Models and Probes of the Early and Dark Universe

Inflation and 21-cm Radiation in Cosmology

by

Michael Sitwell

B.Sc., Queen's University, 2008

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

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Physics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

November 2014

Abstract

The prevailing model of modern cosmology stipulates the existence of exotic substances such as dark matter and dark energy and events such as inflation. However, their underlying nature is not currently known. In this thesis, we explore new models and measurement techniques that may be used to characterize their cosmological effects and shed light on their inner workings.

A model of inflation driven by a substance that may be described macroscopically as a cosmological elastic solid is studied. The proper techniques for the quantization of perturbations within the elastic solid are presented. We find that a sufficiently rigid elastic solid with slowly varying sound speeds can produce an inflationary period. Interestingly, we find models where the elastic solid has an equation of state significantly greater than -1 that nevertheless produces nearly scale-invariant scalar and tensor spectra.

The remaining chapters of this thesis concern the use of 21-cm radiation as a probe of the physics of dark matter and dark energy.

The effects of warm dark matter on the highly-redshifted 21-cm signal is examined. If dark matter is warm instead of cold, its non-negligible velocities may inhibit the formation of low-mass halos, thereby delaying star-formation, which may delay the emission and absorption signals expected in the mean 21-cm signal. The effects of warm dark matter on both the mean 21-cm signal, as well as on its power spectrum, are described and degeneracies between the effects of warm dark matter and other astrophysical parameters are quantified.

One of the primary goals of 21-cm radiation intensity mapping is to measure baryon acoustic oscillations over a wide range of redshifts to constrain the properties of dark energy from the expansion history of the late-time Universe. We forecast the constraining power of the CHIME radio telescope on the matter power spectrum and dark energy parameters. Lastly, we devise new calibration algorithms for the gains of an interferometric radio telescope such as CHIME.

Preface

This thesis contains reprinted material originally found in the following papers:

1. Chapter 4: M. Sitwell, & K. Sigurdson, "Quantization of Perturbations in an Inflating Elastic Solid," *Phys. Rev. D*, vol. 89, 123509, 2014.

2. Chapter 6: M. Sitwell, A. Mesinger, Y. Ma, & K. Sigurdson, "The Imprint of Warm Dark Matter on the Cosmological 21-cm Signal," *MNRAS*, vol. 438, p. 2664, 2014.

3. Chapter 7: J. R. Shaw, K. Sigurdson, M. Sitwell, A. Stebbins, & U. Pen, "Coaxing Cosmic 21cm Fluctuations from the Polarized Sky using m-mode Analysis," arXiv:1401.2095, 2014.

All calculations found in Paper 1 were done by M. Sitwell, which were performed under the supervision of K. Sigurdson. The preparation of this paper was done entirely by MS, with advice from KS.

The work in Paper 2 made heavy use of the 21CMFAST code, which was written and provided by A. Mesinger. Some modifications to the code were made by MS. All analysis done on the output of this code was performed by MS. The forecasts used in this paper were provided by AM. This paper was written entirely by MS with the consultation of AM. Further feedback for this paper was given by Y. Ma and KS. Section 6.2, which does not appear in the published paper, was added to provide additional background information.

Some of the forecasting methods described in Paper 3 can be found in Chaper 7 of this thesis. The majority of the research described in this paper was conducted by J. R. Shaw and KS. The forecasts of distance measurements from the power spectrum, as well as the forecasts for the dark energy parameters, were performed by MS. The preparation of this paper was done almost entirely by JRS, in collaboration with KS. Appendix E of this paper, which was written by MS, describes the forecasting methods covered in Sections 7.6 and 7.7 of this thesis. In addition, forecasts appearing in the

CFI grant proposal for CHIME [10] (specifically those shown in Figs. 3-6) were made by MS using the methods described in Chaper 7.

Table of Contents

Appendix

List of Tables

List of Figures

List of Abbreviations

BAO	Baryon Acoustic Oscillations
BBN	Big Bang Nucleosynthesis
BOSS	Baryon Oscillation Spectroscopic Survey
CDM	Cold Dark Matter
CHIME	Canadian Hydrogen Intensity Mapping Experiment
CL	Galaxy Cluster
CMB	Cosmic Microwave Background
COBE	Cosmic Background Explorer
DM	Dark Matter
EOR	Epoch of Reionization
EPS	Extended Press-Schecther
FLRW	Friedmann-Lemaître-Robertson-Walker
FWHM	Full Width at Half Maximum
FOM	Figure of Merit
IGM	Intergalactic Medium
PS	Press-Schecther
SDSS	Sloan Digital Sky Survey
SKA	Square Kilometre Array
SN	Supernova
SVD	Singular Value Decomposition
UV	Ultraviolet
WDM	Warm Dark Matter
WF	Wouthuysen-Field
WL	Weak Lensing
WMAP	Wilkinson Microwave Anisotropy Probe

Chapter 1

Introduction

1.1 Physical Cosmology

Physical cosmology, the study of the largest scales of the Universe and its fundamental constituents, in its modern form began to take shape in the early 20th century, with such revelations as Albert Einstein's formulation of the theory of general relativity and Edwin Hubble's observational evidence for an expanding universe. Since then, cosmology has grown into a precision science, due to the remarkable measurements of the cosmic microwave background (CMB), the study of galaxies and galaxy clusters, and observations of supernovae, among many other experiments, whose successes have moved cosmology from a largely qualitative field to a quantitative one.

From these theoretical and observational leaps, the standard model of Big Bang cosmology emerged. Chiefly, it describes an expanding universe that on large scales is homogenous and isotropic. The rate of expansion is determined by basic properties of the contents of the Universe. These constituents are divided into the broad categories of matter (or baryons[1]), radiation, dark matter and dark energy. Radiation refers to relativistic species, which in the standard cosmological model include photons and neutrinos. Dark matter (DM) is a non-luminous substance that while acting gravitationally in a similar manner to normal visible matter, does not interact (or at least interacts very weakly) with the photon. Although the idea of such matter dates back to the early 1930s, its exact internal structure is currently not known. The paradigm of dark energy emerged in the 1990s to explain the observed acceleration of the expansion of the Universe. As with the similarly named dark matter, its fundamental nature is currently unknown.[2]

Augmenting the general descriptions given above, dark matter is often assumed to be cold (CDM), denoting that the dark matter should be non-

[1] This is a misnomer, as in cosmology *baryons* commonly refers all types of visible non-relativistic matter, including leptons.

[2] See Refs. [1, 2, 3] for a general introduction to modern cosmology, Ref. [4] for an introduction to particle dark matter and Refs. [5, 6] for an introduction to dark energy,

relativistic, both currently and in the early Universe. One simple and often employed model of dark energy is that of a 'cosmological constant' Λ, which when combined with the above assumptions for dark matter form the standard ΛCDM model of the Universe. While this model has been extremely successful in describing our Universe, it highlights large gaps in our current understanding, most importantly the true nature of dark matter and dark energy. The search to uncover the inner workings of dark matter and dark energy drives a significant amount of research in cosmology, as well as in physics as a whole.

Since the discovery of the expansion of our Universe, researchers have attempted to look further and further back in time, when densities and temperatures were much higher then they are currently. One early success of modern cosmology was that of Big Bang nucleosynthesis (BBN), which describes the production and abundances of the lightest nuclei, occurring at keV to MeV scales [7].

An essential component of modern cosmology is perturbation theory, which in the cosmological context describes small perturbations to the otherwise homogeneous and isotropic Universe [8, 9]. While these perturbations remain small in the early Universe, they become highly non-linear at later times and provide the early structure that eventually grows into dark matter haloes and galaxies. In this sense, these small disturbances in homogeneity and isotropy lay the seeds for the structure that we see all around us in our Universe. Through the use of general relativity, we can track the evolution of these perturbations, and we can thereby extrapolate their properties to earlier and earlier times (as long as we are in a regime where general relativity holds). These perturbations can be seen in the early Universe from the imprint left in the CMB, released approximately 380 000 years after the Big Bang, at a time known as recombination. This imprint is manifested as small anisotropies in the otherwise isotropic signal. Anisotropies in the CMB have been measured to great precision through satellite experiments such as COBE[3], WMAP[4], and Planck[5], ground-based telescopes such as ACT[6] and SPT[7], and balloon-borne experiments such as BOOMERanG[8].

[3] http://lambda.gsfc.nasa.gov/product/cobe/
[4] http://map.gsfc.nasa.gov
[5] http://www.rssd.esa.int/index.php?project=planck
[6] http://www.princeton.edu/act/
[7] http://pole.uchicago.edu
[8] http://www.astro.caltech.edu/ lgg/boomerang/boomerang_front.htm

1.2 The Origin of Perturbations and Inflation

From experiments that measure the CMB or large-scale structure, we can infer some basic properties of these perturbations when extrapolated back into the very early Universe. For example, these very early perturbations are nearly scale invariant, with a very slight preference for larger scales. A natural question to ask is: what is the origin of these perturbations? Since currently there are only a few observables that describe the pre-BBN Universe, this is a difficult question to answer.

