of 21cm cosmology. As the new generation of telescopes sees its first light in the upcoming decade, we are well to be prepared with the adequate tools to face the challenges that accompany them. Here we discuss a short list of still outstanding developments based on the work presented in this thesis which we identify as the most important extensions to our work.

The machine learning paradigm presented in chapter 4 has proven powerful already and a potentially powerful extension should be an application of the emulation strategy to larger scale numerical simulations. In conjunction with the appearance of public data bases, such as Semelin et al. (2017), this presents an immediately attainable, yet important test to our work. If the emulation model, which we have tested on a fast semi-numerical code, can prove to be robust when applied on a fully numerical simulation, then a parameter inference can be used in conjunction with numerical simulations for the first time.

The largest hurdle for 21cm bispectrum observations to overcome is that of systematics induced by the interplay of foregrounds and calibration artefacts. These systematics can potentially swamp the signal and need to be understood well before the bispectrum will become a useful observable. In the same vein, in our analysis presented in chapter 5 we assume a simple model to incorporate foreground residuals, which does not take into account real foreground reduction methods that would be applied to raw data before it could be analysed. Here we propose that an extension of our work should include a thorough analysis of foreground residuals in the context of the bispectrum.

Lastly, in chapter 6 we have introduced the analytic expressions of the 21cm power spectrum in a 3D basis. Due to the highly oscillatory nature of the mixing matrices in the power spectrum, computing the value of the 3D power spectrum has shown to be a computationally intensive exercise. An analysis of the accuracy of the Limber approxi-

mation and an exploration of fast integration methods for Bessel functions is desirable for accurate power spectrum computations.

Bibliography

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, Phys. Rev. Lett., 116, 61102

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, Physical Review Letters, 116, 241103

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, Physical Review Letters, 118, 221101

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017b, Physical Review Letters, 119, 141101

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017c, Physical Review Letters, 119, 161101

Abu-Mostafa, Yaser S., . 2012, Learning from data : a short course (AMLBook.com)

Acquaviva, V., Bartolo, N., Matarrese, S., et al. 2003, Nucl. Phys. B, 667, 119

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2014, ApJS, 211, 17

Alonso, D., Bull, P., Ferreira, P. G., et al. 2015, Mon. Not. Roy. Astron. Soc., 447, 400

Alonso, D., Ferreira, P. G., & Santos, M. G. 2014, Mon. Not. Roy. Astron. Soc., 444, 3183

Anderson, L., Aubourg, E., Bailey, S., et al. 2012, Mon. Not. Roy. Astron. Soc., 427, 3435

Aristizabal Sierra, D. & Fong, C. S. 2018, Physics Letters B, 784, 130

Asirvadam, V., McLoone, S., & Irwin, G. 2004, Proc. 2004 IEEE Int. Conf. Control Appl. 2004., 586

Assassi, V., Simonović, M., & Zaldarriaga, M. 2017, J. Cosmol. Astropart. Phys., 11, 054

Baek, S., Di Matteo, P., Semelin, B., et al. 2009, A&A, 495, 389

Baek, S., Semelin, B., Di Matteo, P., et al. 2010, A&A, 523, A4

Bagla, J. S., Khandai, N., & Datta, K. K. 2010, Mon. Not. Roy. Astron. Soc., 407, 567

Ball, N. M. & Brunner, R. J. 2010, International Journal of Modern Physics D, 19, 1049

Ball, N. M., Brunner, R. J., Myers, A. D., et al. 2006, ApJ, 650, 497

Balogh, M., Eke, V., Miller, C., et al. 2004, Mon. Not. Roy. Astron. Soc., 348, 1355

Bardeen, J. M., Bond, J. R., Kaiser, N., et al. 1986, ApJ, 304, 15

Barkana, R. 2018, Nature, 555, 71

Barkana, R. & Loeb, A. 2005, ApJL, 624, L65

Barkana, R., Outmezguine, N. J., Redigolo, D., et al. 2018, ArXiv e-prints

Bartelmann, M. & Schneider, P. 2001, Physics Reports, 340, 291

Bartolo, N., Komatsu, E., Matarrese, S., et al. 2004, Physics Reports, 402, 103

Battiti, R. 1992, Neural Comput., 4, 141

Battye, R., Browne, I., Chen, T., et al. 2016, ArXiv e-prints

Battye, R. A., Browne, I. W. A., Dickinson, C., et al. 2013, Mon. Not. Roy. Astron. Soc., 434, 1239

Beardsley, A. P., Hazelton, B. J., Sullivan, I. S., et al. 2016, ApJ, 833, 102

Beardsley, A. P., Morales, M. F., Lidz, A., et al. 2015, ApJ, 800, 128

Bennett, C. L., Bay, M., Halpern, M., et al. 2003, ApJ, 583, 1

Bennett, C. L., Larson, D., Weiland, J. L., et al. 2014, ApJ, 794, 135

Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, ApJS, 208, 20

Birrer, S., Treu, T., Rusu, C. E., et al. 2019, Mon. Not. Roy. Astron. Soc.

Bishop, C. M. 2006, Pattern Recognition and Machine Learning (Secaucus, NJ, USA: Springer-Verlag New York)

Bond, J. R., Cole, S., Efstathiou, G., et al. 1991, Astrophys. J., 379, 440

Bonvin, V., Courbin, F., Suyu, S. H., et al. 2017, Mon. Not. Roy. Astron. Soc., 465, 4914

Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2009, ApJ, 695, 183

Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., et al. 2018a, Nature, 555, 67

Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., et al. 2018b, Nature, 564, E35

Bradley, R. F., Tauscher, K., Rapetti, D., et al. 2018, ArXiv e-prints

Braun, R. 2012, ApJ, 749, 87

Bull, P., Camera, S., Kelley, K., et al. 2018, ArXiv e-prints

Bull, P., Camera, S., Raccanelli, A., et al. 2014, in Adv. Astrophys. with Sq. Km. Array, 1–15

Bull, P., Ferreira, P. G., Patel, P., et al. 2015, ApJ, 803, 21

Burns, J. O., Lazio, J., Bale, S., et al. 2012, Advances in Space Research, 49, 433

Camera, S., Santos, M. G., Ferreira, P. G., et al. 2013, Phys. Rev. Lett., 111, 171302

Chang, T.-C., Pen, U.-L., Peterson, J. B., et al. 2008, Physical Review Letters, 100, 091303

Chapman, E., Bonaldi, A., Harker, G., et al. 2014a, in Adv. Astrophys. with Sq. Km. Array

Chapman, E., Bonaldi, A., Harker, G., et al. 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), 5

Chapman, E., Zaroubi, S., & Abdalla, F. 2014b, ArXiv e-prints

Chapman, E., Zaroubi, S., Abdalla, F. B., et al. 2016, Mon. Not. Roy. Astron. Soc., 458, 2928

Chen, X. 2010, Advances in Astronomy, 2010, 638979

Chen, X. 2015, IAU General Assembly, 22, 2252187

Chen, X. & Kamionkowski, M. 2004, Phys. Rev. D, 70, 043502

Cheng, B. & Titterington, D. M. 1994, Statistical Science, 9, 2

Ciardi, B., Stoehr, F., & White, S. D. M. 2003, Mon. Not. Roy. Astron. Soc., 343, 1101

Colless, M., Dalton, G., Maddox, S., et al. 2001, Mon. Not. Roy. Astron. Soc., 328, 1039

Crighton, N. H. M., Murphy, M. T., Prochaska, J. X., et al. 2015, Mon. Not. Roy. Astron. Soc., 452, 217

Crittenden, R. G. & Turok, N. 1996, Physical Review Letters, 76, 575

Cybenko, G. 1989, Mathematics of Control, Signals and Systems, 2, 303

D'Aloisio, A., Zhang, J., Shapiro, P. R., et al. 2013, Mon. Not. Roy. Astron. Soc., 433, 2900

de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., et al. 2013, Experimental Astronomy, 35, 25

DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, PASP, 129, 045001

Deguchi, S. & Watson, W. D. 1985, Astrophys. J., 290, 578

Delhaize, J., Meyer, M. J., Staveley-Smith, L., et al. 2013, Mon. Not. Roy. Astron. Soc., 433, 1398

Di Dio, E., Durrer, R., Maartens, R., et al. 2018, arXiv e-prints

Dillon, J. S., Neben, A. R., Hewitt, J. N., et al. 2015, Phys. Rev. D., 91, 123011

Dodelson, S. 2003, Modern Cosmology (Academic Press)

Dyson, F. W., Eddington, A. S., & Davidson, C. 1920, Philosophical Transactions of the Royal Society of London Series A, 220, 291

Eisenstein, D. J. & Hu, W. 1997, 20

Evoli, C., Mesinger, A., & Ferrara, A. 2014, J. Cosmol. Astropart. Phys., 11, 17

Ferdosi, B. J., Buddelmeijer, H., Trager, S. C., et al. 2011, A&A, 531, A114

Fergusson, J., Liguori, M., & Shellard, E. 2012, J. Cosmol. Astropart. Phys., 2012, 032

Feroz, F., Hobson, M. P., & Bridges, M. 2009, Mon. Not. Roy. Astron. Soc., 398, 1601

Field, G. B. 1958, Proceedings of the IRE, 46, 240

Field, G. B. 1959, ApJ, 129, 536

Fisher, R. A. 1935, J. R. Stat. Soc., 98, 39

Font-Ribera, A., Miralda-Escude, J., Arnau, E., et al. 2012, J. Cosmol. Astropart. Phys., 11, 059

Foreman-Mackey, D., Hogg, D. W., Lang, D., et al. 2013, PASP, 125, 306

Franco-Arcega, A., Flores-Flores, L. G., & Gabbasov, R. F. 2013, in 2013 12th Mexican International Conference on Artificial Intelligence, 181–186

Fraser, S., Hektor, A., Hütsi, G., et al. 2018, Physics Letters B, 785, 159

Fry, J. N. 1984, ApJ, 279, 499

Furlanetto, S. R. 2006, Mon. Not. R. Astron. Soc., 371, 867

Furlanetto, S. R. & Furlanetto, M. R. 2007, Mon. Not. Roy. Astron. Soc., 374, 547

Furlanetto, S. R. & Furlanetto, M. R. 2007, Mon. Not. R. Astron. Soc., 379, 130

Furlanetto, S. R., Mirocha, J., Mebane, R. H., et al. 2017, Mon. Not. Roy. Astron. Soc., 472, 1576

Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, Physics Reports, 433, 181

Furlanetto, S. R. & Pritchard, J. R. 2006, Mon. Not. R. Astron. Soc., 372, 1093

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1

Gal, Y. 2016, PhD thesis, University of Cambridge

Gillon, M., Jehin, E., Lederer, S. M., et al. 2016, Nature, 533, 221

Gillon, M., Triaud, A. H. M. J., Demory, B.-O., et al. 2017, Nature, 542, 456

Giovi, F., Baccigalupi, C., & Perrotta, F. 2003, Phys. Rev. D., 68, 123002

Gnedin, N. Y. & Shaver, P. a. 2004, Astrophys. J., 608, 611

Goldberg, D. M. & Spergel, D. N. 1999, Phys. Rev. D., 59, 103002

Greig, B. & Mesinger, A. 2015, Mon. Not. Roy. Astron. Soc., 449, 4246

Greig, B. & Mesinger, A. 2017, ArXiv e-prints

Greig, B., Mesinger, A., & Koopmans, L. V. E. 2015, ArXiv e-prints

Griffiths, D. J. 2005, Introduction to Quantum Mechanics, 2nd edn. (Pearson Prentice Hall), 468

Hall, A., Bonvin, C., & Challinor, A. 2013, Phys. Rev. D., 87, 064026

Harker, G. J. A., Pritchard, J. R., Burns, J. O., et al. 2012, Mon. Not. Roy. Astron. Soc., 419, 1070

Heavens, A. 2003, Mon. Not. R. Astron. Soc., 343, 1327

Heavens, A. 2009, ArXiv e-prints

Heavens, A., Fantaye, Y., Sellentin, E., et al. 2017, Physical Review Letters, 119, 101301

Heavens, A. F. & Taylor, A. N. 1995, Mon. Not. Roy. Astron. Soc., 275, 483

Heitmann, K., Bingham, D., Lawrence, E., et al. 2016, ApJ, 820, 108

Heitmann, K., Higdon, D., White, M., et al. 2009, ApJ, 705, 156

Heitmann, K., Lawrence, E., Kwan, J., et al. 2013, Astrophys. J., 780, 111

Heitmann, K., Lawrence, E., Kwan, J., et al. 2014, ApJ, 780, 111

Heymans, C., Grocutt, E., Heavens, A., et al. 2013, Mon. Not. Roy. Astron. Soc., 432, 2433

Heymans, C., Van Waerbeke, L., Miller, L., et al. 2012, Mon. Not. Roy. Astron. Soc., 427, 146

Hill, J. C. 2018, Phys. Rev. D., 98, 083542

Hills, R., Kulkarni, G., Meerburg, P. D., et al. 2018, Nature, 564, E32

Hobson, M. P., Jaffe, A. H., Liddle, A. R., et al. 2010, Bayesian Methods in Cosmology

Hogan, C. J. & Rees, M. J. 1979, Mon. Not. R. Astron. Soc., 188, 791

Hogg, D. W. 1999, ArXiv Astrophysics e-prints

Houston, N., Li, C., Li, T., et al. 2018, Physical Review Letters, 121, 111301

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

Iliev, I. T., Ciardi, B., Alvarez, M. A., et al. 2006a, Mon. Not. Roy. Astron. Soc., 371, 1057

Iliev, I. T., Mellema, G., Ahn, K., et al. 2014, Mon. Not. Roy. Astron. Soc., 439, 725

Iliev, I. T., Mellema, G., Pen, U.-L., et al. 2006b, Mon. Not. Roy. Astron. Soc., 369, 1625

Ivezic, Z., J., C. A., T., V. J., et al. 2014, Statistics, data mining, and machine learning in astronomy: a practical python guide for the analysis of survey data (Princeton: Princeton University Press)

Jacobs, D. C., Pober, J. C., Parsons, A. R., et al. 2015, ApJ, 801, 51

Jarvis, M., Sheldon, E., Zuntz, J., et al. 2016, Mon. Not. Roy. Astron. Soc., 460, 2245

Jennings, W. D., Watkinson, C. A., Abdalla, F. B., et al. 2019, Mon. Not. Roy. Astron. Soc., 483, 2907

Jeong, D. & Komatsu, E. 2009, ApJ, 703, 1230

Johnson, M., Moore, L., & Ylvisaker, D. 1990, J. Stat. Plan. Inference, 26, 131

Joliffe, Ian T., . 1986, Principal component analysis (Springer-Verlag New York)

Joudaki, S., Mead, A., Blake, C., et al. 2017, Mon. Not. Roy. Astron. Soc., 471, 1259

Kaiser, N. 1987, Mon. Not. R. Astron. Soc., 227, 1

Karagiannis, D., Lazanu, A., Liguori, M., et al. 2018, Mon. Not. Roy. Astron. Soc., 478, 1341

Kern, N. S., Liu, A., Parsons, A. R., et al. 2017, ArXiv e-prints

Kim, A. G., Linder, E. V., Miquel, R., et al. 2004, Mon. Not. Roy. Astron. Soc., 347, 909

Kim, J., Rotti, A., & Komatsu, E. 2013, Journal of Cosmology and Astroparticle Physics, 2013, 021

Köhlinger, F., Viola, M., Joachimi, B., et al. 2017, Mon. Not. Roy. Astron. Soc., 471, 4412

Komatsu, E. 2010, Classical and Quantum Gravity, 27, 124010

Komatsu, E., Afshordi, N., Bartolo, N., et al. 2009, in ArXiv Astrophysics e-prints, Vol. 2010, astro2010: The Astronomy and Astrophysics Decadal Survey

Komatsu, E. & Spergel, D. N. 2001, Phys. Rev. D, 63, 063002

Komatsu, E., Spergel, D. N., & Wandelt, B. D. 2005, ApJ, 634, 14

Kovetz, E. D., Poulin, V., Gluscevic, V., et al. 2018, ArXiv e-prints

Kragh, H. 2012, Journal for the History of Astronomy, 43, 347

Kremer, J., Stensbo-Smidt, K., Gieseke, F., et al. 2017, IEEE Intelligent Systems, 32, 16

Lah, P., Chengalur, J. N., Briggs, F. H., et al. 2007, Mon. Not. Roy. Astron. Soc., 376, 1357

Le, Q. V., Coates, A., Prochnow, B., et al. 2011, Proc. 28th Int. Conf. Mach. Learn., 265

Levin, D. 1982, Math. Comput., 38, 531

Levin, D. 1996, J. Comput. Appl. Math., 67, 95

Levin, D. 1997, J. Comput. Appl. Math., 78, 131

Lewis, A., Challinor, A., & Hanson, D. 2011, J. Cosmol. Astropart. Phys., 3, 018

Li, Y.-C. & Ma, Y.-Z. 2017, Phys. Rev. D, 96, 063525

Lidz, A., Zahn, O., McQuinn, M., et al. 2008, Astrophys. J., 680, 962

Liguori, M., Sefusatti, E., Fergusson, J. R., et al. 2010, Advances in Astronomy, 2010, 980523

Liu, A. & Parsons, A. R. 2016, Mon. Not. Roy. Astron. Soc., 457, 1864

Liu, A., Tegmark, M., Bowman, J., et al. 2009, Mon. Not. Roy. Astron. Soc., 398, 401

Loeb, A. & Furlanetto, S. R. 2013, The First Galaxies in the Universe (Princeton University Press), 540

Loverde, M. & Afshordi, N. 2008, Phys. Rev. D., 78, 123506

Maartens, R., Abdalla, F. B., Jarvis, M., et al. 2014, in Adv. Astrophys. with Sq. Km. Array, 1–12

Macaulay, E., Nichol, R. C., Bacon, D., et al. 2018, arXiv e-prints

MacCrann, N., Zuntz, J., Bridle, S., et al. 2015, Mon. Not. Roy. Astron. Soc., 451, 2877

Majumdar, S., Pritchard, J. R., Mondal, R., et al. 2018, Mon. Not. Roy. Astron. Soc.

Maldacena, J. 2003, Journal of High Energy Physics, 5, 013

Mangilli, A., Wandelt, B., Elsner, F., et al. 2013, Astron. Astrophys., 555, A82

Mao, Y., D'Aloisio, A., Zhang, J., et al. 2013, Phys. Rev. D., 88, 081303

Martin, A. M., Papastergis, E., Giovanelli, R., et al. 2010, ApJ, 723, 1359

Martinez-Gonzalez, E., Sanz, J. L., & Silk, J. 1990, Astrophys. J., 355, L5

Mashian, N., Oesch, P. A., & Loeb, A. 2016, Mon. Not. Roy. Astron. Soc., 455, 2101

Mason, C. A., Trenti, M., & Treu, T. 2015, Astrophys. J., 813, 21

Mcloone, S. F., Asirvadam, V. S., & Irwin, G. W. 2002, Proc. 2002 IEEE Int. Conf. Neural Networks, 2002., 2, 513

Meiksin, A. 2000, in Perspect. Radio Astron. Sci. with Large Antenna Arrays No. 1, 37–44

Mertens, F. G., Ghosh, A., & Koopmans, L. V. E. 2018, Mon. Not. Roy. Astron. Soc., 478, 3640

Mesinger, A., Ferrara, A., & Spiegel, D. S. 2013, Mon. Not. R. Astron. Soc., 431, 621

Mesinger, A. & Furlanetto, S. 2007, ApJ, 669, 663
Mesinger, A., Furlanetto, S., & Cen, R. 2011, Mon. Not. Roy. Astron. Soc., 411, 955
Mirocha, J. & Furlanetto, S. R. 2018, ArXiv e-prints
Morris, M. D. & Mitchell, T. J. 1995, J. Stat. Plan. Inference, 43, 381
Muñoz, J. B., Ali-Haïmoud, Y., & Kamionkowski, M. 2015, Phys. Rev. D., 92, 083508
Muñoz, J. B., Dvorkin, C., & Loeb, A. 2018, Physical Review Letters, 121, 121301
Munshi, D. & Heavens, A. 2010, Mon. Not. Roy. Astron. Soc., 401, 2406
Munshi, D., Valageas, P., van Waerbeke, L., et al. 2008, Phys. Rep., 462, 67
Nishizawa, A. J. 2014, Prog. Theor. Exp. Phys., 6B110
Noterdaeme, P., Petitjean, P., Carithers, W. C., et al. 2012, A&A, 547, L1
Oh, S. P. 2001, Astrophys. J., 553, 499
Oh, S. P. & Mack, K. J. 2003, Mon. Not. Roy. Astron. Soc., 346, 871
Olivari, L. C., Dickinson, C., Battye, R. A., et al. 2018, Mon. Not. Roy. Astron. Soc., 473, 4242
Ordovás-Pascual, I. & Sánchez Almeida, J. 2014, A&A, 565, A53
Paciga, G., Albert, J. G., Bandura, K., et al. 2013, Mon. Not. Roy. Astron. Soc., 433, 639
Padmanabhan, H., Roy Choudhury, T., & Refregier, A. 2016, Mon. Not. R. Astron. Soc., 458, 781
Parsons, A., Pober, J., McQuinn, M., et al. 2012, ApJ, 753, 81
Parsons, A. R., Backer, D. C., Foster, G. S., et al. 2010, AJ, 139, 1468
Parsons, A. R., Liu, A., Aguirre, J. E., et al. 2014, ApJ, 788, 106
Patil, A. H., Yatawatta, S., Koopmans, L. V. E., et al. 2017, ApJ, 838, 65
Patra, N., Subrahmanyan, R., Raghunathan, A., et al. 2013, Experimental Astronomy, 36, 319
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
Peebles, P. J. E. 1993, Principles of Physical Cosmology (Princeton University Press), 718
Penzias, A. A. & Wilson, R. W. 1965, ApJ, 142, 419
Pillepich, A., Porciani, C., & Matarrese, S. 2007, ApJ, 662, 1
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016a, A&A, 594, A17
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016b, A&A, 594, A13
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018a, ArXiv e-prints
Planck Collaboration, Akrami, Y., Arroja, F., et al. 2018b, ArXiv e-prints
Pober, J. C., Liu, A., Dillon, J. S., et al. 2014, ApJ, 782, 66
Pober, J. C., Parsons, A. R., DeBoer, D. R., et al. 2013, AJ, 145, 65
Pourtsidou, A., Bacon, D., Crittenden, R., et al. 2016, Mon. Not. Roy. Astron. Soc., 459, 863
Pourtsidou, A. & Metcalf, R. B. 2014, Mon. Not. Roy. Astron. Soc., 439, L36
Press, W. H. & Schechter, P. 1974, ApJ, 187, 425
Pritchard, J. & Loeb, A. 2010, Nature, 468, 772
Pritchard, J. R. & Furlanetto, S. R. 2006, Mon. Not. R. Astron. Soc., 367, 1057
Pritchard, J. R. & Loeb, A. 2008, Phys. Rev. D, 78, 1
Pritchard, J. R. & Loeb, A. 2012, Reports on Progress in Physics, 75, 086901

Prochaska, J. X., Herbert-Fort, S., & Wolfe, A. M. 2005, ApJ, 635, 123

Raccanelli, A., Shiraishi, M., Bartolo, N., et al. 2017, Physics of the Dark Universe, 15, 35

Rao, S. M., Turnshek, D. A., & Nestor, D. B. 2006, ApJ, 636, 610

Rasmussen, C. E. & Williams, C. K. I. 2005, Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning Series (Mit Press)

Raveri, M. 2016, Phys. Rev. D., 93, 043522

Rees, M. J. & Sciama, D. W. 1968, Nature, 217, 511

Renn, J., Sauer, T., & Stachel, J. 1997, Science, 275, 184

Repp, A., Szapudi, I., Carron, J., et al. 2015, Mon. Not. Roy. Astron. Soc., 454, 3533

Rhee, J., Zwaan, M. A., Briggs, F. H., et al. 2013, Mon. Not. Roy. Astron. Soc., 435, 2693

Riess, A. G., Casertano, S., Yuan, W., et al. 2018a, ApJ, 855, 136

Riess, A. G., Casertano, S., Yuan, W., et al. 2018b, ApJ, 861, 126

Riess, A. G., Macri, L. M., Hoffmann, S. L., et al. 2016, ApJ, 826, 56

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986, Nature, 323, 533

Sachs, R. K. & Wolfe, A. M. 1967, ApJ, 147, 73

Santos, M., Bull, P., Alonso, D., et al. 2014, in Adv. Astrophys. with Sq. Km. Array, 1–26

Santos, M., Bull, P., Alonso, D., et al. 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), 19

Santos, M. G., Cooray, A., & Knox, L. 2005, ApJ, 625, 575

Seager, S., Sasselov, D. D., & Scott, D. 1999, Astrophys. J., 523, L1

Semelin, B., Eames, E., Bolgar, F., et al. 2017, Mon. Not. Roy. Astron. Soc., 472, 4508

Shaw, J. R., Sigurdson, K., Pen, U.-L., et al. 2014, ApJ, 781, 57

Shimabukuro, H. & Semelin, B. 2017, Mon. Not. Roy. Astron. Soc., 468, 3869

Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1992, ApJL, 396, L1

Smoot, G. F. & Debono, I. 2017, A&A, 597, A136

Sobelman, I. I. 1979, Atomic spectra and radiative transitions, Springer series in chemical physics; 1 (Berlin; New York: Springer-Verlag)

Songaila, A. & Cowie, L. L. 2010, ApJ, 721, 1448

Spergel, D. N. & Goldberg, D. M. 1999, Phys. Rev. D., 59, 103001

Spurio Mancini, A., Taylor, P. L., Reischke, R., et al. 2018, Phys. Rev. D., 98, 103507

Stanley, M. 2003, Isis. Journal of the History of Science Society, 94, 57

Sunyaev, R. A. & Zeldovich, Y. B. 1972, A&A, 20, 189

Takada, M. & Jain, B. 2004, Mon. Not. Roy. Astron. Soc., 348, 897

Taylor, A. N., Kitching, T. D., Bacon, D. J., et al. 2007, Mon. Not. Roy. Astron. Soc., 374, 1377

Tegmark, M., Blanton, M. R., Strauss, M. A., et al. 2004, ApJ, 606, 702

Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, ApJ, 480, 22

Urban, N. M. & Fricker, T. E. 2010, Comput. Geosci., 36, 746

Vanderplas, J. & Connolly, A. 2009, AJ, 138, 1365

Vasconcellos, E. C., de Carvalho, R. R., Gal, R. R., et al. 2011, AJ, 141, 189

Verde, L., Heavens, A. F., & Matarrese, S. 2000, Mon. Not. Roy. Astron. Soc., 318, 584

Verde, L. & Spergel, D. N. 2002, Phys. Rev. D., 65, 043007

Villaescusa-Navarro, F., Genel, S., Castorina, E., et al. 2018, ArXiv e-prints

Voytek, T. C., Natarajan, A., Jáuregui García, J. M., et al. 2014, ApJL, 782, L9

Walsh, D., Carswell, R. F., & Weymann, R. J. 1979, Nature, 279, 381

Watkinson, C. A. & Pritchard, J. R. 2014, Mon. Not. R. Astron. Soc., 443, 3090

Weinberg, S. 2005, Phys. Rev. D., 72, 043514

Wolz, L., Abdalla, F. B., Blake, C., et al. 2014, Mon. Not. Roy. Astron. Soc., 441, 3271

Wouthuysen, S. A. 1952, Astron. J., 57, 31

Zahn, O., Mesinger, A., Mcquinn, M., et al. 2011, Mon. Not. R. Astron. Soc., 414, 727

Zaldarriaga, M., Furlanetto, S. R., & Hernquist, L. 2004, ApJ, 608, 622

Zel'dovich, Y. B. 1970, A&A, 5, 84

Ziegel, E., Press, W., Flannery, B., et al. 1987, Technometrics, 29, 501

Zieser, B. & Merkel, P. M. 2016, Mon. Not. Roy. Astron. Soc., 459, 1586

Zwaan, M. A., Liske, J., Péroux, C., et al. 2015, Mon. Not. Roy. Astron. Soc., 453, 1268

Zwaan, M. A., Meyer, M. J., Staveley-Smith, L., et al. 2005, Mon. Not. Roy. Astron. Soc., 359, L30

Zygelman, B. 2005, Astrophys. J., 622, 1356

Appendix A

Appendix for Chapter 2

A.1 Equation of motion for photons in a perturbed metric

Consider the weakly perturbed FRW metric,

$$ds^2 = \left(1 + \frac{2\Phi}{c^2}\right)c^2dt^2 - \left(1 - \frac{2\Psi}{c^2}\right)a^2(t)[dr^2 + S_k^2(r)d\beta^2], \tag{A.1}$$

where Φ and Ψ denote the perturbations to the Newtonian potential and curvature respectively. We further define the comoving spherical coordinates r, θ and ϕ, and $d\beta \equiv \sqrt{d\theta + \sin^2\theta d\phi}$. Considering the distortions to photons travelling in such a perturbed metric, it is convenient to orient the polar axis of the coordinate system along the photon travel path, and thus we can define the angles $\theta_x = \theta\cos\phi$ and $\theta_y = \theta\sin\phi$. Additionally, using conformal time $d\eta = cdt/a(t)$, the line element becomes

$$ds^2 = a^2\left\{\left(1 + \frac{2\Phi}{c^2}\right)d\eta^2 - \left(1 - \frac{2\Psi}{c^2}\right)\left[dr^2 + S_k^2(r)\left(d\theta_x^2 + d\theta_y^2\right)\right]\right\}. \tag{A.2}$$

Next, the equation of motion for the photons is then given by the geodesic equation, or equivalently the Euler-Lagrange equation,

$$\frac{\partial L^2}{\partial x^\mu} - \frac{d}{dp}\left(\frac{\partial L^2}{\partial \dot{x}^\mu}\right) = 0, \tag{A.3}$$

where p represents monotonically increasing affine parameter and the Lagrangian is defined through

$$L^2 = g_{\mu\nu}\frac{\partial x^\mu}{\partial p}\frac{\partial x^\nu}{\partial p}. \tag{A.4}$$

Using this definition of the Lagrangian, the Euler-Lagrange equation can be written as

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda}\dot{x}^\mu\dot{x}^\nu - \frac{d}{dp}\left(g_{\mu\lambda}\dot{x}^\mu + g_{\lambda\nu}\dot{x}^\nu\right) = 0, \tag{A.5}$$

using the fact that $\partial x^\mu/\partial x^\lambda = \delta^\mu_\lambda$. From the metric in equation (A.2) assuming a flat geometry, the $x^\lambda = \theta_x$ term can be shown to be.

$$\frac{2a^2}{c^2}\frac{\partial\Phi}{\partial\theta_x}\dot{\eta}^2 + \frac{2a^2}{c^2}\frac{\partial\Psi}{\partial\theta_x}\left(\dot{r}^2 + r^2\dot{\theta}_x^2 + r^2\dot{\theta}_y^2\right) - \frac{d}{dp}\left[-2a^2r^2\left(1 - \frac{2\Psi}{c^2}\right)\dot{\theta}_x\right] = 0. \tag{A.6}$$