Currently, the most popular answer is that these perturbations originated as quantum mechanical fluctuations that were stretched to cosmic scales during a brief period of extremely rapid expansion in the very early Universe, known as inflation. The popularity of inflation is due in part to its ability to solve a handful of problems that emerged in classical modern cosmology. One such problem deals with why the CMB temperature is very isotropic over the sky, even though many of the regions where the CMB was released were not in causal contact with one another according to classical modern cosmology. Another problem is why the Universe appears to have very little, if any, spatial curvature.

The persistence of inflationary theories in modern cosmology is largely due to the fact that they both provide solutions to these problems as well as producing the initial set of perturbations in the Universe. On the other hand, due to the small number of observables currently available that can place constraints on models of inflation, if inflation did occur its exact model description is not yet known. However, as inflationary models in general produce propagating gravitational disturbances, known as gravitational waves, measuring these relic gravitational waves may provide crucial evidence of inflation.

1.3 Acoustic Oscillations

Sometime shortly after the events of the very early Universe, the primordial cosmological perturbations found themselves in a radiation-dominated universe. During this extremely hot and dense era, the baryons were strongly coupled to the photons, forming a so-called baryon-radiation fluid, where to a good approximation the baryons and photons moved as one. In overdense areas, the strong radiation pressure of the photons pushes outwards, causing the photons to disperse from the area. Since the baryons are strongly coupled to the photons at this time, they are dragged along with the photons.

Particles rush out of overdense areas in a wave that propagates until the sound speed of these acoustic waves drops to zero, a time labeled as the drag era, occurring after recombination.

The presence of these acoustic waves are embedded in both the distribution of radiation, in the form of the anisotropies in the CMB, and matter, by means of the distribution of galaxies and dark matter haloes. The imprint of these waves on the baryons is known as baryon acoustic oscillations (BAO). As the acoustic waves were only able to propagate from the beginning of the radiation-dominated era until the drag era, the material flowing out of overdense regions propagated a finite distance, leaving extra matter a certain distance away from the location of the original overdensity. This creates a preferential scale in the distribution of matter, occurring at roughly $\sim 150\,\mathrm{Mpc}$.[9] As the primordial perturbation are distributed over a wide range of scales and directions, these waves overlap with one another, making it difficult to see individual signs of these waves. However, as there will be a preferential separation distance of matter at the BAO scale, the BAO signal can be observed statistically, for example as a bump at the BAO scale in the two-point correlation function (ontop of the correlation function that disregards the effect of the baryons) or equivalently as an oscillation in the matter power spectrum.

As the BAO imprints a preferential (comoving) scale into the distribution of matter, it can be used as a statistical standard ruler for measuring the expansion of the Universe, thereby giving BAO great importance in modern cosmology. The BAO scale corresponds to the sound horizon at the drag era. The first BAO detections were made in 2005 from galaxy surveys consisting of 10,000's of galaxies made by the Sloan Digital Sky Survey (SDSS) [11] and also by the Two-degree-Field Galaxy Redshift Survey [12]. Subsequent BAO detections using galaxy surveys have been made by the Six-degree-Field Galaxy Survey [13], WiggleZ [14], and BOSS [15] and has recently been detected at high redshifts in the Lyman-α forest by BOSS [16, 17] and in the cross-correlation of Lyman-α with quasars [18].

By measuring the BAO at various redshifts, we can use the BAO as a standard ruler to track the expansion of the Universe. By using this procedure with redshifts up to $z \sim 3$, a detailed expansion history of the dark energy dominated Universe may be measured. This process can be used to place constraints on models of dark energy, as various models predict slightly different expansion histories.

[9] Unless stated otherwise, all quoted distances are comoving distances chosen to coincide with present-day physical distances.

1.4 21-cm Radiation

A promising new tool for the exploration of cosmology is 21-cm radiation, the radiation emitted by the hyperfine spin-flip of neutral hydrogen (HI), which is emitted with a wavelength of about 21-cm in the rest frame of the hydrogen atom. The low excitation energy for this hyperfine transition gives it some desirable properties: it is sensitive to low temperatures and has a relatively low optical depth so can be used to probe far into the high-redshift Universe. As we can infer radial distances through the redshift of the observed radiation, in addition to the angular distribution of the emission, 21-cm radiation can be used to construct 3D 'tomographic' maps of the HI distribution in our Universe, potentially containing a plethora of new and valuable information. Furthermore, 21-cm radiation may provide our only glimpse into the 'dark ages', a time in which very few structures have formed. However, removing bright foregrounds that may be as high as three to four orders of magnitude larger than the 21-cm signal presents a formidable challenge.

The nature of the 21-cm signal changes throughout cosmic history. The 21-cm signal is measured against the CMB and may appear in either emission or absorption [19]. The 21-cm signal is likely to appear in absorption during the dark ages and slightly afterwards, and in emission shortly before reionization and afterwards. During reionization, regions of ionized hydrogen (HII) form, creating non-emitting 'bubbles' in the intergalactic medium (IGM). These bubbles can grow to the Mpc scale and eventually overlap at the end of reionization. After reionization, when HII regions in the IGM have coalesced, the origin of 21-cm emission is relegated to only dense collapsed halos that contain sufficient amounts of neutral hydrogen.

The post-reionization 21-cm signal may be used to map the underlying distribution of matter, from which the BAO signal may be extracted. Since the BAO scale is on the order of 150 Mpc, high-resolution maps of the matter distribution, such as those made from galaxy surveys, are not necessary to measure the BAO. Lower resolution maps made from the 21-cm signal may be used to measure the BAO at many redshifts, a process potentially easier than conducting vast galaxy surveys. The caveat to this is that for 21-cm measurements of the BAO scale to be successful, the very bright foregrounds comprised mainly of synchrotron radiation must be removed to a sufficient level.

21-cm radiation may also shed light onto the details of exactly how and when reionization took place. Much is currently unknown about how long the epoch of reionization (EOR) lasted and exactly how ionized HII regions

formed. Observations of the Gunn-Peterson trough [20] in the spectrum of distant quasars, caused by the scattering of photons that pass through continuous HI regions while redshifting through the Lyman-α line, places the end of reionization around $z \sim 6$ [21, 22]. However, due to its high cross-section, the Universe is opaque to Lyman-α emission at higher redshifts. On the other hand, the lower optical depth of 21-cm radiation makes it well suited as a probe of the EOR at even higher redshift eras, including when the 21-cm signal may have been in absorption both before and after significant structure formation has taken place.

1.5 Measuring the Effects of Dark Energy

What many consider to be the first substantial evidence of a late-time accelerating Universe came in 1998 with the observations of type Ia supernovae (SN). These SN can act as standard candles, as their peak brightness consistently hits at approximately the same point, and so can be used to trace the expansion of the Universe. The SN observations of Riess et al. [23] and Perlmutter el al. [24] both showed evidence for a late-time acceleration, a result that has since been supported by further observations.

There exist many different models of dark energy (or models that produce a similar real or perceived acceleration), such as a cosmological constant, scalar field models (quintessence), and modified gravity, to name a few. For a model to be consistent with observations, the late-time equation of state of the dark energy $w_{\rm DE}$ must be close to -1. However, different models predict slight departures from $w_{\rm DE} = -1$.[10] With current measurements consistent with $w_{\rm DE} = -1$, a driving force in dark energy research is to obtain more constraining measurements of $w_{\rm DE}$.

As previously mentioned, 21-cm experiments designed to measure the BAO are well suited for this purpose, many of which have recently gone into operation or are to be built in the near future. Many of these experiments are interferometric telescopes that are similar in design to EOR experiments (e.g. LOFAR[11], MWA[12], PAPER[13]). The Canadian Hydrogen Intensity Mapping Experiment[14] (CHIME) is one such radio telescope, being built in Penticton, British Columbia. CHIME is a drift scan telescope

[10] A cosmological constant predicts $w_{\rm DE} = -1$ exactly.
[11] http://www.lofar.org
[12] http://www.mwatelescope.org
[13] http://eor.berkeley.edu
[14] http://chime.phas.ubc.ca

with no moving parts that will consist of five 100m $\times$ 20m cylindrical reflectors with 256 dual-polarization feeds running down the focal line of each cylinder. A smaller scale pathfinder telescope of two 35m long cylinders with 128 feeds on each cylinder, constructed in late 2013, will prototype the full CHIME telescope. The cylinders are aligned with the North-South direction to provide at any one time a wide field of view of the sky in the NS direction and narrow one in the East-West direction. CHIME will be capable of mapping nearly half of the sky in the course of a day. CHIME will observe the sky in the frequency range of $400 - 800\,\mathrm{MHz}$ (wavelengths of $\sim 37 - 75\,\mathrm{cm}$) in order to measure the BAO at redshifts in the range $z \approx 0.8 - 2.5$, a time period when the effects of dark energy first becomes prominent. The measurement of the BAO scale in this redshift range will complement measurements already made at lower redshifts.