For a 0^{th} order analysis, we can see that a photon travelling on a null geodesic implies that the parenthesis reduces to $\dot{\eta}^2$. Further, $\dot{\theta}_x$ is first order in the potentials, which means that $\Psi\dot{\theta}_x$ in the angular brackets can be ignored being a second order term. To complete the computation we solve the Euler-Lagrange equation to 0^{th} order for $x_\lambda = \eta$. We find

$$-2\frac{d}{dp}(a^2\dot{\eta}) = 0, \tag{A.7}$$

which integrated and normalized appropriately gives

$$\dot{\eta} = a^{-2}. \tag{A.8}$$

This result is then also used to write $a^2\dot{\theta}_x = d\theta_x/d\eta$ using the chain rule, and thus equation (A.6) becomes

$$\frac{d^2\theta_x}{d\eta^2} - \frac{2}{r}\frac{d\theta_x}{d\eta} = -\frac{1}{c^2r^2}\frac{\partial}{\partial\theta_x}(\Phi + \Psi). \tag{A.9}$$

One can relate this equation back to the angular comoving coordinates with $x_i \equiv r\theta_i$ for $i = 1, 2$, identifying that

$$\frac{d^2x_i}{d\eta^2} = r\frac{d^2\theta_i}{d\eta^2} - 2\frac{d\theta_i}{d\eta}. \tag{A.10}$$

Solving the Euler-Lagrange equation similarly for θ_y, it is thus easy to see that the equation of motion for the transverse coordinates of a photon in a weakly perturbed, flat Universe becomes

$$\frac{d^2\mathbf{x}}{d\eta^2} = -\frac{1}{c^2}\nabla(\Phi + \Psi). \tag{A.11}$$

Appendix B

Appendix for Chapter 3

B.1 Brightness temperature

We start with the radiative transfer equation ignoring the effects of scattering, which is defined as

$$\frac{dI_\nu}{ds} = -\alpha_\nu I_\nu + j_\nu, \tag{B.1}$$

where $I_\nu(s)$ is the specific intensity of a source at position s, $\alpha_\nu(s)$ is the absorption coefficient at s, and $j_\nu(s)$ is the emission coefficient at s. Furthermore, the optical depth of a medium extending from an initial position s_0 to s is defined to be the integral of its absorption coefficient over the distance,

$$\tau_\nu(s_0, s) \equiv \int_{s_0}^{s} \alpha_\nu(s')ds'. \tag{B.2}$$

Multiplying both sides of equation (B.1) by a factor of $e^{\tau_\nu(s_0,s)}$, rearranging and observing that $\alpha_\nu(s) = \frac{d\tau_\nu(s_0,s)}{ds}$, one finds

$$\frac{d}{ds}\left[e^{\tau_\nu(s_0,s)} I_\nu(s)\right] = j_\nu(s)e^{\tau_\nu(s_0,s)}. \tag{B.3}$$

This result can then integrated from s_0 to s, such that

$$e^{\tau_\nu(s_0,s)} I_\nu(s) - e^{\tau_\nu(s_0,s_0)} I_\nu(s_0) = \int_{s_0}^{s} j_\nu(s')e^{\tau_\nu(s_0,s')}ds'. \tag{B.4}$$

From equation (B.2) we see that

$$e^{\tau_\nu(s_0,s_0)} = 1, \tag{B.5}$$

and furthermore, we write $\tau_\nu = \tau_\nu(s_0, s)$, so that after rearranging one finds

$$I_\nu(s) = I_\nu(s_0)e^{-\tau_\nu} + e^{-\tau_\nu}\int_{s_0}^{s} j_\nu(s')e^{\tau_\nu(s_0,s')}ds'. \tag{B.6}$$

Then, we use the relation between j and α,

$$\frac{j_\nu(s)}{\alpha_\nu(s)} = B_\nu[T(s)], \tag{B.7}$$

where B_ν is the Planck function for black body radiation. This enables us to rewrite the integral in equation (B.6) in terms of the optical depth,

$$I_\nu(s) = I_\nu(s_0)e^{-\tau_\nu} + e^{-\tau_\nu}\int_{0}^{\tau_\nu} B_\nu[T(\tau'_\nu)]e^{\tau'_\nu}d\tau'_\nu. \tag{B.8}$$

Finally, we assume that $B_\nu[T(\tau'_\nu)] \approx B_\nu(T_{\rm ex}) = I_{\rm ex}$ and define the following,

$$T'_{\rm b}(\nu) = \frac{c^2 I_\nu(s)}{2\nu^2 k_{\rm B}}, \tag{B.9}$$

and

$$T'_{\rm R}(\nu) = \frac{c^2 I_\nu(s_0)}{2\nu^2 k_{\rm B}}, \tag{B.10}$$

where equation (3.5) is then simply obtained from equation (B.8),

$$T'_{\rm b}(\nu) = T_{\rm ex}(1 - e^{-\tau_\nu}) + T'_{\rm R}(\nu)e^{-\tau_\nu}. \tag{B.11}$$

B.2 Spin temperature

Here we derive the relation between spin temperature and the various physical effects it is coupled to. This calculation was first done in Field (1958). Starting from equation (3.8), one obtains,

$$\frac{n_1}{n_0} = \frac{C_{01} + P_{01} + B_{01}I_{\rm CMB}}{C_{10} + P_{10} + A_{10} + B_{10}I_{\rm CMB}}. \tag{B.12}$$

Based on the assumption that $T_* \ll 1$, the definition of the spin temperature gives

$$\frac{n_1}{n_0} = \frac{g_1}{g_0}e^{-\frac{T_*}{T_{\rm S}}} \approx 3\left(1 - \frac{T_*}{T_{\rm S}}\right). \tag{B.13}$$

Then, for collisions the transition rate is proportional to the number of particles times some function of the kinetic temperature $T_{\rm K}$ of the particles. Since the ratio of rates is a

function of $T_{\rm K}$ only, the Boltzmann law for thermodynamic equilibrium holds,

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{-\frac{T_*}{T_{\rm K}}}. \tag{B.14}$$

In addition conservation from the steady state implies that,

$$n_0 C_{01} = n_1 C_{10} = n_0 \frac{g_1}{g_0} e^{-\frac{T_*}{T_{\rm K}}} C_{10}. \tag{B.15}$$

Thus, assuming $T_* \ll T_{\rm K}$,

$$\frac{C_{01}}{C_{10}} = \frac{g_1}{g_0} e^{-\frac{T_*}{T_{\rm K}}} \approx 3\left(1 - \frac{T_*}{T_{\rm K}}\right). \tag{B.16}$$

Furthermore, we define the effective colour temperature $T_{\rm c}$ via the following relation,

$$\frac{P_{01}}{P_{10}} \equiv 3\left(1 - \frac{T_*}{T_{\rm c}}\right). \tag{B.17}$$

Lastly, we look at the terms involving the Einstein B coefficients. Taking some background radiation at the 21cm frequency with specific intensity I_ν, the absorption and stimulated emission probabilities are given by the Einstein B rates times I_ν. In this context the only background radiation there is at this frequency comes from the CMB so we write $I_\nu = I_{\rm CMB}$. One can relate these quantities to the CMB temperature T_γ via the Rayleigh-Jeans limit, equation (3.3), and the relation between A and B (Griffiths 2005),

$$B_{10} I_{\rm CMB} = \frac{\pi^2 c^3}{\omega \hbar} A_{10} I_{\rm CMB} \tag{B.18a}$$

$$= \frac{\lambda^2 c}{2\nu h} A_{10} I_{CMB} \tag{B.18b}$$

$$= \frac{\lambda^2 c}{2\nu h} \frac{2\nu^2 k_{\rm B} T_\gamma}{c^2} A_{10} \tag{B.18c}$$

$$= \frac{c k_{\rm B}}{h\nu} T_\gamma A_{10}. \tag{B.18d}$$

Using the definition for T_* from equation (3.2) and setting $c = 1$, one then finds

$$B_{10} I_{\rm CMB} = A_{10} \frac{T_\gamma}{T_*}, \tag{B.19}$$

and

$$B_{01} I_{\rm CMB} = \frac{g_1}{g_0} B_{10} I_{\rm CMB} = 3 A_{10} \frac{T_\gamma}{T_*}. \tag{B.20}$$

Introducing these developments into equation (B.12),

$$3\left(1-e^{T_*/T_S}\right)=\frac{3\left(1-\frac{T_*}{T_K}\right)C_{10}+3\left(1-\frac{T_*}{T_c}\right)P_{10}+3A_{10}\frac{T_\gamma}{T_*}}{C_{10}+P_{10}+A_{10}+A_{10}\frac{T_\gamma}{T_*}}, \tag{B.21}$$

where we can rearrange terms to find

$$\frac{T_*}{T_S}=\frac{C_{10}+P_{10}+A_{10}\left(1+\frac{T_\gamma}{T_*}\right)-\left(1-\frac{T_*}{T_K}\right)C_{10}-\left(1-\frac{T_*}{T_c}\right)P_{10}-A_{10}\frac{T_\gamma}{T_*}}{C_{10}+P_{10}+A_{10}\left(1+\frac{T_\gamma}{T_*}\right)} \tag{B.22a}$$

$$=\frac{A_{10}+\frac{T_*}{T_K}C_{10}+\frac{T_*}{T_c}P_{10}}{C_{10}+P_{10}+A_{10}\left(1+\frac{T_\gamma}{T_*}\right)} \tag{B.22b}$$

$$=\frac{1+\frac{T_*}{T_K}\frac{C_{10}}{A_{10}}+\frac{T_*}{T_c}\frac{P_{10}}{A_{10}}}{\frac{C_{10}}{A_{10}}+\frac{P_{10}}{A_{10}}+1+\frac{T_\gamma}{T_*}}. \tag{B.22c}$$

At this point it is convenient to introduce the coupling coefficients x_α and x_c,

$$x_\alpha\equiv\frac{P_{10}}{A_{10}}\frac{T_*}{T_\gamma},\quad\text{and}\quad x_c\equiv\frac{C_{10}}{A_{10}}\frac{T_*}{T_\gamma}, \tag{B.23}$$

to obtain

$$\frac{T_*}{T_S}=\frac{1+x_c\frac{T_\gamma}{T_K}+x_\alpha\frac{T_\gamma}{T_c}}{x_c\frac{T_\gamma}{T_*}+x_\alpha\frac{T_\gamma}{T_*}+1+\frac{T_\gamma}{T_*}}. \tag{B.24}$$

Dividing through by T_* reveals

$$T_S^{-1}=\frac{T_\gamma^{-1}+x_cT_K^{-1}+x_\alpha T_c^{-1}}{1+x_c+x_\alpha+\frac{T_*}{T_\gamma}}. \tag{B.25}$$

Finally, under the assumption that $T_*\ll 1$, the fraction in the denominator can be ignored and the result is obtained,

$$T_S^{-1}=\frac{T_\gamma^{-1}+x_cT_K^{-1}+x_\alpha T_c^{-1}}{1+x_c+x_\alpha}. \tag{B.26}$$

B.3 Lyman-α coupling coefficient

Here we show that,

$$P_{10}=4P_\alpha/27, \tag{B.27}$$

and from there, it is easy to show why equation (3.15) holds.

We start by labelling the hyperfine transitions relevant to the Wouthuysen-Field effect in ascending order in energy, referring to Fig. 3.2, thus:

$$ {}_0S_{1/2} \equiv 0,\ {}_1S_{1/2} \equiv 1 $$

$$ {}_0P_{1/2} \equiv 2,\ {}_1P_{1/2} \equiv 3 $$

$$ {}_0P_{3/2} \equiv 4,\ {}_1P_{3/2} \equiv 5. $$

Now, we let A_{ji} be the spontaneous emission coefficient for transitions between these energy levels. Then, we write the background flux at the frequency corresponding to the $j \to i$ transition as J_{ji}. Deguchi & Watson (1985) write

$$ P_{10} = \sum_{j=2,3,4,5} \left(\frac{T^{\mathrm{B}}_{j1}}{T_{j1}} \right) \left(\frac{g_j A_{j1} J_{j1}}{g_1} \right) \frac{A_{j0}}{\sum\limits_{i=0,1} A_{ji}}, \tag{B.28} $$

$$ P_{01} = \sum_{j=2,3,4,5} \left(\frac{T^{\mathrm{B}}_{j0}}{T_{j0}} \right) \left(\frac{g_j A_{j0} J_{j0}}{g_0} \right) \frac{A_{j1}}{\sum\limits_{i=0,1} A_{ji}}. \tag{B.29} $$

In these equations, we defined the brightness temperature of the $j \to i$ transition, T^{B}_{ji}, via the Rayleigh-Jeans law, and the temperature corresponding to the energy difference between the two states, T_{ji}, as follows

$$ T^{\mathrm{B}}_{ji} = \frac{c^2}{2 k_{\mathrm{B}} \nu^2_{ji}} I_{ji} \quad \text{and} \quad T_{ji} = \frac{h \nu_{ji}}{k_{\mathrm{B}}}. \tag{B.30} $$

Further, g_k denotes the statistical weight of the level k and is obtained by

$$ g_k = 2F + 1, \tag{B.31} $$

and its value is related to k as shown by the table below.

k	0	1	2	3	4	5
g_k	1	3	1	3	3	5

Consider equation (B.28) to interpret this result. Looking at a single term in the sum over j, say $j = 3$, on obtains

$$ P_{10,j=3} = \frac{T^{\mathrm{B}}_{31}}{T_{31}} \frac{g_3 J_{31}}{g_1} A_{31} \frac{A_{30}}{A_{30} + A_{31}}. \tag{B.32} $$

The combination of the first three factors denotes the rate at which the atom excites from level 1 to level 3, and the last fraction is the probability for it to de-excite into the ground

state. Therefore, the de-excitation rate of the ground state triplet state due to Lyman-α scattering is composed of the various possibilities for the atom to excite away from the triplet state and de-excite down to the singlet state.

There are a few steps that can now be taken to simplify the above equations. First, we take the background flux to be equal for all transitions and equal to the average Lyman-α background flux, so for all i, j:

$$J_{ji} \approx J_\alpha.$$

Next, we ignore pairs that contain forbidden transitions, ie. $A_{20} = A_{50} = 0$. Note that this automatically eliminates pairs that do not mix the ground state levels, ie. pairs containing A_{51} and A_{21}. We can then rewrite equations (B.28) and (B.29) as follows,

$$P_{10} = J_\alpha \left(\frac{T_{31}^{\mathrm{B}}}{T_{31}} \frac{A_{31}A_{30}}{A_{31}+A_{30}} + \frac{T_{41}^{\mathrm{B}}}{T_{41}} \frac{A_{41}A_{40}}{A_{41}+A_{40}} \right), \tag{B.33}$$

$$P_{01} = 3J_\alpha \left(\frac{T_{30}^{\mathrm{B}}}{T_{30}} \frac{A_{31}A_{30}}{A_{31}+A_{30}} + \frac{T_{40}^{\mathrm{B}}}{T_{40}} \frac{A_{41}A_{40}}{A_{41}+A_{40}} \right). \tag{B.34}$$

We want to relate the A_{ji}'s to A_α, the total spontaneous emission rate of the Lyman-α transition due to all the hyperfine states. This can be achieved using the summation rule for transitions which states that the sum of the intensities of all transitions of given nFJ to all and any given $n'J'$, summed over F', is proportional to $2F+1$ (Meiksin 2000; Loeb & Furlanetto 2013).

$$\sum_{F'} I_{nFJ\to n'F'J'} \propto 2F+1. \tag{B.35}$$

So, we can write these sums down for the 4 different states that have downward transitions, and for the 4 different upward transitions:

- $nFJ = 22\frac{3}{2} : I_{22\frac{3}{2}\to 11\frac{1}{2}} = I_{51} \propto 5$
- $nFJ = 21\frac{3}{2} : I_{21\frac{3}{2}\to 11\frac{1}{2}} + I_{21\frac{3}{2}\to 10\frac{1}{2}} = I_{41} + I_{40} \propto 3$
- $nFJ = 21\frac{1}{2} : I_{21\frac{1}{2}\to 11\frac{1}{2}} + I_{21\frac{1}{2}\to 10\frac{1}{2}} = I_{31} + I_{30} \propto 3$
- $nFJ = 20\frac{1}{2} : I_{20\frac{1}{2}\to 11\frac{1}{2}} = I_{21} \propto 1$

From the downward transition we can write down the following ratios,

$$\frac{I_{51}}{I_{41}+I_{40}} = \frac{5}{3}, \quad \frac{I_{41}+I_{40}}{I_{31}+I_{30}} = 1, \quad \frac{I_{31}+I_{30}}{I_{21}} = 3. \tag{B.36}$$

- $nFJ = 10\frac{1}{2} \to n'J' = 2\frac{1}{2} : I_{10\frac{1}{2}\to 21\frac{1}{2}} = I_{03} = I_{30} \propto 1$
- $nFJ = 10\frac{1}{2} \to n'J' = 2\frac{3}{2} : I_{10\frac{1}{2}\to 21\frac{3}{2}} = I_{40} \propto 1$

- $nFJ = 11\frac{1}{2} \rightarrow n'J' = 2\frac{1}{2} : I_{11\frac{1}{2}\rightarrow 20\frac{1}{2}} + I_{11\frac{1}{2}\rightarrow 21\frac{1}{2}} = I_{21} + I_{31} \propto 3$
- $nFJ = 11\frac{1}{2} \rightarrow n'J' = 2\frac{3}{2} : I_{11\frac{1}{2}\rightarrow 21\frac{3}{2}} + I_{11\frac{1}{2}\rightarrow 22\frac{3}{2}} = I_{41} + I_{51} \propto 3$

From the upward transition we can write down the following ratios,

$$\frac{I_{40}}{I_{41} + I_{51}} = \frac{1}{3}, \quad \frac{I_{30}}{I_{21} + I_{31}} = \frac{1}{3}. \tag{B.37}$$

Having derived all the ratios above, we use

$$\frac{I_{ji}}{I_\alpha} = \frac{g_j A_{ji}}{g_{\rm tot} A_\alpha}, \tag{B.38}$$

where I_α is the total Lyman-α decay intensity summed over all the hyperfine transitions, $I_\alpha = \sum_{i,j} I_{ji}$, and $g_{\rm tot} = \sum_j g_j$, to obtain the ratios of A_{ji}/A_α. E.g. for A_{21},

$$I_{51} = \frac{5}{3}(I_{40} + I_{41}) \Rightarrow I_\alpha = I_{21} + I_{30} + I_{31} + \frac{8}{3}(I_{40} + I_{41}), \tag{B.39a}$$

$$I_{40} + I_{41} = I_{30} + I_{31} \Rightarrow I_\alpha = I_{21} + \frac{11}{3}(I_{30} + I_{31}), \tag{B.39b}$$

$$I_{30} + I_{31} = I_{21} \Rightarrow I_\alpha = 12 I_{21}, \tag{B.39c}$$

$$\therefore \frac{I_{21}}{I_\alpha} = \frac{1}{12} = \frac{g_2}{g_2 + g_3 + g_4 + g_5}\frac{A_{21}}{A_\alpha} = \frac{1}{12}\frac{A_{21}}{A_\alpha} \tag{B.39d}$$

$$\therefore \frac{A_{21}}{A_\alpha} = 1. \tag{B.39e}$$

Similarly, for the other transitions, one gets (Meiksin 2000)

$$\frac{A_{21}}{A_\alpha} = \frac{A_{51}}{A_\alpha} = 1, \tag{B.40}$$

$$\frac{A_{30}}{A_\alpha} = \frac{A_{41}}{A_\alpha} = \frac{1}{3}, \tag{B.41}$$

$$\frac{A_{31}}{A_\alpha} = \frac{A_{40}}{A_\alpha} = \frac{2}{3}. \tag{B.42}$$

We are now finally in a situation to relate P_{10} to P_α. From (B.33),

$$P_{10} = J_\alpha A_\alpha \left(\frac{T_{31}^{\rm B}}{T_{31}} \frac{\frac{2}{3\times 3}}{\frac{1}{3} + \frac{2}{3}} + \frac{T_{41}^{\rm B}}{T_{41}} \frac{\frac{2}{3\times 3}}{\frac{1}{3} + \frac{2}{3}} \right) = \frac{2}{9} J_\alpha A_\alpha \left(\frac{T_{31}^{\rm B}}{T_{31}} + \frac{T_{41}^{\rm B}}{T_{41}} \right). \tag{B.43}$$

Another approximation that can be made at this point is to say that $\frac{T_{31}^{\rm B}}{T_{31}} + \frac{T_{41}^{\rm B}}{T_{41}} \approx 2\frac{T_\alpha^{\rm B}}{T_\alpha}$, such that

$$P_{10} = \frac{4}{9} J_\alpha A_\alpha \frac{T_\alpha^{\rm B}}{T_\alpha}. \tag{B.44}$$

Writing down the total scattering rate of Lyman-α photons,

$$P_\alpha = \frac{T_\alpha^B}{T_\alpha} \sum_{i=0,1} \sum_{j=2,3,4,5} \frac{g_j A_{ji} J_{ji}}{g_0 + g_1} = \frac{T_\alpha^{\mathrm{B}}}{T_\alpha} \frac{J_\alpha}{4} \sum_{i=0,1} \sum_{j=2,3,4,5} g_j A_{ji} \tag{B.45}$$

$$\Rightarrow P_\alpha = \frac{T_\alpha^{\mathrm{B}}}{T_\alpha} \frac{J_\alpha}{4} (A_{21} + 3A_{30} + 3A_{31} + 3A_{40} + 3A_{31} + 5A_{51}). \tag{B.46}$$

Once again, one can apply the ratios found above,

$$P_\alpha = 3 J_\alpha A_\alpha \frac{T_\alpha^{\mathrm{B}}}{T_\alpha}, \tag{B.47}$$

and the final result can be obtained by dividing equation (B.44) by equation (B.47),

$$P_{10} = \frac{4}{27} P_\alpha. \tag{B.48}$$

Finally, (3.15) is obtained combining this result with the definition of x_α in equation (B.23).

Appendix C

Appendix for Chapter 5

C.1 Spherical harmonics

The spherical harmonics form a set of orthonormal basis functions, where the normalization is chosen such that the spherical harmonics $Y_{\ell m}$ adhere to the following orthonormality condition,

$$\int d^2\hat{\boldsymbol{n}} Y_{\ell m}(\hat{\boldsymbol{n}}) Y^*_{\ell' m'}(\hat{\boldsymbol{n}}) = \delta^{\mathrm{K}}_{\ell' \ell} \delta^{\mathrm{K}}_{m' m}, \tag{C.1}$$

where δ^K denotes the Kronecker delta.

C.2 Tree-level bispectrum due to NLG

For the non-linear gravity bispectrum, we look at perturbations in the density field up to second order as the linear perturbations are Gaussian, thus rendering the bispectrum zero,

$$\delta(\boldsymbol{x}, t) = \delta^{(1)}(\boldsymbol{x}, t) + \delta^{(2)}(\boldsymbol{x}, t), \tag{C.2}$$

where $\delta^{(1)}$ and $\delta^{(2)}$ denote the first and second order contributions to the density fluctuations. In the following, we suppress the explicit expression of the position and time dependence of the density fluctuations for notational ease. The bispectrum is the ensemble average of the Fourier transform of δ, such that in order to determine the bispectrum, we must first determine the Fourier transform of the first and second order contributions to the density perturbations, namely $\tilde{\delta}^{(1)}$ and $\tilde{\delta}^{(2)}$.

Let us first determine δ to first and second order. The second order equation of motion governing the evolution of the density perturbations in the matter dominated limit can be

shown to be (Fry 1984)

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - 4\pi G\bar{\rho}\delta = 4\pi G\bar{\rho}\delta^2 + \frac{1}{a^2}\nabla_i\delta\nabla_i\phi + \frac{1}{a^2}\nabla_i\nabla_j[(1+\delta)v_p^i v_p^j], \tag{C.3}$$

where the peculiar velocity, v_p^i, relative to the Hubble flow and the gravitational potential, ϕ, are related to δ via the continuity equation (2.44) and the Poisson equation (2.50) respectively.

The first order solution is determined by the linear perturbation theory result of chapter 2, and we have that

$$\delta^{(1)}(\boldsymbol{x}, t) = \delta_0(\boldsymbol{x})A(t), \tag{C.4}$$

where δ_0 denotes the initial perturbation and $A(t) \propto t^{2/3}$ is the growing solution to the evolution equation. The solutions for the first order peculiar velocity and potential can be found by solving the continuity and Poisson equations to first order, defining a function Δ as $\nabla^2\Delta = \delta$, such that

$$\Delta(\boldsymbol{x}) = -\frac{1}{4\pi}\int d^3\boldsymbol{x}'\frac{\delta(\boldsymbol{x}')}{|\boldsymbol{x}-\boldsymbol{x}'|}. \tag{C.5}$$

One thus finds

$$v_p^{(1),i} = -a\left(\frac{\dot{A}}{A}\right)\nabla^i\Delta^{(1)}, \tag{C.6}$$

and

$$\phi^{(1)} = 4\pi G\bar{\rho}a^2\Delta^{(1)}. \tag{C.7}$$

The second order solution for the density perturbation can then be found by solving equation (C.3), ie solving

$$\ddot{\delta}^{(2)} + 2\frac{\dot{a}}{a}\dot{\delta}^{(2)} - 4\pi G\bar{\rho}\delta^{(2)} = 4\pi G\bar{\rho}\delta^{(1),2} + \frac{1}{a^2}\nabla_i\delta^{(1)}\nabla_i\phi^{(1)} + \frac{1}{a^2}\nabla_i\nabla_j[(1+\delta)v_p^{(1),i}v_p^{(1),j}]. \tag{C.8}$$

Substituting the solutions for the peculiar velocity and the gravitational potential, the equation can be expressed in terms of the first order perturbation only,

$$\ddot{\delta}^{(2)}+2\frac{\dot{a}}{a}\dot{\delta}^{(2)}-4\pi G\bar{\rho}\delta^{(2)} = \left[4\pi G\bar{\rho} + \left(\frac{\dot{A}}{A}\right)^2\right]\delta^{(1),2}+\left[4\pi G\bar{\rho} + 2\left(\frac{\dot{A}}{A}\right)^2\right]\nabla_i\delta^{(1)}\nabla_i\Delta^{(1)}+\left(\frac{\dot{A}}{A}\right)^2\nabla_i\nabla_j\Delta^{(1),2}. \tag{C.9}$$

Next, we use the linear solution to express

$$\frac{\dot{A}}{A} = \frac{2}{3}t^{-1} = \frac{\dot{a}}{a},\ \frac{\ddot{A}}{A} = -\frac{2}{9}t^{-2},\ \text{and}\ 4\pi G\bar{\rho} = \frac{2}{3}t^{-2}. \tag{C.10}$$

Equation (C.9) can then be solved using the Ansatz, $\delta^{(2)} \propto A^2(t)$, and expressing both sides as the same power of t. We finally find

$$\delta^{(2)} = \frac{5}{7}\delta^{(1),2} + \delta^{(1)}_{,i}\Delta^{(1)}_{,i} + \frac{2}{7}\Delta^{(1),2}_{,i}. \tag{C.11}$$

We now have an expression for the first and second order perturbations, and are able to find their Fourier transform, defined by

$$\tilde{\delta}(\boldsymbol{k}) = \int d^3\boldsymbol{x}\delta(\boldsymbol{x})e^{i\boldsymbol{k}\cdot\boldsymbol{x}}, \text{ and } \delta(\boldsymbol{x}) = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}\tilde{\delta}(\boldsymbol{k})e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \tag{C.12}$$

Whereas the first order perturbation is straight forwardly found to be

$$\tilde{\delta}^{(1)}(\boldsymbol{k}, z) = D_+(z)\tilde{\delta}_0(\boldsymbol{k}), \tag{C.13}$$

where we have expressed the time component in redshift space via the growth function D_+, the second order perturbation involves the multiplication of differing factors, which result in a convolution in Fourier space, as

$$FT(f_1 f_2)(\boldsymbol{k}) = \int \frac{d^3\boldsymbol{k}_1}{(2\pi)^3}\frac{d^3\boldsymbol{k}_2}{(2\pi)^3}\left[(2\pi)^3\delta^{\mathrm{D}}(\boldsymbol{k}_1 + \boldsymbol{k}_2 - \boldsymbol{k})\right] FT(f_1)(\boldsymbol{k}_1)FT(f_2)(\boldsymbol{k}_2). \tag{C.14}$$

In order to transform the second order perturbation, we first find the transform of each factor in equation (C.11) and find,

$$\delta^{(1)}(\boldsymbol{x}) \to \tilde{\delta}^{(1)}(\boldsymbol{k}), \tag{C.15a}$$

$$\delta^{(1)}_{,i}(\boldsymbol{x}) \to ik^i\tilde{\delta}^{(1)}(\boldsymbol{k}), \tag{C.15b}$$

$$\Delta^{(1)}(\boldsymbol{x}) \to -\frac{1}{k^2}\tilde{\delta}^{(1)}(\boldsymbol{k}), \tag{C.15c}$$

$$\Delta^{(1)}(\boldsymbol{x})_{,i} \to -\frac{ik^i}{k^2}\tilde{\delta}^{(1)}(\boldsymbol{k}), \tag{C.15d}$$

$$\Delta^{(1)}(\boldsymbol{x})_{,ij} \to \frac{k^i k^j}{k^2}\tilde{\delta}^{(1)}(\boldsymbol{k}). \tag{C.15e}$$

$$\tag{C.15f}$$

Using these transforms, we can then use equation (C.14) to find

$$\tilde{\delta}^{(2)}(\boldsymbol{k}) = \int \frac{d^3\boldsymbol{k}_1 d^3\boldsymbol{k}_2}{(2\pi)^3}\delta^{\mathrm{D}}(\boldsymbol{k}_1+\boldsymbol{k}_2-\boldsymbol{k})\tilde{\delta}^{(1)}(\boldsymbol{k}_1)\tilde{\delta}^{(1)}(\boldsymbol{k}_2)\left[\frac{5}{7} + \frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_2^2} + \frac{2}{7}\left(\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1 k_2}\right)^2\right]. \tag{C.16}$$

The order of terms in the second can be interchanged and one therefore writes

$$FT(\delta^{(1)}_{,i}\Delta^{(1)}_{,i}) = \frac{1}{2}FT(\delta^{(1)}_{,i}\Delta^{(1)}_{,i} + \Delta^{(1)}_{,i}\delta^{(1)}_{,i}), \tag{C.17}$$

which then results in the following form of the transform of the second order density perturbation,

$$\tilde{\delta}^{(2)}(\boldsymbol{k}) = \int \frac{d^3\boldsymbol{k}_1 d^3\boldsymbol{k}_2}{(2\pi)^3}\delta^{\mathrm{D}}(\boldsymbol{k}_1+\boldsymbol{k}_2-\boldsymbol{k})\tilde{\delta}^{(1)}(\boldsymbol{k}_1)\tilde{\delta}^{(1)}(\boldsymbol{k}_2)\left[\frac{5}{7} + \frac{1}{2}\left(\frac{\boldsymbol{k}_1}{\boldsymbol{k}_2} + \frac{\boldsymbol{k}_2}{\boldsymbol{k}_1}\right)\cos\theta_{12} + \frac{2}{7}\cos^2\theta_{12}\right], \tag{C.18}$$

where we have defined the angle between $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$ as θ_{12}. The term in brackets is defined as $\mathcal{K}(\boldsymbol{k}_1, \boldsymbol{k}_2)$ by equation (5.22) in section 5.4.1. The bispectrum of the density perturbations is thus defined as

$$\begin{aligned}(2\pi)^3 B^{\mathrm{NLG}}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)\delta^{\mathrm{D}}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3) &= \left\langle \tilde{\delta}(\boldsymbol{k}_1)\tilde{\delta}(\boldsymbol{k}_2)\tilde{\delta}(\boldsymbol{k}_3)\right\rangle \\ &= \left\langle \left(\tilde{\delta}^{(1)}_1 + \tilde{\delta}^{(2)}_1\right)\left(\tilde{\delta}^{(1)}_2 + \tilde{\delta}^{(2)}_2\right)\left(\tilde{\delta}^{(1)}_3 + \tilde{\delta}^{(2)}_3\right)\right\rangle,\end{aligned} \tag{C.19}$$

where we have introduced the short-hand notation $\tilde{\delta}_i \equiv \tilde{\delta}(\boldsymbol{k}_i)$ for $i = 1, 2, 3$. When multiplying our the right hand side of this expression, we find that the lowest order contribution is the linear bispectrum which is zero as the field is Gaussian at this order,

$$\left\langle \tilde{\delta}^{(1)}_1\tilde{\delta}^{(1)}_2\tilde{\delta}^{(1)}_3\right\rangle = 0, \tag{C.20}$$

such that the lowest non-zero contribution to the bispectrum as terms involving a single second order term,

$$B^{\mathrm{NLG}}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) \sim \left\langle \tilde{\delta}^{(1)}_1\tilde{\delta}^{(1)}_2\tilde{\delta}^{(2)}_3\right\rangle + \text{cyclic terms}. \tag{C.21}$$