1.6 Cosmological History in Brief

As cosmology studies the evolution of the Universe from its birth to the present day, there have been many important events that have occurred in the history of the Universe. To conclude this introduction, I give a brief overview of cosmic history. In the following list, time is demarcated by either a temperature T or redshift z, where the former is more convenient at early times and the latter at later times.

- $T \gtrsim \mathrm{few\,MeV}$, $z \gtrsim 10^9$ *The very early Universe*: Many significant events might have occurred during this time, for example grand unification or baryogenesis. Inflation, if it occurred, would belong to this time period (see Chapters 3 and 4).
- $T \sim 0.1{-}10\,\mathrm{MeV}$, $z \sim 4 \times 10^8{-}4 \times 10^{10}$ *Big-Bang Nucleosynthesis*: Light nuclei are formed.
- $T \sim 1\,\mathrm{MeV}$, $z \sim 4 \times 10^9$ *Neutrino Decoupling*: Neutrinos decouple from other species and free-stream thereafter.
- $T \sim 0.5\,\mathrm{MeV}$, $z \sim 2 \times 10^9$ *Electron-Positron Annihilation*: Electrons and positrons annihilate with one another, a relatively small number of electrons persist past the annihilation.
- $T \sim 1\,\mathrm{eV}$, $z \sim 4250$ *Matter-Radiation Equality*: The Universe becomes matter dominated past this point.

- $T \sim 0.26\,\mathrm{eV}$, $z \sim 1100$ *Recombination*: Free electrons and protons combine to form hydrogen, the CMB is released.
- $T \sim 1.6{-}2.6\,\mathrm{meV}$, $z \sim 6{-}10$ *Reionization*: Radiation from early astrophysical sources ionize hydrogen in the IGM (see Chapters 5 and 6).
- $T \sim 0.33\,\mathrm{meV}$, $z \sim 0.4$ *Matter-Dark Energy Equality*: The Universe is dominated by dark energy past this point (see Chapter 7).

Chapter 2

The Universe: Background, Linear Perturbations, Nonlinear Structures

2.1 The Unperturbed Universe

2.1.1 The FLRW Spacetime

On large scales ($\gtrsim 100\,\mathrm{Mpc}$) the Universe is very homogenous and isotropic. While this was established empirically in the late 20th century, this was a commonly used assumption well before this time. As such, much insight can be gained from perfectly homogenous and isotropic models of the Universe, which can later be extended to allow for small amounts of inhomogeneity and anisotropy.

The most general homogenous and isotropic spacetime admitted from the Einstein equations is the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, which can be represented by

$$ds^2 = \bar{g}_{\mu\nu} dx^\mu dx^\nu = dt^2 - a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2 d\Omega^2 \right], \tag{2.1}$$

where $\bar{g}_{\mu\nu}$ is the background FLRW metric and $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$. The spatial curvature constant K can assume the values $0, 1, -1$ for a spatially flat, open, and closed spacetime, respectively. The expansion of the spacetime is controlled by the scale factor $a(t)$ and its evolution is given by the Hubble rate $H(t) = \dot{a}(t)/a(t)$, where an overdot denotes differentiation with respect to t. Often it will be more convenient to work with the conformal time $\eta \equiv \int dt/a(t)$ instead of the coordinate time t. In these cases, we will also make use of the conformal Hubble rate $\mathcal{H} \equiv a'/a = aH$, where prime $'$ stands for $\partial/\partial\eta$.

It is easy to show that the stress-energy tensor for a perfect fluid

$$T^\alpha_\beta = (\rho + P) u^\alpha u_\beta - P \delta^\alpha_\beta \tag{2.2}$$

yields a homogenous and isotropic spacetime, where ρ is the energy density, P is the pressure scalar, and u^α is the 4-velocity. Solving the Einstein equations with this metric and stress-energy tensor yields the Friedmann equations

$$H^2 + \frac{K}{a^2} = \frac{8\pi G}{3}\rho, \tag{2.3a}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}G\rho(1+3w). \tag{2.3b}$$

In the above equation, we have introduced the equation of state parameter $w \equiv P/\rho$. It is important to note that, from (2.3b), if the equation of state is larger than $-1/3$, the expansion of the spacetime will be decelerating, while a value of w smaller than $-1/3$ leads to an accelerating expansion. Eliminating ρ in the above equations yields the useful differential equation

$$\mathcal{H}' = -\frac{1+3w}{2}\mathcal{H}^2. \tag{2.4}$$

In the case where w is constant, this differential equation can easily be solved as

$$\mathcal{H} = \frac{2}{(1+3w)(\eta - \eta_c)}, \tag{2.5}$$

where η_c is a constant of integration.

For a noninteracting perfect fluid, the covariant conservation of the stress-energy tensor yields the energy (density) conservation equation

$$\dot{\rho} + 3H\rho(1+w) = 0. \tag{2.6}$$

For constant w, this implies that $\rho \propto a^{-3(1+w)}$.

The energy density is typically decomposed as a sum of components with different equations of state. In the ΛCDM model, all substances have one of three equations of state: nonrelativistic matter that has $w = 0$, radiation which has the relativistic equation of state $w = 1/3$, and the cosmological constant Λ with $w = -1$. The present-day energy density of each substance i is often expressed as a fraction Ω_i of the critical density ρ_{cr}, where the critical density is the density that yields a flat spacetime with a Hubble rate matching the present-day value. The Friedmann equation (2.3a) can then be neatly expressed as

$$H^2 = H_0^2(\Omega_{\mathrm{m}}a^{-3} + \Omega_{\mathrm{r}}a^{-4} + \Omega_\Lambda + \Omega_{\mathrm{k}}a^{-2}), \tag{2.7}$$

where the subscripts m and r denote matter and radiation, respectively, the subscript 0 denotes the present-day value, and $\Omega_{\mathrm{k}} = -K/(a_0H_0)^2$. The

scale factor is normalized as $a_0 = 1$. Current observed values for these parameters are very roughly $\Omega_\Lambda \approx 0.73, \Omega_\mathrm{m} \approx 0.27, \Omega_\mathrm{k} \approx 0, \Omega_\mathrm{r} \approx 8 \times 10^{-5}$, and $h \approx 0.7$, where h is the present-day Hubble parameter H_0 expressed in units of $100\,\mathrm{km\,s^{-1}Mpc^{-1}}$ [25]. In certain situations it is convenient to use the parameters $\omega_\mathrm{m} = \Omega_\mathrm{m} h^2$ and $\omega_\mathrm{b} = \Omega_\mathrm{b} h^2$.

2.1.2 Distances and Times in Cosmology

When dealing with an expanding universe, the notion of a distance can be ambiguous and requires a more precise definition then when thought of in the Newtonian sense. Furthermore, when using units where $c = 1$, distances and times share the same units and in many contexts can be thought of interchangeably.

One of the most basic of such measures is the redshift $z = (\lambda_\mathrm{ob} - \lambda_\mathrm{em})/\lambda_\mathrm{em}$ of light emitted with wavelength λ_em and observed with wavelength λ_ob. The redshift is commonly used in place of the scale factor a, related by $a_0/a = 1 + z$.

An important distinction is the difference between physical (proper) distances L_ph and comoving distances L, where comoving coordinates denote coordinates that are defined such that they remain constant with respect to the motion of particular objects. In other words, for objects that are comoving with the Hubble flow, the separation in their comoving distance remains constant. Physical and comoving distances are related by $L_\mathrm{ph} = aL$. Up to an integration constant, the conformal time η measures the comoving distance that a massless particle travels in a certain duration. More precisely, a particle traveling at the speed of light travels a comoving distance $\chi = \eta(t_2) - \eta(t_1)$ between the times t_1 and t_2. Unless otherwise stated, the later time t_2 is assumed to be the present day and then χ is a function of a single time parameter. Thus, if the zero point of η is chosen appropriately, $\eta(t)$ measures the particle horizon of a massless particle at time t.

By expressing the conformal time as the integral

$$\eta = \int \frac{d\tilde{a}}{\tilde{a}} \mathcal{H}^{-1}(\tilde{a}) \tag{2.8}$$

we can see that massless particles can travel roughly a comoving distance $\mathcal{H}^{-1}$ in the time that the scale factor increases by a factor of e. In this light, the Hubble radius H^{-1} (or $\mathcal{H}^{-1}$ in comoving coordinates) is commonly referred to as the 'horizon'.

The angular diameter distance D_A is used to relate the physical size L_ph of a very distant object to the angle θ that it subtends by $D_\mathrm{A} = L_\mathrm{ph}/\theta$. In

general, the angular diameter distance is given by

$$D_{\mathrm{A}}(z) = \frac{1}{1+z}\frac{1}{H_0\sqrt{\Omega_{\mathrm{k}}}}\sinh\left(\sqrt{\Omega_{\mathrm{k}}}H_0\chi(z)\right), \tag{2.9}$$

but greatly simplifies in a flat universe to $D_{\mathrm{A}}(z) = a\chi(z)$.