Now we can apply our result for the second order transform to get

$$\left\langle \tilde{\delta}^{(1)}_1\tilde{\delta}^{(1)}_2\tilde{\delta}^{(2)}_3\right\rangle = \int \frac{d^3\boldsymbol{k}_a d^3\boldsymbol{k}_b}{(2\pi)^3}\delta^{\mathrm{D}}(\boldsymbol{k}_a + \boldsymbol{k}_b - \boldsymbol{k}_3)\mathcal{K}(\boldsymbol{k}_a, \boldsymbol{k}_b)\left\langle \tilde{\delta}^{(1)}_1\tilde{\delta}^{(1)}_2\tilde{\delta}^{(1)}_a\tilde{\delta}^{(1)}_b\right\rangle, \tag{C.22}$$

and the angular bracket can be expanded using Wick's theorem,

$$\left\langle \tilde{\delta}^{(1)}_1\tilde{\delta}^{(1)}_2\tilde{\delta}^{(1)}_a\tilde{\delta}^{(1)}_b\right\rangle = \left\langle \tilde{\delta}^{(1)}_1\tilde{\delta}^{(1)}_2\right\rangle\left\langle \tilde{\delta}^{(1)}_a\tilde{\delta}^{(1)}_b\right\rangle + \left\langle \tilde{\delta}^{(1)}_1\tilde{\delta}^{(1)}_a\right\rangle\left\langle \tilde{\delta}^{(1)}_2\tilde{\delta}^{(1)}_b\right\rangle + \left\langle \tilde{\delta}^{(1)}_1\tilde{\delta}^{(1)}_b\right\rangle\left\langle \tilde{\delta}^{(1)}_2\tilde{\delta}^{(1)}_a\right\rangle. \tag{C.23}$$

Finally, using the definition of the power spectrum,

$$\left\langle \tilde{\delta}_1\tilde{\delta}_2\right\rangle = (2\pi)^3 P(k_1)\delta^{\mathrm{D}}(\boldsymbol{k}_1 + \boldsymbol{k}_2), \tag{C.24}$$

and combining the results above as well as reintroducing the explicit temporal dependence, one finds the tree-level bispectrum due to the non-linear collapse of structure to be

$$B^{\rm NLG}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, z_1, z_2, z_3) = 2\mathcal{K}(\boldsymbol{k}_1, \boldsymbol{k}_2) D_+^2(z_1) D_+(z_2) D_+(z_3) P(k_1) P(k_2) + \text{cycl.}. \quad \text{(C.25)}$$

Equation (5.21) is then realised by using the fact that the brightness temperature fluctuations are related to the density fluctuations through a bias factor.

C.3 Angular bispectrum due to non-linear gravity.

For the derivation of the angular bispectrum due to non-linear gravitational collapse, we expand the brightness temperature fluctuations in spherical harmonic space and define the angular coefficient as per equation (5.13),

$$a_{\ell m}^{\nu} = \int d^2\hat{\boldsymbol{n}} \delta T_{\rm b}^{\rm obs}(\hat{\boldsymbol{n}}, \nu) Y_{\ell m}(\hat{\boldsymbol{n}}), \quad \text{(C.26)}$$

with the relevant notation defined in section 5.3. The angular bispectrum is then

$$B_{\ell_1\ell_2\ell_3}^{{\rm NLG},m_1m_2m_3} = \left\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \right\rangle, \quad \text{(C.27)}$$

where we have suppressed the explicit frequency dependence for notational ease. By expanding the harmonic coefficients as well as introducing the window function defined in equation (5.7), we can write the angular bispectrum in terms of the three-point correlation function of brightness temperature fluctuations,

$$B_{\ell_1\ell_2\ell_3}^{{\rm NLG},m_1m_2m_3} = \int d^3\boldsymbol{r}_1 d^3\boldsymbol{r}_2 d^3\boldsymbol{r}_3 W_1 W_2 W_3 Y_{\ell_1 m_1}(\hat{\boldsymbol{n}}_1) Y_{\ell_2 m_2}(\hat{\boldsymbol{n}}_2) Y_{\ell_3 m_3}(\hat{\boldsymbol{n}}_3) \left\langle \delta T_{\rm b,1} \delta T_{\rm b,2} \delta T_{\rm b,3} \right\rangle, \quad \text{(C.28)}$$

where $W_i \equiv W(\boldsymbol{r}_i)$ and $\delta T_{\rm b,i} \equiv \delta T_{\rm b}(\boldsymbol{r}_i)$ for $i = 1, 2, 3$. We can introduce the tree-level bispectrum into this expression by introducing the Fourier transform of the three-point function. Using equation (5.20) to introduce the power spectrum and expanding the Fourier modes in spherical harmonics, the bispectrum can be expressed via

$$B_{\ell_1\ell_2\ell_3}^{{\rm NLG},m_1m_2m_3}(z) = B_{\ell_1\ell_2\ell_3}(z) \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \quad \text{(C.29)}$$

with the cyclic nature of the tree-level bispectrum resulting in the form of equation (5.24).

C.4 Limber approximation

In the Limber approximation, for large ℓ, Bessel functions are taken to be sharply peaked and are approximated by a Dirac delta function (Loverde & Afshordi 2008), such that

$$\begin{aligned}\int dk k^2 f(k) j_\ell[kr(z)] j_\ell[kq(z')] \\ \simeq f\left[\frac{\ell+1/2}{r(z)}\right]\frac{\pi}{2r^2(z)}\frac{\delta^{\rm D}(z-z')}{|r'(z)|}.\end{aligned} \tag{C.30}$$

In order to integrate equation (5.25), we use the Limber approximation and look at each ℓ term in turn.

C.4.1 The $\ell = 0$ case

We begin with the $\ell = 0$ term in (5.25), which can be written as

$$B_{12}^{\ell=0} = A_{\ell_1\ell_2\ell_3} b_{\rm HI} \int dz D_+^2(z)\delta\bar{T}_{\rm b}(z) W_\nu(z)\theta_{\ell_1}(z)\theta_{\ell_2}(z), \tag{C.31}$$

where we have defined

$$\begin{aligned}A_{\ell_1\ell_2\ell_3} =& \frac{16}{\pi}\sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{(4\pi)^3}}(2\ell_1+1) \\ & (2\ell_2+1)B_0 \begin{Bmatrix}\ell_1 & \ell_2 & \ell_3\\ \ell_2 & \ell_1 & 0\end{Bmatrix}\begin{pmatrix}\ell_1 & \ell_1 & 0\\ 0 & 0 & 0\end{pmatrix} \\ & \begin{pmatrix}\ell_2 & \ell_2 & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\ell_3 & \ell_1 & \ell_2\\ 0 & 0 & 0\end{pmatrix},\end{aligned} \tag{C.32}$$

and

$$\theta_\ell(z) = b_{\rm HI}\int dz' D_+(z')\delta\bar{T}_{\rm b}(z') W_\nu(z')\beta_\ell(z,z'), \tag{C.33}$$

with

$$\beta_\ell(z,z') = \int dk k^2 P(k) j_\ell[kr(z)] j_\ell[kr(z')]. \tag{C.34}$$

Applying the Limber approximation (C.30) to β_ℓ gives

$$\beta_\ell(z,z') \simeq \frac{\pi}{2r^2(z)r'(z)} P\left[\frac{\ell+1/2}{r(z)}\right]\delta^{\rm D}(z-z'), \tag{C.35}$$

such that

$$\theta_\ell(z) \simeq \frac{\pi b_{\rm HI}}{2r^2(z)r'(z)} P\left[\frac{\ell+1/2}{r(z)}\right] D_+(z)\delta\bar{T}_{\rm b}(z) W_\nu(z). \tag{C.36}$$

C.4.2 The $\ell = 1$ case

For the $\ell = 1$ case, we have that $\beta_1(k_1, k_2) = 2A_1\,(k_1/k_2 + k_2/k_1)$. Therefore, (5.25) contains two terms with k integrals of the form, $\int dk_1 dk_2 k_1^3 k_2 \cdots$ and $\int dk_1 dk_2 k_1 k_2^3 \cdots$. Defining functions similar to the $\ell = 0$ case, we find

$$\begin{aligned} B_{12}^{\ell=1} =& b_{\rm HI} \sum_{\ell'\ell''} A^{\ell'\ell''}_{\ell_1\ell_2\ell_3} \int dz W_\nu(z) T_{\rm b}(z) D_+^2(z) \\ & \left[\theta^1_{\ell_1\ell'}(z)\theta^{-1}_{\ell_2\ell''}(z) + \theta^{-1}_{\ell_1\ell'}(z)\theta^1_{\ell_2\ell''}(z)\right], \end{aligned} \tag{C.37}$$

where we define

$$\begin{aligned} A^{\ell'\ell''}_{\ell_1\ell_2\ell_3} =& -\frac{16}{\pi}\sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{(4\pi)^3}}(2\ell'+1) \\ & (2\ell''+1) i^{\ell_1+\ell_2+\ell'+\ell''} 2A_1 \begin{Bmatrix} \ell_1 & \ell_2 & \ell_3 \\ \ell'' & \ell' & 1 \end{Bmatrix} \\ & \begin{pmatrix} \ell_1 & \ell' & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_2 & \ell'' & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_3 & \ell' & \ell'' \\ 0 & 0 & 0 \end{pmatrix}, \end{aligned} \tag{C.38}$$

and

$$\theta^q_{\ell\ell'}(z) = b_{\rm HI} \int dz' D_+(z') \delta\bar{T}_{\rm b}(z') W_\nu(z') \beta^q_{\ell\ell'}(z, z'), \tag{C.39}$$

with

$$\beta^q_{\ell\ell'}(z, z') = \int dk k^{2+q} P(k) j_\ell[kr(z)] j_{\ell'}[kr(z')]. \tag{C.40}$$

Importantly, the Wigner symbols in (C.38) reduce the sum in (C.37) to 4 terms, which all incidentally render the powers of i even. Only terms with $\ell' = \ell_1 - 1, \ell_1 + 1$ and $\ell'' = \ell_2 - 1, \ell_2 + 1$, as shown in the table below are non-zero.

ℓ'/ℓ''	$\ell_2 - 1$	$\ell_2 + 1$
$\ell_1 - 1$	$\ell_1 - 1, \ell_2 - 1$	$\ell_1 - 1, \ell_2 + 1$
$\ell_1 + 1$	$\ell_1 + 1, \ell_2 - 1$	$\ell_1 + 1, \ell_2 + 1$

The difference between the ℓ indices in (C.37) is one, such that we approximate $\ell \pm 1 \sim \ell$ for Bessel function indices here. We find this approximation to work well as most of the signal comes from large ℓ-modes. We thus apply the Limber approximation (C.30) with $f(k) = k^q P(k)$ to (C.40), and find

$$\beta^q_{\ell\ell'}(z, z') \simeq \frac{\pi(\ell + 1/2)^q}{2r^{2+q}(z) r'(z)} P\left[\frac{\ell + 1/2}{r(z)}\right] \delta^{\rm D}(z - z'), \tag{C.41}$$

such that

$$\theta^q_{\ell\ell'}(z) \simeq \frac{\pi b_{\rm HI}(\ell+1/2)^q}{2r^{2+q}(z)r'(z)} P\left[\frac{\ell+1/2}{r(z)}\right] D_+(z)\delta\bar{T}_{\rm b}(z)W_\nu(z). \tag{C.42}$$

C.4.3 The $\ell = 2$ case

Similar to the $\ell = 0$ case, B_2 is independent of k, and thus we can write

$$\begin{aligned} B_{12}^{\ell=2} =& b_{\rm HI}\sum_{\ell'\ell''} A^{\ell'\ell''}_{\ell_1\ell_2\ell_3}\int dz W_\nu(z)T_{\rm b}(z)D_+^2(z) \\ & \theta_{\ell_1\ell'}(z)\theta_{\ell_2\ell''}(z), \end{aligned} \tag{C.43}$$

where we define

$$\begin{aligned} A^{\ell'\ell''}_{\ell_1\ell_2\ell_3} =& \frac{16}{\pi}\sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{(4\pi)^3}}(2\ell'+1) \\ & (2\ell''+1)\beta_2 i^{\ell_1+\ell_2+\ell'+\ell''}\begin{Bmatrix}\ell_1 & \ell_2 & \ell_3\\ \ell'' & \ell' & 2\end{Bmatrix} \\ & \begin{pmatrix}\ell_1 & \ell' & 2\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\ell_2 & \ell'' & 2\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\ell_3 & \ell' & \ell''\\ 0 & 0 & 0\end{pmatrix}, \end{aligned} \tag{C.44}$$

and

$$\theta_{\ell\ell'}(z) = b_{\rm HI}\int dz' D_+(z')\delta\bar{T}_{\rm b}(z')W(z')\beta_{\ell\ell'}(z,z'), \tag{C.45}$$

with

$$\beta_{\ell\ell'}(z,z') = \int dk k^2 P(k) j_\ell[kr(z)] j_{\ell'}[kr(z')]. \tag{C.46}$$

Similar to the $\ell = 1$ case, the Wigner symbols in (C.44) reduce the sum in (C.43) to 9 non-zero terms, which all result in even powers of i. The terms are non-zero for combinations of $\ell' = \ell_1 - 2, \ell_1, \ell_1 + 2$ and $\ell'' = \ell_2 - 2, \ell_2, \ell_2 + 2$ as shown below.

ℓ'/ℓ''	ℓ_2-2	ℓ_2	ℓ_2+2
ℓ_1-2	ℓ_1-2, ℓ_2-2	ℓ_1-2, ℓ_2	ℓ_1-2, ℓ_2+2
ℓ_1	ℓ_1, ℓ_2-2	ℓ_1, ℓ_2	ℓ_1, ℓ_2+2
ℓ_1+2	ℓ_1+2, ℓ_2-2	ℓ_1+2, ℓ_2	ℓ_1-1, ℓ_2+1

Although at large ℓ, we have $\ell \pm 2 \sim \ell$, we find that this approximation does not give robust results when applying the Limber approximation. Instead we assume that $P(k)$ varies slowly across the range of the peaks of both Bessel functions such that we can effectively evaluate it at either peak. Similarly, we assume that $D_+\delta\bar{T}_{\rm b}$ varies slowly across the window, such that we may evaluate it at the window centre. Hence, for combinations

involving $\ell_1 \pm 2$ and $\ell_2 \pm 2$, we have

$$\begin{aligned}\theta_{\ell\ell\pm2}(z) \simeq &\frac{\pi b_{\rm HI}}{2r^2(z)|r'(z)|} P\left[\frac{\ell+1/2}{r(z)}\right] D_+(z)\delta\bar{T}_{\rm b}(z) \\ &\int dz' dk k^2 j_\ell[kr(z)] j_{\ell\pm2}[kr(z')] W_\nu(z') \\ &\times \frac{2r^2(z')|r'(z')|}{\pi}.\end{aligned} \tag{C.47}$$

We need to include the factor of $2r^2(z')|r'(z')|/\pi$ into the integral, as evaluating $P(k)$ at the peak of the Bessel function introduces the inverse term when setting the k-integral to a delta function, and since we are evaluating the integral exactly here, we need to cancel out this normalization. When $\ell' = \ell_1$ and $\ell'' = \ell_2$, we apply (C.30) similarly to the $\ell = 0$ case, and recover (C.35) and (C.36).

C.5 Lensing coefficient derivation

The brightness temperature fluctuations projected onto the sky are perturbed along the line of sight by the ISW effect, and in angle by gravitational lensing,

$$\begin{aligned}\delta T_{\rm b}^{\rm obs}(\hat{\boldsymbol{n}}, \nu) =&\delta T_{\rm b,0}^{\rm obs}(\hat{\boldsymbol{n}}, \nu) + \nabla\delta T_{\rm b,0}^{\rm obs}(\hat{\boldsymbol{n}}, \nu)\cdot\nabla\theta(\hat{\boldsymbol{n}}, \nu) \\ &+ \nu\frac{d\delta T_{\rm b,0}^{\rm obs}}{d\nu}(\hat{\boldsymbol{n}}, \nu)\frac{\Delta\nu}{\nu}(\hat{\boldsymbol{n}}, \nu),\end{aligned} \tag{C.48}$$

where the 0-index indicates the unperturbed field. These fluctuations can then be transformed into harmonic space,

$$\begin{aligned}a_{\ell m}^{\nu} = \int d^2\hat{\boldsymbol{n}} Y_{\ell m}(\hat{\boldsymbol{n}}) \Big[&\delta T_{\rm b,0}^{\rm obs}(\hat{\boldsymbol{n}}, \nu) + \nabla\delta T_{\rm b,0}^{\rm obs}(\hat{\boldsymbol{n}}, \nu)\cdot\nabla\theta(\hat{\boldsymbol{n}}, \nu) \\ &+\nu\frac{d\delta T_{\rm b,0}^{\rm obs}}{d\nu}(\hat{\boldsymbol{n}}, \nu)\frac{\Delta\nu}{\nu}(\hat{\boldsymbol{n}}, \nu)\Big].\end{aligned} \tag{C.49}$$

We can separate out each term in equation (C.49). Then, according to eq (32), we define

$$a_{\ell m}^{{\rm L},\nu} = \int d^2\hat{\boldsymbol{n}} Y_{\ell m}(\hat{\boldsymbol{n}})\nabla\delta T_{\rm b,0}^{\rm obs}(\hat{\boldsymbol{n}}, \nu)\cdot\nabla\theta(\hat{\boldsymbol{n}}, \nu) \tag{C.50}$$

with

$$\theta(\hat{\boldsymbol{n}}, \nu) = \sum_{\ell' m'} \theta_{\ell' m'}^{\nu} Y_{\ell' m'}^*(\hat{\boldsymbol{n}}), \tag{C.51}$$

and

$$\delta T_{\rm b,0}^{\rm obs}(\hat{\boldsymbol{n}},\nu) = \sum_{\ell' m'} a_{\ell' m'}^{0,\nu} Y_{\ell' m'}^{*}(\hat{\boldsymbol{n}}). \tag{C.52}$$

We thus find

$$\begin{aligned} a_{\ell m}^{\rm L,\nu} = & \sum_{\ell'\ell'' m' m''} \int d^2\hat{\boldsymbol{n}} a_{\ell' m'}^{*0,\nu} \theta_{\ell'' m''}^{*\nu} \\ & \times Y_{\ell m}^{*}(\hat{\boldsymbol{n}}) \nabla Y_{\ell' m'}^{*}(\hat{\boldsymbol{n}}) \cdot \nabla Y_{\ell'' m''}^{*}(\hat{\boldsymbol{n}}), \end{aligned} \tag{C.53}$$

Where we have used the fact that the fluctuations are real. Further, one can use the properties of the spherical harmonics and the following identity for functions A, B, and C,

$$\int d\hat{\boldsymbol{n}} C \nabla A \cdot \nabla B = \frac{1}{2} \int d\hat{\boldsymbol{n}} \left(AB\nabla^2 C - AC\nabla^2 B - BC\nabla^2 A \right), \tag{C.54}$$

to show that the angular integral becomes,

$$\int d\hat{\boldsymbol{n}} Y_{\ell m}^{*}(\hat{\boldsymbol{n}}) \nabla Y_{\ell' m'}^{*}(\hat{\boldsymbol{n}}) \cdot \nabla Y_{\ell'' m''}^{*}(\hat{\boldsymbol{n}}) = W_{\ell\ell'\ell''}^{m m' m''}, \tag{C.55}$$

where

$$W_{\ell\ell'\ell''}^{m m' m''} \equiv \frac{1}{2} (-1)^{m+m'+m''} L_{\ell\ell'\ell''} \mathcal{H}_{\ell\ell'\ell''}^{m m' m''}, \tag{C.56}$$

with

$$L_{\ell\ell'\ell''} \equiv -\ell(\ell+1) + \ell'(\ell'+1) + \ell''(\ell''+1). \tag{C.57}$$

Therefore, we find

$$a_{\ell m}^{\rm L,\nu} = \sum_{\ell'\ell'' m' m''} W_{\ell\ell'\ell''}^{m m' m''} a_{\ell' m'}^{*0,\nu} \theta_{\ell'' m''}^{*\nu}. \tag{C.58}$$

The harmonic transforms can be related to the 3D fields via

$$a_{\ell m}^{\nu} = \int d^2\hat{\boldsymbol{n}} \delta T_{\rm b}^{\rm obs}(\hat{\boldsymbol{n}},\nu) Y_{\ell m}(\hat{\boldsymbol{n}}), \tag{C.59}$$

where the orthonormality condition for spherical harmonics can be applied to obtain

$$\begin{aligned} a_{\ell m}^{\nu} = & 4\pi i^{\ell} \int dz W_{\nu}(z) \delta\bar{T}_{\rm b}(z) b_{\rm HI}(z) D_{+}(z) \\ & \times \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} \tilde{\delta}(\boldsymbol{k}) j_{\ell}[kr(z)] Y_{\ell m}(\hat{\boldsymbol{k}}). \end{aligned} \tag{C.60}$$

Here

$$\begin{aligned} \theta_{\ell m}^{\nu} = & \int d^2\hat{\boldsymbol{n}} \theta(\hat{\boldsymbol{n}},\nu) Y_{\ell m}(\hat{\boldsymbol{n}}) \\ = & \int d^2\hat{\boldsymbol{n}} dz W_{\nu}(z) \theta\left[r(z)\hat{\boldsymbol{n}}, z\right] Y_{\ell m}(\hat{\boldsymbol{n}}), \end{aligned} \tag{C.61}$$

with $\theta\left[r(z)\hat{\boldsymbol{n}}, z\right]$ given by equation (5.28),

$$\begin{aligned}\theta^{\nu}_{\ell m} = -\frac{2}{c^2}\int d^2\hat{\boldsymbol{n}}dz W_\nu(z)Y_{\ell m}(\hat{\boldsymbol{n}})\times \\ \int_0^{r(z)} dr' \frac{S_k\left[r(z)-r'\right]}{S_k\left[r(z)\right]S_k(r')}\Phi(r'\hat{\boldsymbol{n}}).\end{aligned} \tag{C.62}$$

C.6 LISW power spectrum

Let us first write down an expression for the ISW coefficients. From (C.49),

$$a^{\mathrm{ISW},\nu}_{\ell m} = \int d^2\hat{\boldsymbol{n}}Y^*_{\ell m}(\hat{\boldsymbol{n}})\underbrace{\nu\frac{d\delta T^{\mathrm{obs}}_{\mathrm{b},0}}{d\nu}(\hat{\boldsymbol{n}})}_{\eta^\nu(\hat{\boldsymbol{n}})}\frac{\Delta\nu}{\nu}(\hat{\boldsymbol{n}},\nu) \tag{C.63}$$

We relate the projection on the sky to the 3D field,

$$\begin{aligned}a^{\mathrm{ISW},\nu}_{\ell m} &= \int d^2\hat{\boldsymbol{n}}Y^*_{\ell m}(\hat{\boldsymbol{n}})\eta^\nu(\hat{\boldsymbol{n}})\int dz W_\nu(z)\frac{\Delta\nu}{\nu}\left[r(z)\hat{\boldsymbol{n}},z\right] \\ &= \frac{2}{c^3}\int d^2\hat{\boldsymbol{n}}dz Y^*_{\ell m}(\hat{\boldsymbol{n}})\eta^\nu(\hat{\boldsymbol{n}})W_\nu(z) \\ &\times\int_0^{r(z)} dr'\frac{\partial\Phi}{\partial t}(r'\hat{\boldsymbol{n}},z),\end{aligned} \tag{C.64}$$

where we assume $\eta^\nu(\hat{\boldsymbol{n}}) = \eta(z) = \nu(z)\frac{d\bar{T}_\mathrm{b}}{d\nu}(z)$ to lowest order.

We then define $Q_\ell(\nu,\nu)$ via equation (5.39). Applying the Kronecker deltas, we find

$$\begin{aligned}Q_\ell(\nu,\nu) = \Bigg\langle -\frac{2}{c^2}\int d\hat{\boldsymbol{n}}dz W_\nu(z)Y_{\ell m}(\hat{\boldsymbol{n}}) \\ \times\int_0^{r(z)} dr'\frac{S_k[r(z)-r']}{S_k[r(z)]S_k(r')} \\ \times\frac{2}{c^3}\int d\hat{\boldsymbol{n}}'dz' W_\nu(z')Y_{\ell m}(\hat{\boldsymbol{n}}')\eta(z') \\ \times\int_0^{r(z')} dr''\frac{\partial\Phi}{\partial t}(r''\hat{\boldsymbol{n}}',z')\Bigg\rangle.\end{aligned} \tag{C.65}$$

We then write Φ in terms of its Fourier transform and expand the exponential according to equation (5.11). The resulting expression can be summed over using the spherical harmonics closure relations and through the definition of the power spectrum for the gravitational potential,

$$\left\langle\frac{\partial\Phi}{\partial t}(\boldsymbol{k},z)\Phi(\boldsymbol{k}',z')\right\rangle = \frac{(2\pi)^3}{2}\frac{\partial P_\Phi}{\partial t}(k,z,z')\delta^D(\boldsymbol{k}+\boldsymbol{k}'), \tag{C.66}$$

we find

$$\begin{aligned} Q_\ell(\nu,\nu) = \frac{2(4\pi)^2}{c^5}\int dz dz' W_\nu(z) W_\nu(z')\eta(z') \\ \times \int_0^{r(z)} dr' \int_0^{r(z')} dr'' \frac{S_k[r(z)-r']}{S_k[r(z)]S_k(r')} \\ \times \int \frac{k^2 dk}{(2\pi)^3}\frac{\partial P_\Phi}{\partial t}(k,z,z') j_\ell(kr') j_\ell(kr''). \end{aligned} \tag{C.67}$$

We then apply the Limber approximation (see Appendix C.4), integrate out the delta function introduced, and change integration variable to obtain,

$$\begin{aligned} Q_\ell(\nu,\nu) = \frac{2}{c^4}\int dz W_\nu(z)\eta(z)\int dz' W_\nu(z') \\ \times \int_0^{z'} dz'' \frac{S_k[r(z')-r(z'')]}{S_k[r(z')]S_k[r(z'')]r(z'')^2} \\ \times \frac{\partial P_\Phi}{\partial z}(k,z'')\Big|_{k=\ell/r(z'')}, \end{aligned} \tag{C.68}$$

where

$$\eta(z) = -(1+z)\frac{d\delta\bar{T}_\mathrm{b}}{dz}(z), \tag{C.69}$$

and

$$P_\Phi(k,z) = \left(\frac{3}{2}\Omega_{\mathrm{M},0}\right)^2\left(\frac{H_0}{k}\right)^4 P(k,z)(1+z)^2. \tag{C.70}$$

Finally, we assume that both ν and the integral of the power spectrum vary slowly over the width of the window, which results in equation (5.40).

C.7 Bispectrum data vector for Fisher analysis

Based on the idea that a normalised sum of independent random variables tends towards a Gaussian distribution via the central limit theorem, we show here that taking the bispectrum data vector to be

$$\boldsymbol{x}^\nu_{\ell_1\ell_2\ell_3} = \sum_{m_1 m_2 m_3} a^\nu_{\ell_1 m_1} a^\nu_{\ell_2 m_2} a^\nu_{\ell_3 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \tag{C.71}$$

results in a Gaussian distributed data vector, to which the application of the Fisher analysis is valid.

For large ℓ-modes, the central limit theorem can be applied to determine that a weighted average of the product of three angular coefficients

$$A^\nu_{\lambda\mu} \equiv a^\nu_{\ell_1 m_1} a^\nu_{\ell_2 m_2} a^\nu_{\ell_3 m_3}, \tag{C.72}$$

where $\lambda = (\ell_1, \ell_2, \ell_3)$ and $\mu = (m_1, m_2, m_3)$, namely

$$\hat{b}^\nu_\lambda = \sum_\mu A^\nu_{\lambda\mu} w_{\lambda\mu}, \tag{C.73}$$

where $w_{\lambda\mu}$ denote some weighting function, is Gaussian distributed as desired. Now, we assume full-sky coverage to eliminate any ℓ-mode mixing and determine the weights for a minimal-variance unbiased estimator. We have that the bispectrum, in this notation, takes the following form,

$$\left\langle A^\nu_{\lambda\mu} \right\rangle = b^\nu_\lambda W_{\lambda\mu}, \tag{C.74}$$

where $W_{\lambda\mu}$ is the Wigner-3J symbol and b^ν_λ is the reduced bispectrum. The expectation value of the data vector is then

$$\left\langle \hat{b}^\nu_\lambda \right\rangle = b^\nu_\lambda \sum_\mu w_{\lambda\mu} W_{\lambda\mu}, \tag{C.75}$$

where the requirement for an unbiased estimator results in the condition

$$\sum_\mu w_{\lambda\mu} W_{\lambda\mu} = 1. \tag{C.76}$$

Next, we compute the variance of the data vector with the aim to discover the weights which minimize it. The variance is

$$\begin{aligned}
\left\langle \left(\hat{b}^\nu_\lambda\right)^2 \right\rangle - \left\langle \hat{b}^\nu_\lambda \right\rangle^2 &= \sum_{\mu\eta} \left\langle A^\nu_{\lambda\mu} A^\nu_{\lambda\eta} \right\rangle w_{\lambda\mu} w_{\lambda\eta} - \left(b^\nu_\lambda\right)^2 \\
&= \sum_{\mu\eta} \left\langle a^\nu_{\ell_1 m_1} a^\nu_{\ell_2 m_2} a^\nu_{\ell_3 m_3} a^{*\nu}_{\ell_1 m_4} a^{*\nu}_{\ell_2 m_5} a^{*\nu}_{\ell_3 m_6} \right\rangle w_{\lambda\mu} w_{\lambda\eta} - \left(b^\nu_\lambda\right)^2 \\
&= \sum_{\mu\eta} \left(\left\langle a^\nu_{\ell_1 m_1} a^{*\nu}_{\ell_1 m_4} \right\rangle \left\langle a^\nu_{\ell_2 m_2} a^{*\nu}_{\ell_2 m_5} \right\rangle \left\langle a^\nu_{\ell_3 m_3} a^{*\nu}_{\ell_3 m_6} \right\rangle + \text{cycl.} \right) w_{\lambda\mu} w_{\lambda\eta} \\
&= C^\nu_{\ell_1} C^\nu_{\ell_2} C^\nu_{\ell_3} \sum_{\mu\eta} \delta^{\mathrm{K}}_{\mu\eta} w_{\lambda\mu} w_{\lambda\eta} + \text{cycl.} \\
&= \Delta_\lambda C^\nu_{\ell_1} C^\nu_{\ell_2} C^\nu_{\ell_3} \sum_\mu w^2_{\lambda\mu},
\end{aligned} \tag{C.77}$$

where $\Delta_\lambda = 6, 2, \text{or } 1$, if all ℓ modes are the same, two are the same, or all are different respectively, and we have cancelled the $\left(b^\nu_\lambda\right)^2$ with the $\langle aaa \rangle \langle aaa \rangle$ term after applying Wick's theorem. To obtain the minimal variance, unbiased estimator, we minimize the