2.2 Thermodynamics

Macroscopic thermodynamic quantities are ubiquitous in cosmology, as it is commonplace to find substances in thermodynamic equilibrium.

The number density n, energy density ρ, and pressure P can be expressed as [1, 2]

$$n = g\int \frac{d^3p}{(2\pi)^3} f(E), \tag{2.10a}$$

$$\rho = g\int \frac{d^3p}{(2\pi)^3} f(E)E, \tag{2.10b}$$

$$P = g\int \frac{d^3p}{(2\pi)^3} f(E)\frac{p^2}{3E}, \tag{2.10c}$$

where g is the number of degrees of freedom and f is the distribution function (Bose-Einstein or Fermi-Dirac).[15] The entropy density s can be found via the thermodynamic identity as

$$s = \frac{\rho + P - \mu n}{T}. \tag{2.11}$$

In most situations, substances are either non-relativistic or ultra-relativistic. An ultra-relativistic substance with temperature T and particle mass m satisfies $T \gg m$, in which case for a boson the above relations simplify to

$$n^B = \frac{\zeta(3)}{\pi^2} gT^3, \qquad \rho^B = \frac{\pi^2}{30} gT^4, \qquad P^B = \frac{\pi^2}{90} gT^4, \tag{2.12}$$

where ζ is the zeta function. For a fermion, we get $n^F = (3/4)n^B$, $\rho^F = (7/8)\rho^B$, $P^F = (7/8)P^B$. From these expressions we can see that indeed $w = 1/3$ for an ultra-relativistic species, for both bosons and fermions. The entropy density for a boson is then

$$s^B = \frac{2\pi^2}{45} gT^3 \tag{2.13}$$

[15] We set $c = \hbar = k_{\mathrm{b}} = 1$ in this section.

and $s^F = (7/8)s^B$ for a fermion.

In the nonrelativistic limit $T \ll m$, for both bosons and fermions we have

$$n = g\left(\frac{mT}{2\pi}\right)^{3/2} e^{(\mu-T)/T}, \qquad \rho = mn, \qquad P = nT. \tag{2.14}$$

For nonrelativistic matter we see that $w \approx 0$, as asserted in the previous section.

An important application of the above thermodynamic relations is to a plasma composed of relativistic particles. Suppose the plasma contains N_b (N_f) relativistic bosons (fermions), with each species b (f) having g_b (g_f) degrees of freedom and is at equilibrium temperature T_b (T_f). Using Eqs. (2.12) and (2.13), we can write the total energy, entropy, and number densities of the relativistic plasma in terms of temperature dependent effective degrees of freedom as

$$\rho_r(T) = \frac{\pi^2}{30} g(T) T^4, \tag{2.15a}$$

$$s_r(T) = \frac{2\pi^2}{45} g_s(T) T^3, \tag{2.15b}$$

$$n_r(T) = \frac{\zeta(3)}{\pi^2} g_n(T) T^3, \tag{2.15c}$$

where T is the photon temperature, and g, g_s, and g_n are the number of effective relativistic degrees of freedom contributing towards the energy, entropy, and number densities, respectively, given by

$$g(T) = \sum_b g_b \left(\frac{T_b}{T}\right)^4 + \frac{7}{8}\sum_f g_f \left(\frac{T_f}{T}\right)^4, \tag{2.16a}$$

$$g_s(T) = \sum_b g_b \left(\frac{T_b}{T}\right)^3 + \frac{7}{8}\sum_f g_f \left(\frac{T_f}{T}\right)^3, \tag{2.16b}$$

$$g_n(T) = \sum_b g_b \left(\frac{T_b}{T}\right)^3 + \frac{3}{4}\sum_f g_f \left(\frac{T_f}{T}\right)^3. \tag{2.16c}$$

Note that $g = g_s$ if all species are in thermodynamic equilibrium at temperature $T = T_b = T_f$.

These expressions are useful for describing the energy and entropy densities in the early Universe, since the early Universe is radiation dominated.

For example, in the standard model of particle physics, at temperatures well above the top mass ($T > 173\,\mathrm{GeV}$) all particles in the standard model will be relativistic and we will have $g = g_s = 106.75$. As the temperature drops, the values of the effective degrees of freedom decrease as particles become nonrelativistic. All relativistic particles in the plasma are in thermodynamic equilibrium with each other (so $g = g_s$) until $T \sim \mathrm{MeV}$ when neutrinos decouple. In addition, just after neutrino decoupling, electrons and positrons annihilate, which dumps their entropy into the photons but not into the neutrinos, since by this time they have decoupled, which results in a heating of the photons relative to the neutrinos. Thus after electron-positron annihilation, not all relativistic species share the same temperature and $g \neq g_s$. For $T \ll \mathrm{MeV}$, after electron-positron annihilation, we will have $g \approx 3.36$ and $g_s = 43/11 \approx 3.91$.

Throughout much of the history of the Universe, local thermal equilibrium held and thus the total entropy density remained constant. A useful consequence of this is that for these times, we can use the conservation of entropy to relate the expansion of the Universe to the temerpature by

$$\frac{a(T_1)}{a(T_2)} = \left(\frac{g_s(T_2)}{g_s(T_1)}\right)^{1/3} \frac{T_2}{T_1}, \tag{2.17}$$

where T_1 and T_2 are photon temperatures at two different times.

2.3 Linear Perturbation Theory

2.3.1 Notation and Conventions

We now extend the results from the previous section to allow for small perturbations that break homogeneity and isotropy. Since these deviations remain small for a large part of cosmic history on many relevant scales, linear perturbation has become an essential tool in cosmology.[16]

We begin by perturbing the metric

$$ds^2 = a^2(\eta)(\eta_{\alpha\beta} - h_{\alpha\beta})dx^\alpha dx^\beta, \tag{2.18}$$

where $\eta_{\alpha\beta}$ is the metric for Minkowski space and $h_{\alpha\beta}$ represents small perturbations about the background. It will prove useful to further parameterize the perturbations $h_{\alpha\beta}$. To do this, we first separate the time-like and spatial parts as

$$ds^2 = a^2(\eta)\left[(1+2\phi)d\eta^2 - 2B_i dx^i d\eta - h_{ij}dx^i dx^j\right]. \tag{2.19}$$

[16] This section largely uses the notation of Ref. [9]. See also Ref. [26].

This introduces a scalar ϕ, a vector B_i, and a tensor h_{ij}. The vector and tensor perturbations can be further decomposed into scalar, vector and tensor parts according to how each part transforms under spatial transformations. B_i decomposes as $B_i = B_{,i} + S_i$ (where $_{,i} = \partial_i$), comprised of a scalar B and a divergenceless vector S_i. For the tensor h_{ij}, we first decompose it as $h_{ij} = (1 + h/3)\delta_{ij} + 2E_{ij}$, where h is the trace of h_{ij} and E_{ij} is traceless. E_{ij} is subsequently decompose into scalar, vector, and tensor components, where the full tensor formed from each component is denoted by E^{S}_{ij}, E^{V}_{ij}, and E^{T}_{ij}, respectively. The tensors formed from the scalar component E and vector component E_i are given by

$$E^{\mathrm{S}}_{ij} = E_{,ij} - \frac{1}{3}\delta_{ij}\nabla^2 E \tag{2.20a}$$

$$E^{\mathrm{V}}_{ij} = E_{(i,j)} \tag{2.20b}$$

where the curved brackets in the subscript denotes symmetrization (in other words $E_{(i,j)} = \frac{1}{2}(E_{i,j} + E_{j,i})$). In addition to being traceless, the tensor E^{T}_{ij} is transverse meaning $E^{\mathrm{Ti}}_{j,i} = 0$. Lastly, we make two notational changes to conform to popular conventions by denoting $2E^{\mathrm{T}}_{ij}$ by h^{T}_{ij} and using the scalar perturbation $\psi = -\frac{1}{6}h + \frac{1}{3}\nabla^2 E$ in place of h.

Since perfect fluids are often used to model substances in cosmology, we now perturb the stress-energy tensor of a perfect fluid given in Eq. (2.2) as

$$\delta T^0_0 = \delta\rho, \tag{2.21a}$$

$$\delta T^i_0 = (\rho + P)v^i, \tag{2.21b}$$

$$\delta T^i_j = -(\delta P \delta^i_j + P\Pi^i_j), \tag{2.21c}$$

where $\delta\rho$ and δP are the energy density and pressure perturbations, respectively, v^i is the velocity perturbation, and Π^i_j is the anisotropic stress. The velocity perturbation v^i and anisotropic stress Π^i_j can be decomposed into scalar, vector, and tensor parts in the same manner as the metric perturbations B_i and E_{ij}, and denote the scalar parts by v and Π, respectively.

We will often transform from position space into Fourier space with wave vectors $\mathbf{k}$.[17] When doing so, we use the convention of including an extra factor of $k = |\mathbf{k}|$ in the Fourier variables for the scalar component of vectors (such as B and v) and a factor of k^2 for the scalar component of tensors (such as E and Π), so that perturbations in Fourier space all have the same dimensions. A real, homogenous, and isotropic Gaussian field f can be

[17] Here we use the Fourier conventions $f(\mathbf{x}) = (2\pi)^{-3/2}\int d^3k f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}$.

described by a power spectrum $\mathcal{P}_f(k)$, which describes the variance of its Fourier components and is given by

$$\langle f_{\mathbf{k}} f_{\tilde{\mathbf{k}}} \rangle = (2\pi^2/k^3)\mathcal{P}_f(k)\delta(\mathbf{k}+\tilde{\mathbf{k}}). \tag{2.22}$$

The correlation function ξ_f for f is then

$$\xi_f(|\mathbf{x}-\tilde{\mathbf{x}}|) = \int \frac{dk}{k}\mathcal{P}_f(k)\frac{\sin(k|\mathbf{x}-\tilde{\mathbf{x}}|)}{k|\mathbf{x}-\tilde{\mathbf{x}}|}. \tag{2.23}$$

2.3.2 Choosing a Gauge

Implicit in decomposing our spacetime into background and perturbed spacetimes are the coordinate systems used on each [26]. If one first defines a coordinate system on the background spacetime, there can be many mappings of points on the background spacetime to points on the perturbed spacetime. For functions defined on the perturbed spacetime, each choice of mapping will yield a different value for the function for the same point on the background spacetime. In other words, the perturbation variables defined in the previous section may change their values for different coordinate systems in the perturbed spacetime. We can relate two such coordinate systems x^α and $\tilde{x}^\alpha$ by $\tilde{x}^\alpha = x^\alpha + \xi^\alpha$. Switching coordinate systems in this manner is known as a gauge transformation. The spatial part of the vector relating the two coordinate systems ξ^α can be decomposed as $\xi^i = \zeta^{,i} + \xi^i_\perp$, where $\xi_\perp$ is divergenceless. When changing coordinates, the scalar metric perturbations transform as [18]

$$\tilde{\phi} = \phi - \mathcal{H}\xi^0 - (\xi^0)', \tag{2.24a}$$

$$\tilde{\psi} = \psi + \mathcal{H}\xi^0, \tag{2.24b}$$

$$\tilde{B} = B - \xi^0 + \zeta', \tag{2.24c}$$

$$\tilde{E} = E + \zeta, \tag{2.24d}$$

where perturbations in the coordinate system $\tilde{x}^\alpha$ (x^α) are denoted with (without) a tilde. A scalar variable q defined in the perturbed spacetime that is decomposed into a background component $\bar{q}$ and a perturbed component δq transforms as

$$\delta\tilde{q} = \delta q - \bar{q}'\xi^0. \tag{2.25}$$

[18] We remind the reader that a prime represents a partial derivative with respect to conformal time.

Relevant examples of this are the energy density ρ and pressure P. A 4-vector w^i, such as the 4-velocity, will transform as

$$\tilde{w}^0 = w^0 + \bar{w}^0(\xi^0)' - (\bar{w}^0)'\xi^0, \qquad \tilde{w}^i = w^i + \bar{w}^0(\xi^i)'. \tag{2.26}$$

Vector perturbations are generally not considered, as in most situations they decay very rapidly. Lastly, we note that since the perturbation to the spatial part of a tensor transforms as $\delta\tilde{C}^i_j = \delta C^i_j - \frac{1}{3}\delta^i_j(\bar{C}^k_k)'\xi^0$, the traceless part of the tensor δC^i_j, given by $\delta C^i_j - \frac{1}{3}\delta^i_j\delta C^k_k$, will be unchanged by a gauge transformation and thus the tensor perturbations are gauge-invariant.

Choosing a particular coordinate system in the perturbed spacetime corresponds to picking a gauge for the perturbation variables. One can then move between gauges by using the 4-vector ξ^α that relates the gauges and the transformations listed above. Choosing a coordinate system for the time and spatial variables is referred to a slicing and threading, respectively. Often it is convenient to pick a gauge where certain perturbations vanish. The usefulness of a gauge usually depends on the situation. A few popular choices of gauge are listed in Table 2.1.

Gauge	Condition
Conformal-Newtonian	$B = E = 0$
Synchronous	$\phi = B = 0$
Comoving	$v = B = 0$
Off-Diagonal	$\psi = E = 0$

Table 2.1: Popular gauge choices for the scalar perturbations.

Instead of picking a gauge to work in, in some situations it is helpful to use gauge-invariant variables. Although there are multiple ways of defining such variables, they are most commonly defined as

$$\Phi = \phi + \mathcal{H}(B - E') + (B - E')', \tag{2.27a}$$

$$\Psi = \psi - \mathcal{H}(B - E'), \tag{2.27b}$$

$$\delta\rho^{(\mathrm{gi})} = \delta\rho + \rho'(B - E'), \tag{2.27c}$$

$$\delta P^{(\mathrm{gi})} = \delta P + P'(B - E'), \tag{2.27d}$$

$$v^{(\mathrm{gi})} = v + B - E'. \tag{2.27e}$$

In the above equations, ρ and P denote their background quantities, a convention which we use from here onwards unless stated otherwise.

2.3.3 Linear Einstein Equations

Perturbing the Einstein equations is a straightforward but somewhat lengthly procedure (see Ref. [9] for more details). Using the stress-energy tensor in Eq. (2.21), the scalar gauge-invariant equations are

$$\nabla^2\Psi - 3\mathcal{H}(\Psi' + \mathcal{H}\Phi) = \frac{3}{2}l^2a^2\delta\rho^{(\mathrm{gi})}, \tag{2.28a}$$

$$\Psi' + \mathcal{H}\Phi = \beta v^{(\mathrm{gi})}, \tag{2.28b}$$

$$\Psi'' + \mathcal{H}(\Phi' + 2\Psi') + (\mathcal{H}^2 + 2\mathcal{H}')\Phi + \frac{1}{3}\nabla^2(\Phi - \Psi) = \frac{3}{2}l^2a^2\delta P^{(\mathrm{gi})}, \tag{2.28c}$$

$$\Psi - \Phi = 3l^2a^2P\Pi, \tag{2.28d}$$

where $\beta = \mathcal{H}^2 - \mathcal{H}' = \frac{3}{2}l^2a^2(\rho + P)$. While the energy-momentum conservation equations gained from the covariant conservation of the stress-energy tensor are not independent of the Einstein equations listed above, they are often useful and their gauge-invariant form is given by

$$\delta^{(\mathrm{gi})\prime} - (1+w)(\nabla^2 v^{(\mathrm{gi})} + 3\Psi') + 3\mathcal{H}\left(\frac{\delta P^{(\mathrm{gi})}}{\rho} - w\delta^{(\mathrm{gi})}\right) = 0, \tag{2.29a}$$

$$v^{(\mathrm{gi})\prime} + \mathcal{H}(1-3w)v^{(\mathrm{gi})} + \frac{w'}{1+w}v^{(\mathrm{gi})} - \frac{\delta P^{(\mathrm{gi})}}{\rho + P} - \Phi - \frac{2}{3}\frac{w}{1+w}\nabla^2\Pi = 0, \tag{2.29b}$$

where $\delta = \delta\rho/\rho$ is the density contrast.

The Einstein equations are very simple for the tensor perturbations and yield the sole equation

$$(h^{\mathrm{T}})^{i\prime\prime}_j + 2\mathcal{H}(h^{\mathrm{T}})^{i\prime}_j - \nabla^2(h^{\mathrm{T}})^i_j = 6l^2a^2P(\Pi^{\mathrm{T}})^i_j, \tag{2.30}$$

where $(\Pi^{\mathrm{T}})^i_j$ is the tensor part of the anisotropic stress tensor.

2.3.4 Adiabatic and Entropy Modes

Another useful decomposition of perturbations is the separation into adiabatic and entropic parts. This divides perturbations into adiabatic mode(s) with $(\delta\rho \neq 0, \delta s = 0)$ and entropy mode(s) with $(\delta\rho = 0, \delta s \neq 0)$. The pressure of a 'fluid-like' substance [19] in general is a function of both the energy and entropy densities so that

$$\delta P = \left.\frac{\partial P}{\partial \rho}\right|_s \delta\rho + \left.\frac{\partial P}{\partial s}\right|_\rho \delta s. \tag{2.31}$$

[19] By 'fluid-like' we are not necessarily referring to a perfect fluid, but to a substance whose stress-energy tensor can be parameterized by Eq. (2.21).

2.4 Linear Perturbations in Our Universe

Two of the most fruitful pursuits in modern cosmology have been the study of linear perturbations in the matter and in the radiation permeating our Universe, which are manifested in large-scale structure and anisoptropies in the CMB, respectively. In this section, we briefly describe the evolution of these perturbations.