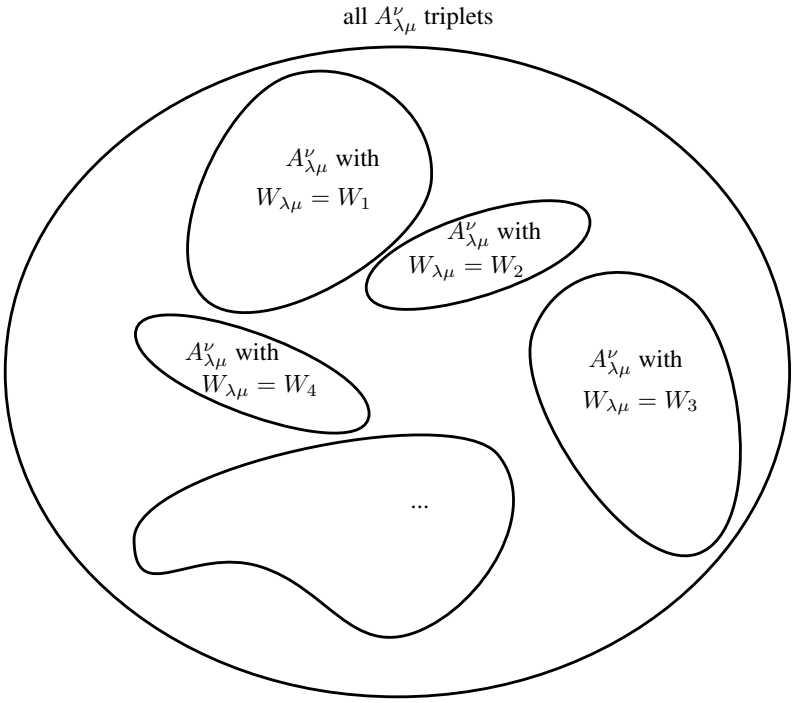

Figure C.1: We separate the ensemble of all triplets $A^{\nu}_{\lambda\mu}$ into groups within which the Wigner-3J symbols of the index groups λ and μ are identical. Then, when performing the sum over each group, each sum reduces to a weighted average with identical weight.

variance subject to the condition (C.76), using a Lagrange multiplier,

$$\begin{aligned}
&\frac{\partial}{\partial w_{\lambda\rho}}\left[\Delta_\lambda C^{\nu}_{\ell_1} C^{\nu}_{\ell_2} C^{\nu}_{\ell_3} \sum_{\mu} w^2_{\lambda\mu} - k\left(\sum_{\mu} w_{\lambda\mu} W_{\lambda\mu} - 1\right)\right] = 0 \\
&\Rightarrow 2\Delta_\lambda C^{\nu}_{\ell_1} C^{\nu}_{\ell_2} C^{\nu}_{\ell_3} w_{\lambda\rho} - k W_{\lambda\rho} = 0 \\
&\Rightarrow w_{\lambda\rho} = \text{const.}\, W_{\lambda\rho}.
\end{aligned} \tag{C.78}$$

The condition (C.76) together with the result (C.80) show that the weights taken to be the Wigner-3J symbol would result in the unbiased minimal estimator we are after.

Finally, we need to verify that such a data vector is indeed Gaussian. Let us group the triplets of $a_{\ell m}$ such that within each group the Wigner-3J symbol is the same for all triplets, see Fig C.1. Then, for large ℓ, the central limit theorem can be applied as the elements within each sum are independent and identically distributed random variables, making each group a Gaussian random variable itself. Finally, the sum of all groups is a sum of Gaussian random variables and itself Gaussian, thus showing that the minimal variance, unbiased estimator, which we have derived above, is Gaussian distributed for sufficiently large ℓ.

C.8 Orthogonality relations of Wigner-3J symbol

The Wigner-3J symbols obey the following orthogonality relation (Sobelman 1979):

$$\sum_{m_1,m_2} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell'_3 \\ m_1 & m_2 & m'_3 \end{pmatrix} = \frac{\delta_{\ell_3\ell'_3}\delta_{m_3m'_3}}{(2\ell_3+1)}, \tag{C.79}$$

for ℓ_1, ℓ_2 and ℓ_3 obeying the triangle conditions. From this result we find a corollary by summing over the last m,

$$\sum_{m_1,m_2,m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}^2 = 1, \tag{C.80}$$

where again the ℓ modes need to satisfy the triangle conditions, otherwise the sum is zero.

Appendix D

Appendix for Chapter 6

D.1 Deriving a relation between u and δ

In linear perturbation theory, the following relation for the evolution of the matter perturbations can be derived from the continuity equation for the mass density (see Peebles 1993),

$$\dot{\delta} + \frac{1}{a}\nabla \cdot \mathbf{u} = 0. \tag{D.1}$$

It is customary to relate $\dot{\delta}$ to δ directly

$$\dot{\delta} = H(a)\delta \frac{d\ln\delta}{d\ln a}. \tag{D.2}$$

Then, we define

$$f(\Omega_{\rm m}) \equiv \frac{d\ln\delta}{d\ln a}. \tag{D.3}$$

Reintroducing these equations into (D.1) yields

$$\nabla \cdot \mathbf{u} = -aH(a)f(\Omega_{\rm m})\delta. \tag{D.4}$$

Let us first consider an expansion in Fourier space

$$\nabla \cdot \mathbf{u} = \frac{1}{(2\pi)^3}\int d^3k\; i\mathbf{k}\cdot\mathbf{u}_k e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{D.5}$$

and,

$$\delta = \frac{1}{(2\pi)^3}\int d^3k\; \delta_k e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{D.6}$$

Introducing these expansions into (D.4) one finds

$$i\mathbf{k}\cdot\mathbf{u}_k = -aH(a)f(\Omega_{\rm m})\delta_k. \tag{D.7}$$

This can be used along with the fact that $\mathbf{u}_k \propto \mathbf{k}$, to show that

$$\mathbf{u}_k = -\frac{a}{ik^2} H(a) f(\Omega_\mathrm{m}) \delta_k \mathbf{k}. \tag{D.8}$$

So finally,

$$\mathbf{u}(\mathbf{r}) = \frac{ia}{(2\pi)^3} H(a) f(\Omega_\mathrm{m}) \int d^3k \delta_k \mathbf{k} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{D.9}$$

Now, for the expansion in harmonic space it is instructive to change notation from $\mathbf{u}$ to $\mathbf{v}$ as radial and transverse components are related to different transformation functions. Here, we relate the velocity vector $\mathbf{v}$ to the potential Ψ,

$$\mathbf{v} = -\nabla\Psi. \tag{D.10}$$

Since k^2 is an eigenvalue of the Laplacian operator ∇^2 we can relate the coefficients $\Psi_{\ell m}(k)$ and $\delta_{\ell m}(k)$ directly via

$$\begin{aligned} (D.4) &\Rightarrow -\nabla \cdot \mathbf{v} = \nabla^2\Psi = aH(a) f(\Omega_\mathrm{m}) \delta \\ &\Rightarrow -k^2 \Psi_{\ell m}(k) = aH(a) f(\Omega_\mathrm{m}) \delta_{\ell m}(k). \end{aligned} \tag{D.11}$$

Now, we are only interested in the radial velocity,

$$u(\mathbf{r}) = \mathbf{v} \cdot \hat{r} = -\hat{r} \cdot \nabla\Psi. \tag{D.12}$$

In spherical polar coordinated the gradient operator can be expressed as

$$\nabla f = \hat{r}\frac{\partial f}{\partial r} + \hat{\theta}\frac{1}{r}\frac{\partial f}{\partial \theta} + \hat{\phi}\frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}. \tag{D.13}$$

Hence,

$$u(\mathbf{r}) = \mathbf{v} \cdot \hat{r} = -\frac{\partial\Psi}{\partial r}. \tag{D.14}$$

Then, Ψ can be expanded in harmonic space using (6.2),

$$u(\mathbf{r}) = -\sqrt{\frac{2}{\pi}} \int dk k^2 \sum_{\ell m} \Psi_{\ell m}(k) \frac{d}{dr} j_\ell(kr) Y_{\ell m}(\theta, \phi). \tag{D.15}$$

Finally, the factor of $-k^2\Psi_{\ell m}(k)$ can be replaced by relation (D.11) and the result is obtained,

$$u(\boldsymbol{r}) = aH(a) f(\Omega_\mathrm{m}) \sqrt{\frac{2}{\pi}} \int dk \sum_{\ell m} \delta^{(r)}_{\ell m}(k) \frac{d}{dr} j_\ell(kr) Y_{\ell m}(\theta, \phi). \tag{D.16}$$

Appendix E

Permissions

E.1 Permission to reproduce Paper I

Request

From: Claude Schmit <claudejpschmit@gmail.com>
To: Journals Permissions <journals.permissions@oup.com>
Date: 2/11/2018
Time: 10:47

Dear Sir/Madam,

I would like to request permission to reproduce the contents of the following paper of which I am the author in my PhD thesis: "Emulation of reionization simulations for Bayesian inference of astrophysics parameters using neural networks, C J Schmit & J R Pritchard, Monthly Notices of the Royal Astronomical Society, Volume 475, Issue 1, 21 March 2018, Pages 1213{1223, https://doi.org/10.1093/mnras/stx3292. Published: 23 December 2017"

I am writing this thesis at Imperial College London and it will not be published commercially.

Kind Regards,

Claude Schmit

Response

From: Journals Permissions <journals.permissions@oup.com>

To: Claude Schmit <claudejpschmit@gmail.com>
Date: 2/11/2018
Time: 12:27

Dear Claude Schmit,

RE. C J Schmit & J R Pritchard. Emulation of reionization simulations for Bayesian inference of astrophysics paramters using neural networks. MNRAS (2018) 475 (1): 1213-1223

Thank you for your email requesting permission to reuse all or part of your article in a thesis/dissertation.

As part of your copyright agreement with Oxford University Press you have retained the right, after publication, to use all or part of the article and abstract, in the preparation of derivative works, extension of the article into a booklength work, in a thesis/dissertation, or in another works collection, provided that a full acknowledgement is made to the original publication in the journal. As a result, you should not require direct permission from Oxford University Press to reuse your article.

Authors may upload a PDF of the accepted manuscript to institutional and/or centrally organized repositories and/or in free public servers, upon acceptance for publication in the journal. Authors may upload the version of record to institutional and/or centrally organized repositories and/or in free public servers, upon publication in the journal.

Please Note: Inclusion under a Creative Commons License or any other Open-Access License allowing onward reuse is prohibited.

For full details of our publication and rights policy, including credit-lines, please see the attached link to our website:

https://academic.oup.com/journals/pages/access_purchase/rights_and_permissions/self_archiving_policy_p

If you have any other queries, please feel free to contact us.

Kind regards,

Katie

Katie Randall | Permissions Assistant | Rights Department
Academic and Journals Divisions | Global Business Development
Oxford University Press | Great Clarendon Street | Oxford | OX2 6DP

E.2 Permissions to reproduce third-party figures

E.2.1 Figures taken from articles published in MNRAS

These permissions include Figs 3.1, 3.3, 3.4, and 3.6.

Request

From: Claude Schmit <claude.schmit13@imperial.ac.uk>
To: Journals Permissions <journals.permissions@oup.com>
Date: 9/11/2018
Time: 14:11

Dear Sir/Madam,

I was wondering if you'd be able to help me. I am currently completing my PhD thesis at Imperial College London entitled "Analytic and numerical analysis of the cosmic 21cm signal". I seek permissions to include in my thesis several figures which were published across various articles in your journal. I have already tried the automated RightsLink service, but was informed to send an inquiry to you directly.

In my thesis I am wishing to include the following figures:

Figure 2 of "Spin exchange rates in proton-hydrogen collisions, Steven R. Furlanetto & Michael R. Furlanetto, Monthly Notices of the Royal Astronomical Society, Volume 379, Issue 1, 21 July 2007, Pages 130{134, https://doi.org/10.1111/j.1365-2966.2007.11921.x"

Figure 1 of "The scattering of Lyman-series photons in the intergalactic medium, Steven R. Furlanetto & Jonathan R. Pritchard, Monthly Notices of the Royal Astronomical Society, Volume 372, Issue 3, 1 November 2006, Pages 1093{1103, https://doi.org/10.1111/j.1365-2966.2006.10899.x"

Figure 12 of "21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal, Andrei Mesinger, Steven Furlanetto, Renyue Cen, Monthly Notices of the Royal Astronomical Society, Volume 411, Issue 2, 21 February 2011, Pages 955{972, https://doi.org/10.1111/j.1365-2966.2010.17731.x"

As well as an adaptation to:

Figure 2 of "Descending from on high: Lyman-series cascades and spin-kinetic temperature coupling in the 21cm line, Jonathan R. Pritchard & Steven R. Furlanetto, Monthly Notices of the Royal Astronomical Society, Volume 367, Issue 3, 11 April 2006, Pages 1057{1066, https://doi.org/10.1111/j.1365-2966.2006.10028.x"

The figures will be referenced as follows in the figure caption: "Taken with permission from <article reference>." or "Adapted with permission from <article reference>."

My thesis will be added to Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/ and made available to the public under a Creative Commons Attribution-NonCommercial-NoDerivs licence. If you are happy to grant me all the permissions requested, please reply to this email.

Yours sincerely,

Claude Schmit

Response

From: Journals Permissions <journals.permissions@oup.com>
To: Claude Schmit <claude.schmit13@imperial.ac.uk>
Date: 14/11/2018
Time: 10:45

Dear Claude,

RE.
Fig 2. Steven R. Furlanetto & Michael R. Furlanetto. Spin exchange rates in proton-hydrogen collisions. MNRAS (2007) 379 (1): 130-134

Fig 1. Steven R. Furlanetto & Jonathan R. Pritchard. The scattering of Lyman-series photons in the intergalactic medium. MNRAS (2006) 372 (3): 1093-1103

Fig 12. Andrei Mesinger et al. 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal. MNRAS (2011) 411 (2): 955-972

Fig 2 (adapted). Jonathan R. Pritchard & Steven R. Furlanetto. Descending from on high: Lyman-series cascades and spin-kinetic temperature coupling in the 21-cm line. MNRAS (2006) 367 (3): 1057-1066

Thank you for your response. As it is a requirement of your college to publish your thesis under a Creative Commons license, you would be permitted to include the above MNRAS figures. We cannot tell you which CC license to choose however OUPs preference would be the CC-BY-NC-ND license as it is the most restrictive.

This reuse, however, is on the proviso that you include the following credit line and disclaimer for each figure stating that the OUP material is not included under the CC license:

e.g.
Reproduced with permission from Fig 2. Steven R. Furlanetto & Michael R. Furlanetto. Spin exchange rates in proton-hydrogen collisions. MNRAS (2007) 379 (1): 130-134, DOI: 10.1111/j.1365-2966.2007.11921.x. By permission of Oxford University Press on behalf of the Royal Astronomical Society, available online at:
https://academic.oup.com/mnras/article/379/1/130/1132800?searchresult=1.
This figure is not included under the Creative Commons license of this publication. For permissions, please contact journals.permissions@oup.com

Would you be able to accommodate this for each figure?