Species capable of free-streaming can be described by the use of a set of multipole moments. For example, this decomposition can be done with the temperature field of the photons [26]. The evolution of the multipole moments can be found from Boltzmann equations (see Ref. [1, 26] for details). In the context of a fluid, the density, velocity, and anisotropy perturbations are associated with the first three moments of such a decomposition. Before recombination, the baryons and photons were tightly-coupled by Compton scattering, which suppresses higher moments of the photon's temperature field and thus the baryons and photons can be well described by a fluid. At early times, the quadrupole of the radiation is small due to the tight-coupling between the photons and baryons. After decoupling, radiation is a subdominant component in the Universe and so the quadrupole remains small. As such, we can safely neglect the quadrupole in many cases. An important consequence is that in such cases we have $\Psi \approx \Phi$.

We will now examine some of the basics of the evolution of the matter density contrast δ in the Newtonian gauge. At times past recombination, we must revert back to using the full set of multipole moments to describe the photon distribution. However, as these times are far into matter domination, the effect of the radiation on the matter distribution is negligible at this point. Consequently, the Einstein equations in Section 2.3.3 are sufficient for describing the matter perturbations during these times.

At late times, most modes of interest are inside the horizon. The Einstein equations in Section 2.3.3 imply that for these modes, the matter density contrast δ evolves as [1]

$$\frac{d^2\delta_{\mathbf{k}}}{da^2} + \left(\frac{d\ln H}{da} + \frac{3}{a}\right)\frac{d\delta_{\mathbf{k}}}{da} - \frac{3\Omega_{\mathrm{m}}}{2a^5(H/H_0)^2}\delta_{\mathbf{k}} = 0. \tag{2.32}$$

Note that for these late-time sub-horizon modes, the evolution of $\delta_{\mathbf{k}}$ is independent of $\mathbf{k}$. We can then separate the evolution of $\delta_{\mathbf{k}}$ into two regimes: an early scale-dependent evolution and a late scale-independent evolution. The expression for $\delta_{\mathbf{k}}$ at late times is most often expressed through its relation to the metric perturbation ϕ, which in the current limit from Eq. (2.28a) implies $-k^2\phi_k = (3/2)l^2a^2\rho_{\mathrm{m}}\delta_k$.

A transfer function $T(k)$ is used to describe the early scale-independent evolution and is defined as

$$T(k) = \frac{\phi(k, a_{\rm late})}{\phi(k_{\rm LS}, a_{\rm late})}, \tag{2.33}$$

where $a_{\rm late}$ is some late time well into the scale-independent regime. The transfer function is normalized so that it equals unity for some large-scale mode $k_{\rm LS}$. It can be shown from the Einstein equations without too much difficulty that for large-scale superhorizon modes, ϕ_k decreases by a factor of 9/10 from its primordial value [9]. Although the transfer function can be found analytically in small and large scale limits, expressions for the transfer function valid for both small and large scales are typically expressed as a fitting formula found numerically. Two of the most popular fitting formulas for the transfer function are that of Bardeen, Bond, Kaiser, and Szalay [27] and Eisenstein and Hu [28].

The growth function $G(a)$ parameterizes the late scale-independent evolution of ϕ and δ. It is defined as

$$G(a) = a\frac{\phi(a)}{\phi(a_{\rm late})}, \tag{2.34}$$

for $a > a_{\rm late}$.[20] The growth function can be found by solving Eq. (2.32) and only retaining the growing mode. With appropriate initial conditions, the growth function is found to be

$$G(a) = \frac{5}{2}\Omega_{\rm m}\frac{H(a)}{H_0}\int_0^a d\tilde{a}(\tilde{a}H(\tilde{a})/H_0)^{-3}. \tag{2.35}$$

The primary descriptive statistic of the matter density field is its two-point correlation function, or as more commonly used, its Fourier transform, the matter power spectrum. The last remaining piece before we write down the linear matter power spectrum is specifying the primordial power spectrum for ϕ. This is conventionally parameterized as [21]

$$P_{\phi,\rm I}(k) = \frac{50\pi^2}{9k^3}\delta_H^2(k/H_0)^{n_{\rm s}-1}\Omega_{\rm m}^2/G(a=1)^2. \tag{2.36}$$

The amplitude of the power spectrum is set by the parameter δ_H and its scale dependence is specified by the scalar spectral index $n_{\rm s}$. Most often

[20] The extra factor of a is added to the definition of the growth function so that $\delta \propto G(a)$.

[21] The form of this parameterization is chosen to simplify the expression for the matter power spectrum evaluated at the present.

n_s is taken to be independent of k and from observational constraints is slightly less than unity, while the amplitude is roughly $\delta_H \sim 10^{-5}$. With this, we now arrive at the expression for the linear matter power spectrum for $a > a_\mathrm{late}$

$$P_\mathrm{m}(k,a) = 2\pi^2 \delta_H^2 \frac{k^{n_\mathrm{s}}}{H_0^{n_\mathrm{s}+3}} T^2(k) \left(\frac{G(a)}{G(a=1)} \right)^2 . \tag{2.37}$$

A related statistic often employed in cosmology is the expectation value of the variance of the linear overdensity within a sphere of radius R, symbolized by $\sigma_R^2 = \langle \delta_R^2 \rangle$, where $\delta_R(\mathbf{x}) = \int d^3\tilde{x}\delta(\tilde{\mathbf{x}})W_R(\mathbf{x}-\tilde{\mathbf{x}})$ is the linear overdensity smoothed on the scale R with the top hat window function W_R, which in Fourier space is given by

$$W_R(k) = \frac{3(\sin(kR) - kR\cos(kR))}{(kR)^3}. \tag{2.38}$$

This variance can be written in terms of the the power spectrum $P(\mathbf{k})$ by

$$\sigma_R^2 = \int \frac{d^3k}{(2\pi)^3} P(\mathbf{k}) |W_R(k)|^2. \tag{2.39}$$

Its value at $R = 8\,\mathrm{Mpc}\,h^{-1}$, denoted by σ_8, is a frequently measured parameter (measured to by about $\sigma_8 \sim 0.8$ [25, 29]) and is often used to normalize the linear power spectrum.

2.5 Collapse into Nonlinear Structures

So far we have examined our Universe approximated as homogeneous and isotropic and then considered linear perturbations about the homogeneous and isotropic background. On large scales, one can go far with this model. On the other hand, on smaller scales the behaviour of the perturbations is highly nonlinear, as evident from the galaxies, stars, planets, and other astrophysical structures present in our Universe. In this section we give a short review of some simple but powerful models for describing the collapse of linear perturbations into nonlinear structures. In particular, we examine collapse into dark matter halos, which subsequently act as the breeding ground for galaxies. In this section, we assume that the dark matter is cold and will examine some of the effects of relaxing this assumption in Chapter 6.

2.5.1 Spherical Collapse

Before we are able to predict quantities like the abundances of collapsed structures, we must be able to track a perturbation from the linear to the nonlinear regime. To accomplish this, we aim to find the value of the overdensity predicted in linear theory when the full nonlinear perturbation has collapsed.

To start, we consider an isolated, spherical, and uniform overdensity of cold, pressureless matter. In this simple model, particles move in spherical shells without crossing one another until far into its collapse, after which the motion of the particles will be chaotic, eventually relaxing into a virialized state [5, 30]. We focus our attention on times when the Universe is matter dominated. As we are considering a region smaller than the horizon size, Newtonian dynamics should be reasonably accurate, so that a shell of matter a distance R away from the centre of the overdensity moves according to

$$\frac{d^2R}{dt^2} = -\frac{GM}{R^2} = -\frac{4}{3}\pi G\rho R, \tag{2.40}$$

where $M = (4/3)\pi\rho R^3$, until shell crossing occurs. Since in a matter-dominated universe we have $\rho \propto a^{-3}$, the full nonlinear overdensity $\delta_{\rm nl}$ is given by

$$\delta_{\rm nl} = \left(\frac{a(t)}{R(t)/R_0}\right)^3 - 1, \tag{2.41}$$

where R_0 is the initial size of the overdense region. By substituting $\delta_{\rm nl}$ into Eq. (2.40) and solving for the overdensity yields the following parametric solutions

$$R = GM(1-\cos\tau)C^{-1}, \tag{2.42a}$$

$$t = GM(\tau - \sin\tau)C^{-3/2}, \tag{2.42b}$$

$$\delta_{\rm nl} = \frac{9(\tau-\sin\tau)^2}{2(1-\cos\tau)^3} - 1, \tag{2.42c}$$

$$\delta = \frac{3}{5}\left(\frac{3}{4}(\tau-\sin\tau)\right)^{2/3}, \tag{2.42d}$$

where $\tau \in (0, 2\pi)$ is a parametric variable, C is an integration constant, and we have used the fact that $a \propto t^{2/3}$ in a matter-dominated universe.[22] In the above equations, δ is the solution for the overdensity by linearizing all equations in the overdensity, while $\delta_{\rm nl}$ is the solution without any such

[22] Similar expressions exist for the evolution of an underdense region.

approximations. From Eq. (2.42a), we can see that initially the size of the overdense region expands with the background until it reaches a turnaround point where the overdense region starts to collapse. This simplified model of collapse should be useful until late in the collapse, when significant shell crossing occurs.

We can now use the evolution equations for δ and $\delta_{\rm nl}$ given in Eqs. (2.42) to map the evolution of the linear perturbations to the full nonlinear behaviour (at least in this simplified case). From Eq. (2.42a), we can see that the turn around occurs when $\tau = \pi$ and collapse is complete when $\tau = 2\pi$. At final collapse, the linear overdensity is $\delta = \delta_{\rm c} \approx 1.69$, where $\delta_{\rm c}$ is referred to as the (linear) critical collapse threshold. Conveniently, we have found that in a matter dominated universe, the critical collapse threshold for spherical collapse is a constant. For the same model but in a universe with both matter and dark energy (with constant equation of state), a similar analysis can be done, but the collapse threshold now evolves with time [31]. The critical collapse threshold plays a central role in the Press-Schecther model, which will be the focus of the next section.

2.5.2 The Press-Schecther model

The Press-Schecther (PS) model [32] is a simple but powerful tool that predicts the abundance of dark matter halos, which has been relatively successful matching predictions to observations and simulations.[23] The PS model considers a Gaussian random (linear) density field that is considered collapsed into a halo when it reaches a critical collapse threshold.

More precisely, we would like to know when a region of size R and mass $M = (4\pi/3)\rho R^3$ collapses into a halo. To this end, we smooth the density contrast on a scale R to yield the field δ_M, which will have a variance $\sigma_M^2(z)$, given by Eq. (2.39).[24] Since the field is Gaussian, the fraction of collapsed regions (known as the collapse fraction) with mass M or above is simply

$$f_{\rm coll}(z) = 2\int_{\delta_{\rm c}}^{\infty} d\delta_M \frac{1}{\sqrt{2\pi}\sigma_M(z)} \exp\left(-\frac{\delta_M^2}{2\sigma_M^2({\rm z})}\right) = {\rm erfc}\left(\frac{\delta_{\rm c}}{\sqrt{2}\sigma_M({\rm z})}\right). \tag{2.43}$$

Above, a somewhat precarious factor of 2 was added whose inclusion was originally justified to allow for underdense regions with $\delta_M < 0$ to be incor-

[23] This agreement improves significantly with small modifications to the formulation, such as accounting for non-spherical collapse, as considered by Ref. [33].

[24] In general, we will write the smoothing scale in terms of the mass M within a region of size R instead of R itself.

porated into larger halos. This factor of 2 was later more rigorously justified and we will return to the matter in Section 2.5.3.

By differentiating the collapse fraction, we find that the mass function, the number density dn of halos with mass M between M and $M + dM$, is given by

$$\frac{dn}{dM} = -\frac{\rho}{M}\frac{d\ln\sigma}{dM}F(\nu), \tag{2.44}$$

where $\nu(z) = \delta_c/\sigma_M(z)$ and for the PS model $F(\nu)$ is

$$F_{\mathrm{PS}}(\nu) = \sqrt{\frac{2}{\pi}}\nu e^{-\nu^2/2}. \tag{2.45}$$

2.5.3 The Excursion Set Formalism

The Press-Schecther model of collapse provides a starting point for a powerful analysis tool known as the excursion set or extended Press-Schecther (EPS) formalism [34, 35]. By reframing the standard Press-Schecther model, the EPS formalism adds many useful extensions to the standard PS model, as well as providing new insights into our simple model of collapse.

To motivate the excursion set formalism, we first discuss a conceptual drawback of the standard PS theory. This drawback concerns how to properly form halo statistics to account for the situation when there is a smaller region below the collapse threshold (when the density field is smoothed on a smaller scale), which is contained in a larger region that is above the collapse threshold (when smoothed on a larger scale). One would expect that the smaller region would be amalgamated with the matter in the larger region into a collapsed structure [35]. This is known as the 'cloud-in-cloud' problem.

In light of the cloud-in-cloud problem, we slightly alter the objective of the standard PS method: We would like to find the largest smoothing scale where the smoothed density field exceeds the critical threshold. This is accomplished by starting at a very large smoothing scale, where $\delta_M \approx 0$, so that the probability of collapse at this scale is negligible, and decrease the smoothing scale until we find the first point where the smoothed density field exceeds the critical threshold. The largest scale that exceeds the critical threshold is marked as a collapsed halo and any smaller scales inside this region that surpasses the critical threshold is considered part of the larger halo.

As the smoothing scale decreases, more Fourier modes become relevant for the collapse and the probability of collapse increases. At this point, we

may do this process numerically with a particular realization of the density field. Alternatively, we may proceed to calculate halo statistics analytically using the probability distributions given in the problem, a description of which follows. We can imagine adding Fourier modes to the density field as the smoothing scale decreases, which, since we are considering a Gaussian random field where the modes are independent of one another, has the same statistics as a diffusion process. This amounts to the density contrast taking a random walk as the scale decreases, starting from a value of zero at large scales. The goal is to find the probability of the first 'up-crossing' through the critical threshold at a particular scale. If we choose the window function used in the smoothing to be a spherical top hat in k-space, each step in the random walk will be independent of one another, yielding a simple analytic solution. However, in Section 2.5.1 we calculated the collapse threshold assuming that our overdensity had the profile of a spherical top hat in real space. Thus using both the critical threshold as previously derived and the spherical top-hat smoothing window function in k-space is not fully consistent with one another. Fortunately, predictions using the aforementioned method match well to simulations and observations and using more self-consistent approaches seem to yield little improvement while making the analysis more cumbersome. In this light, we continue with the method stated above with less trepidation. With uncorrelated steps in our random walk, the expression for the collapse fraction can be found by calculating the fraction of random walks trajectories that remain below the critical threshold for all modes with k less than the the cut-off scale, set by M in our k-space window function. The resulting expression for the collapse fraction coincides with that of Eq. (2.43), including the addition of the factor of 2.

The excursion set formalism allows us to tackle many more problems, such as how halos accrete mass and merge over time, the length of time for formation, and many other similar questions.[25] We may now also calculate the spatial biasing of halos. Until now, we have only examined global quantities, but now we wish to determine halo statistics in a particular region of space with a finite size that encompasses a mass $\tilde{M}$ and has density contrast δ. This local collapse fraction, known as the biased collapse fraction, can be calculated in a similar manner as described above for the global collapse fraction, expect that instead of starting the random walk process at $\delta_M = 0$ and $M \to \infty$, we start from $\delta_M = \delta$ and $M = \tilde{M}$. The effect of this is to simply make the replacement $\delta_\mathrm{c} \to \delta_\mathrm{c} - \delta$ in Eq. (2.43). We can now find

[25] For a comprehensive review of these subjects, see Ref. [30].

the biased mass function dn/dM, which is a function of δ. For cases where δ is small, it is useful to expand the mass function as a Taylor series, which to linear order can be expressed as

$$\frac{dn}{dM}(\delta) = \frac{dn}{dM}\Big(1 + b(M)\delta\Big), \tag{2.46}$$

where b is referred to as the halo bias. For the PS mass function, the halo bias $b_{\rm PS}$ is given by

$$b_{\rm PS}(M) = 1 + \frac{\nu^2(M,z) - 1}{\delta_{\rm c}}. \tag{2.47}$$

2.5.4 Improvements to the Mass Function

The EPS formalism provides a simple but powerful analysis tool for examining the basic properties of halos. However, as formulated above, the PS mass function underestimates the number of high-mass halos and overestimates the number of low-mass halos, as compared to numerical simulations. One of the most successful extensions to the PS model is to allow for ellipsoidal collapse. The mass function of Sheth and Tormen [33] allows for such deviations from spherical collapse. Conveniently, the resulting mass function can still be expressed in terms of the critical collapse threshold $\delta_{\rm c}$ for spherical collapse and has a similar form as in the PS model. The Sheth-Tormen mass function can be expressed using Eq. (2.44), but where F is now given by

$$F_{\rm ST} = A\sqrt{\frac{2}{\pi}}\hat{\nu}(1 + \hat{\nu}^{-2p})e^{-\hat{\nu}^2/2}, \tag{2.48}$$

where $\hat{\nu} = \sqrt{a}\nu$ and A, a, and p are fitting parameters.

2.5.5 Halo Virialization

The ESP formalism has proved to be a valuable tool for statistically describing the collapse of matter into halos. However, if we would like to examine basic characteristics of the halo after significant shell-crossing has occured, we require new tools that can accommodate for the chaotic behaviour of the matter as it enters its final stages of collapse and its subsequent relaxation. In this section we describe how basic halo properties can be estimated by using the virial theorem.