Many thanks for your time and I look forward to hearing from you.

Kind regards,

Katie

Katie Randall | Permissions Assistant | Rights Department
Academic and Journals Divisions | Global Business Development
Oxford University Press | Great Clarendon Street | Oxford | OX2 6DP

Reply

From: Claude Schmit <claude.schmit13@imperial.ac.uk>
To: Journals Permissions <journals.permissions@oup.com>
Date: 14/11/2018
Time: 10:51

Dear Katie,

I certainly will include the credit line with each figure. Thank you so much for your help!

Best,

Claude Schmit.

Response

From: Journals Permissions <journals.permissions@oup.com>
To: Claude Schmit <claude.schmit13@imperial.ac.uk>
Date: 14/11/2018
Time: 10:54

Dear Claude,

I am glad this is acceptable. Please find your permission for these figures below.

Kind regards,

Katie

GratPerm

Dear Claude Schmit,

RE.
Fig 2. Steven R. Furlanetto & Michael R. Furlanetto. Spin exchange rates in proton-hydrogen collisions. MNRAS (2007) 379 (1): 130-134

Fig 1. Steven R. Furlanetto & Jonathan R. Pritchard. The scattering of Lyman-series photons in the intergalactic medium. MNRAS (2006) 372 (3): 1093-1103

Fig 12. Andrei Mesinger et al. 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal. MNRAS (2011) 411 (2): 955-972

Fig 2 (adapted). Jonathan R. Pritchard & Steven R. Furlanetto. Descending from on high: Lyman-series cascades and spin-kinetic temperature coupling in the 21-cm line. MNRAS (2006) 367 (3): 1057-1066

Thank you for your email requesting permission to reprint the above material. Our permission is granted without fee to reproduce the material.

Use of the OUP Material is restricted to: Inclusion in the forthcoming Ph.D. dissertation/thesis titled 'Analytic and numerical analysis of the cosmic 21cm signal' by Claude Schmit to be submitted in PDF format to Imperial College London in 2018, and uploaded to the university's research repository, URL http://spiral.imperial.ac.uk/ (NB: the copyright line must appear on the same page as the OUP material). Territory: World. Language: English.

This permission is limited to this particular use and does not allow you to use it elsewhere or in any other format other than specified above.

Permission is granted on the proviso that you include the credit line and disclaimer previously provided that excludes any OUP material from the Creative Commons license of the new publication.

If the credit line or acknowledgement in our publication indicates that material including any illustrations/figures etc was drawn or modified from an earlier source it will be necessary for you to also clear permission

with the original publisher. If this permission has not been obtained, please note that this material cannot be included in your publication.

Please do not hesitate to contact me if I can be of any further assistance.

Kind regards,

Katie

Katie Randall | Permissions Assistant | Rights Department

Academic and Journals Divisions | Global Business Development

Oxford University Press | Great Clarendon Street | Oxford | OX2 6DP

E.2.2 Figures taken from articles published in IEEE

These permissions include Figs. 3.2 and 4.1.

Explicit permission not required, statement obtained via RightsLink on the article page on ieee.org.

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.

3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

E.2.3 Figures taken from articles published in Nature

This permission concerns Fig. 3.5. See the following pages for details.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Nov 12, 2018

This Agreement between Mr. Claude Schmit ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4466470112328
License date	Nov 12, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature
Licensed Content Title	Cosmology: Hydrogen was not ionized abruptly
Licensed Content Author	Jonathan Pritchard, Abraham Loeb
Licensed Content Date	Dec 8, 2010
Licensed Content Volume	468
Licensed Content Issue	7325
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Analytic and numerical analysis of the cosmic 21cm signal
Institution name	Imperial College
Expected presentation date	Feb 2019
Portions	Figure 1
Requestor Location	Mr. Claude Schmit Exhibition Road London, SW7 2AZ United Kingdom Attn: Mr. Claude Schmit
Billing Type	Invoice
Billing Address	Mr. Claude Schmit Exhibition Road London, United Kingdom SW7 2AZ Attn: Mr. Claude Schmit

Total 0.00 USD

Terms and Conditions

Springer Nature Terms and Conditions for RightsLink Permissions

Springer Nature Customer Service Centre GmbH (the Licensor) hereby grants you a non-exclusive, world-wide licence to reproduce the material and for the purpose and requirements specified in the attached copy of your order form, and for no other use, subject to the conditions below:

1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

   If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Where **print only** permission has been granted for a fee, separate permission must be obtained for any additional electronic re-use.

3. Permission granted **free of charge** for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

4. A licence for 'post on a website' is valid for 12 months from the licence date. This licence does not cover use of full text articles on websites.

5. Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

6. Permission granted for books and journals is granted for the lifetime of the first edition and does not apply to second and subsequent editions (except where the first edition permission was granted free of charge or for signatories to the STM Permissions Guidelines http://www.stm-assoc.org/copyright-legal-affairs/permissions/permissions-guidelines/), and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence.

7. Rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

8. The Licensor's permission must be acknowledged next to the licensed material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

9. Use of the material for incidental promotional use, minor editing privileges (this does not include cropping, adapting, omitting material or any other changes that affect the meaning, intention or moral rights of the author) and copies for the disabled are permitted under this licence.

10. Minor adaptations of single figures (changes of format, colour and style) do not require the Licensor's approval. However, the adaptation should be credited as shown in Appendix below.

Appendix — Acknowledgements:

For Journal Content:

Reprinted by permission from [**the Licensor**]: [**Journal Publisher**

(e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer /Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.1

E.2.4 Figures taken from articles published in ApJ

Request for Fig 2.6

From: Claude Schmit <claude.schmit13@imperial.ac.uk>
To: Michael Blanton <Blanton@physics.nyu.edu>
Date: 21/11/2018
Time: 14:41

Dear Prof. Blanton

I was wondering if you'd be able to help me. I am currently completing my PhD thesis at Imperial College London entitled "Analytic and numerical analysis of the cosmic 21cm signal" supervised by Jonathan Pritchard and Alan Heavens. I am seeking permission to include a figure from one of your papers in my thesis into my introduction chapter and was informed by the journal to contact you directly.

The figure in question is Figure 37 of "The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey, Max Tegmark, Michael R. Blanton, Michael A. Strauss, Fiona Hoyle, David Schlegel, Roman Scoccimarro, Michael S. Vogeley, David H. Weinberg, Idit Zehavi, Andreas Berlind, The Astrophysical Journal, Volume 606, Number 2, 2004, http://iopscience.iop.org/article/10.1086/382125/meta".

The figure will be referenced as follows in the figure caption: "Taken with permission from <article reference>."

According to the Imperial College regulation my thesis will be made publicly available on Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/, under a non-commercial Creative Commons license.

If you are happy to grant me the permission to reproduce this figure, please reply to this email.

Best Regards,

Claude Schmit

Response

From: Michael Blanton <Blanton@physics.nyu.edu>
To: Claude Schmit <claude.schmit13@imperial.ac.uk>
Date: 21/11/2018
Time: 15:30

You certainly have my permission! Best of luck!

Request for Fig 3.7

From: Claude Schmit <claude.schmit13@imperial.ac.uk>
To: Adam Lidz <alidz@sas.upenn.edu>
Date: 12/11/2018
Time: 13:53

Dear Dr Lidz,

I was wondering if you'd be able to help me. I am currently completing my PhD thesis at Imperial College London entitled "Analytic and numerical analysis of the cosmic 21cm signal" supervised by Jonathan Pritchard and Alan Heavens. I am seeking permission to include a figure from one of your papers in my thesis into my introduction chapter and was informed by the journal to contact you directly.

The figure in question is Figure 1 from "Detecting the Rise and Fall of 21cm Fluctuations with the Murchison Widefield Array, Adam Lidz, Oliver Zahn, Matthew McQuinn, Matias Zaldarriaga, and Lars Hernquist, The Astrophysical Journal, Volume 680, Number 2, 2008, http://iopscience.iop.org/article/10.1086/587618/meta".

The figure will be referenced as follows in the figure caption: "Taken with permission from <article reference>."

According to the Imperial College regulation my thesis will be made publically available on Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/, under a non-commercial Creative Commons license.

If you are happy to grant me the permission to reproduce this figure, please reply to this email.

Best Regards,

Claude Schmit

Response

From: Adam Lidz <alidz@sas.upenn.edu>
To: Claude Schmit <claude.schmit13@imperial.ac.uk>
Date: 12/11/2018
Time: 21:49

Hi Claude,

Sure, you are very welcome to use this figure.

Best of luck with completing your thesis! Please also send my regards to Jonathan!

Best,

Adam

E.2.5 Figures taken from articles published in Phys. Rev. D

This permission concerns Fig 3.7. See the following pages for details.

American Physical Society
Reuse and Permissions License

12-Nov-2018

This license agreement between the American Physical Society ("APS") and Claude Schmit ("You") consists of your license details and the terms and conditions provided by the American Physical Society and SciPris.

Licensed Content Information

License Number:	**RNP/18/NOV/009304**
License date:	12-Nov-2018
DOI:	10.1103/PhysRevD.78.103511
Title:	Evolution of the 21 cm signal throughout cosmic history
Author:	Jonathan R. Pritchard and Abraham Loeb
Publication:	Physical Review D
Publisher:	American Physical Society
Cost:	USD $ 0.00

Request Details

Does your reuse require significant modifications: No

Specify intended distribution locations:	UK & Commonwealth (excluding Canada)
Reuse Category:	Reuse in a thesis/dissertation
Requestor Type:	Student
Items for Reuse:	Figures/Tables
Number of Figure/Tables:	1
Figure/Tables Details:	Figure 3
Format for Reuse:	Electronic and Print
Total number of print copies:	Up to 1000

Information about New Publication:

University/Publisher:	Imperial College London
Title of dissertation/thesis:	Analytic and numerical analysis of the cosmic 21cm signal
Author(s):	Claude Schmit
Expected completion date:	Feb. 2019

License Requestor Information

Name:	Claude Schmit
Affiliation:	Individual
Email Id:	claudejpschmit@gmail.com
Country:	United Kingdom

American Physical Society
Reuse and Permissions License

TERMS AND CONDITIONS

The American Physical Society (APS) is pleased to grant the Requestor of this license a non-exclusive, non-transferable permission, limited to Electronic and Print format, provided all criteria outlined below are followed.

1. You must also obtain permission from at least one of the lead authors for each separate work, if you haven't done so already. The author's name and affiliation can be found on the first page of the published Article.
2. For electronic format permissions, Requestor agrees to provide a hyperlink from the reprinted APS material using the source material's DOI on the web page where the work appears. The hyperlink should use the standard DOI resolution URL, http://dx.doi.org/{DOI}. The hyperlink may be embedded in the copyright credit line.
3. For print format permissions, Requestor agrees to print the required copyright credit line on the first page where the material appears: "Reprinted (abstract/excerpt/figure) with permission from [(FULL REFERENCE CITATION) as follows: Author's Names, APS Journal Title, Volume Number, Page Number and Year of Publication.] Copyright (YEAR) by the American Physical Society."
4. Permission granted in this license is for a one-time use and does not include permission for any future editions, updates, databases, formats or other matters. Permission must be sought for any additional use.
5. Use of the material does not and must not imply any endorsement by APS.
6. APS does not imply, purport or intend to grant permission to reuse materials to which it does not hold copyright. It is the requestor's sole responsibility to ensure the licensed material is original to APS and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover or table does not indicate it was reprinted by APS with permission from another source.
7. The permission granted herein is personal to the Requestor for the use specified and is not transferable or assignable without express written permission of APS. This license may not be amended except in writing by APS.
8. You may not alter, edit or modify the material in any manner.
9. You may translate the materials only when translation rights have been granted.
10. APS is not responsible for any errors or omissions due to translation.
11. You may not use the material for promotional, sales, advertising or marketing purposes.
12. The foregoing license shall not take effect unless and until APS or its agent, Aptara, receives payment in full in accordance with Aptara Billing and Payment Terms and Conditions, which are incorporated herein by reference.
13. Should the terms of this license be violated at any time, APS or Aptara may revoke the license with no refund to you and seek relief to the fullest extent of the laws of the USA. Official written notice will be made using the contact information provided with the permission request. Failure to receive such notice will not nullify revocation of the permission.
14. APS reserves all rights not specifically granted herein.
15. This document, including the Aptara Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

E.2.6 Figures taken from article pre-prints published on arxiv.org

Request for Fig 2.1

From: Claude Schmit <claude.schmit13@imperial.ac.uk>
To: George Efstathiou <gpe@ast.cam.ac.uk>
Date: 12/11/2018
Time: 14:31

Dear Prof. Efstathiou,

I was wondering if you'd be able to help me. I am currently completing my PhD thesis at Imperial College London entitled "Analytic and numerical analysis of the cosmic 21cm signal" supervised by Jonathan Pritchard and Alan Heavens. I am seeking permission to include a figure from one of the Planck collaboration papers in my thesis into my introduction chapter and I found you to be one of the corresponding authors.

The figure in question is Figure 1 from "Planck 2018 results. VI. Cosmological Parameters, Planck Collaboration, July 2018, https://arxiv.org/abs/1807.06209".

The figure will be referenced as follows in the figure caption: "Taken with permission from <article reference>."

According to the Imperial College regulation my thesis will be made publically available on Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/, under a non-commercial Creative Commons license.

If you are happy to grant me the permission to reproduce this figure, please reply to this email.

Best Regards,

Claude Schmit

Response

From: George Efstathiou <gpe@ast.cam.ac.uk>
To: Claude Schmit <claude.schmit13@imperial.ac.uk>

Date: 13/11/2018
Time: 10:05

Dear Claude

No problems -- you have permission to use the figure.

Regards

George

George Efstathiou FRS Email: gpe@ast.cam.ac.uk Professor of Astrophysics
Kavli Institute for
Cosmology Cambridge
Madingley Road Tel (Work): 01223-337530 (44 1223- outside UK)
Cambridge Tel (Home): 01223-574001 (Mobile): 07900491495
CB3 OHA
England

Request for Figs 3.8 and 3.9

From: Claude Schmit <claude.schmit13@imperial.ac.uk>
To: Amanda Weltman <amanda.weltman@uct.ac.za>
Date: 12/11/2018
Time: 14:51

Dear Prof. Weltman,

I was wondering if you'd be able to help me. I am currently completing my PhD thesis at Imperial College London entitled "Analytic and numerical analysis of the cosmic 21cm signal" supervised by Jonathan Pritchard and Alan Heavens. I am seeking permission to include two figures from a recent SKA paper and I found you to be one of the editors and convenors of the work.

The figures in question are Figures 3 and 13 from "Fundamental Physics with the Square Kilometer Array, Bull et al., October 2018,

https://arxiv.org/abs/1810.02680".

The figures will be referenced as follows in the figure caption: "Taken with permission from <article reference>."

According to the Imperial College regulation my thesis will be made publically available on Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/, under a non-commercial Creative Commons license.

If you are happy to grant me the permission to reproduce these figures, please reply to this email.

Best Regards,

Claude Schmit

Response

From: Amanda Weltman <amanda.weltman@uct.ac.za>
To: Claude Schmit <claude.schmit13@imperial.ac.uk>
Date: 12/11/2018
Time: 16:05

Hi Claude

Curiously that link takes me to the Planck results. Regardless though, you would need to get permission from the original place they are published. Most of our figures come from another publication so it is best to check which in each case and ask them. Otherwise I am fine with it of course! Congrats on finishing your PhD.

All the best
Amanda

A/Prof Amanda Weltman
South African Research Chair in Physical Cosmology
Department of Mathematics and Applied Mathematics
University of Cape Town, South Africa