For a self-gravitating system, the virial theorem relates the time-averaged kinetic and potential energy, K and U, respectively, by $U = -2K$. Assuming conservation of energy within the system, the energy of the system

at turnaround, given simply by U at this time, will equal the energy of the relaxed system $U_{\rm vir} + K_{\rm vir} = U_{\rm vir}/2$. Since for a spherical overdensity $U = -(3/5)GM/r$, the (physical) virial radius $r_{\rm vir}$ will be half the value of the radius at turnaround and the volume at virialization will decrease by a factor of 8 compared to that at turnaround. We can approximate the time of virialization as occurring when the overdensity would collapse completely according to Eqs. (2.42) (at $\tau = 2\pi$). Since $a \propto t^{2/3}$ in a matter-dominated universe, from Eq. (2.42b) we see that a expands by a factor of $2^{2/3}$ between turnaround and virilization, and consequently $\rho_{\rm cr}$ will decrease by a factor of 4 during this time. Putting all of these factors together, we approximate the ratio $\Delta_{\rm c} = \rho_{\rm vir}/\bar{\rho}_{\rm cr}$ of the density of the virialized halo $\rho_{\rm vir}$ to the critical density at virialization as $\Delta_{\rm c} = 32[1 + \delta_{\rm nl}(\tau = \pi)] = 18\pi^2$. This result can be generalized to a flat universe with both matter and cosmological constant ($\Omega_{\rm m} + \Omega_\Lambda = 1$) with the fitting formula [36]

$$\Delta_{\rm c} = 18\pi^2 + 82d - 39d^2, \tag{2.49}$$

where $d \equiv \Omega_{\rm m}^z - 1$ and $\Omega_{\rm m}^z$ is the matter density parameter at redshift z, which in this case is given by

$$\Omega_{\rm m}^z = \frac{\Omega_{\rm m}(1+z)^3}{\Omega_{\rm m}(1+z)^3 + \Omega_\Lambda}. \tag{2.50}$$

We can now write the the virialized radius as

$$r_{\rm vir} = 1.49 \left(\frac{h}{0.7}\right)^{-2/3} \left(\frac{\Omega_{\rm m}}{0.3}\right)^{-1/3} \left(\frac{1}{\Omega_{\rm m}^z}\frac{\Delta_c}{18\pi^2}\right)^{-1/3} \times \left(\frac{1+z}{10}\right)^{-1} \left(\frac{M}{10^8\,{\rm M}_\odot}\right)^{1/3} {\rm kpc} \tag{2.51}$$

and can subsequently find the corresponding circular velocity $V_{\rm c} = \sqrt{GM/r_{\rm vir}}$ and can define the virial temperature as $T_{\rm vir} = \mu m_p V_{\rm c}^2/2k_{\rm b}$, where μ is the mean molecular weight and m_p is the proton mass. The halo mass as a function of its virial temperature can then be written as

$$M = 9.37 \times 10^7 \left(\frac{\mu}{0.6}\right)^{-3/2} \left(\frac{h}{0.7}\right)^{-1} \left(\frac{\Omega_{\rm m}}{0.3}\right)^{-1/2} \times \left(\frac{1}{\Omega_{\rm m}^z}\frac{\Delta_{\rm c}}{18\pi^2}\right)^{-1/2} \left(\frac{1+z}{10}\right)^{-3/2} \left(\frac{T_{\rm vir}}{10^4\,{\rm K}}\right)^{3/2} {\rm M}_\odot. \tag{2.52}$$

Chapter 3

A Brief Tour Through Cosmological Inflation

3.1 Introduction

From the start of modern cosmology in the early 20th century through the 1960s, a standard cosmological model emerged that described the expansion of our Universe and its basic constituents. However, beginning in the 1970s, puzzling questions arose that made this standard model seem incongruent with observations. Among others, one such question was why the CMB was so isotropic in spite of the fact that, according to the prevailing cosmological model of the time, many CMB photons coming from different directions would have originated from locations what were not yet in causal contact with one another. Alan Guth proposed the theory of inflation in 1980 as a solution to these problems [37], later developed by Linde [38], Albrecht and Steinhardt [39], among others. It was later realized that inflation also provided a mechanism that seeds the perturbations in our Universe, which later imprinted themselves as anisotropies in the CMB and sourced large-scale structures. In this chapter, we give a brief introduction to inflationary theory.

3.2 Problems with the Standard Cosmological Model

The problems alluded to in the previous section all in some way deal with the initial conditions set in the early Universe. Here we outline these problems.

The Horizon Problem

We have already briefly touched on the horizon problem, one view of which asks why the Universe is isotropic to such a high degree. We can formulate this problem more precisely by comparing the comoving size of the present-

day observable Universe to that of a causal patch at some early time. In the standard cosmological model, where the total equation of state w lies between 0 and 1/3 throughout, the comoving particle horizon at time t is of the order of $\mathcal{H}^{-1}(t)$. We can then estimate this ratio by

$$\frac{\mathcal{H}_0^{-1}}{\mathcal{H}_\mathrm{i}^{-1}} = \frac{a_\mathrm{i} H_\mathrm{i}}{a_0 H_0} \sim 10^{28} \frac{T_\mathrm{i}}{m_\mathrm{p}}, \tag{3.1}$$

where the subscript i denotes quantities evaluated at some early 'initial' time t_i and have assumed that the Universe is radiation dominated at this time with temperature T_i. If t_i is near the Planck scale, then the comoving length scale of the present-day observable Universe was about 10^{28} times bigger than that of causal regions at t_i, so that the present-day observable Universe encloses 10^{84} different regions that were causally disconnected from one another at t_i. Decreasing T_i doesn't help the situation much; for T_i at the GeV scale, the present-day horizon volume would still encompass around 10^{30} causally disconnected regions. With this in mind, it is unusual that our Universe would be so isotropic as well as homogenous on large scales if the regions that comprise the present-day horizon volume were not in causal contact with one another at some time in the distant past.

The Flatness Problem

The present-day Universe is very flat, in the sense that Ω_k is currently bounded by roughly $|\Omega_\mathrm{k}| < 0.04$. However, a fine-tuning problem arises with the realization that in the standard cosmological model Ω_k increases with time, such that Ω_k must have been initially fine-tuned to an extremely small value. We can compare the present-day value of Ω_k to that at t_i by the fraction

$$\frac{\Omega_\mathrm{k}(t_0)}{\Omega_\mathrm{k}(t_\mathrm{i})} = \left(\frac{H_\mathrm{i} a_\mathrm{i}}{H_0 a_0}\right)^2 \sim 10^{56} \left(\frac{T_\mathrm{i}}{m_\mathrm{p}}\right)^2, \tag{3.2}$$

again assuming that the Universe is dominated by radiation at t_i. Thus, $\Omega_\mathrm{k}(t_\mathrm{i})$ must be fine-tuned to an extremely small value at the Planck scale, when one might expect it to be of order unity at this time.

3.3 The Basics

All of the aforementioned problems in some way deal with the horizon size in the very early Universe. The problems stem from the fact that in the

standard cosmological model, the particle horizon[26] is of the same order of magnitude as the Hubble length, which can be thought of as the length scale over which particles can communicate with one another at a certain time (within the time that the scale factor grows by a factor of e). From Eq. (2.5), we can see that if w is bounded by 0 and 1/3 throughout, as in the standard cosmological model, the comoving Hubble length $\mathcal{H}^{-1}$ monotonically increases with η, and η and $\mathcal{H}^{-1}$ are of the same order of magnitude. In other words, in the standard model, when a scale enters the horizon, it is the first time there can be causal contact on this scale.

Inflation resolves these issues by creating a large difference between the Hubble length and particle horizon. Unlike the particle horizon, which increases monotonically (this is why we can use η as a time variable), the Hubble length can decrease. Examining Eq. (2.5) again, if $w < -1/3$ then $\mathcal{H}^{-1}$ would decrease as η increases, so that a sufficiently long stage with such an equation of state would create a drastic difference between $\mathcal{H}^{-1}$ and η. During inflation, the comoving Hubble length 'zooms in' to a much smaller scale than at the start of inflation. Spacetimes with $-1 < w < -1/3$ have an event horizon [27], whose comoving length is of order $\mathcal{H}^{-1}$. A (comoving) scale k 'leaves the horizon' when $k \sim \mathcal{H}$; communication on this scale is possible before this time but not after. After its rapid decrease during inflation, $\mathcal{H}^{-1}$ begins to grow again as the subsequent evolution of the Universe proceeds as described in the standard cosmological model. With inflation, when a scale reenters the horizon well after inflation, the particle horizon is many orders of magnitude larger than the Hubble length and so although communication can only commence on this scale once it enters the horizon, communication could have taken place on this scale well before this time (i.e. before it left the horizon during inflation). Standard inflationary models assume that the inflationary spacetime is nearly de Sitter (w is close to -1) so that H is nearly constant during inflation. The parallel view in terms of physical scales has rapidly growing physical scales during inflation pass through a nearly constant Hubble length H^{-1}.

We can now revisit the problems discussed in Section 3.2 with an inflationary period assumed to have occurred in the very early Universe. As before, to fit the present-day observable Universe into a causal region of space during an 'initial' time t_{i}, we require $\mathcal{H}_0^{-1} \leq \mathcal{H}_{\mathrm{i}}^{-1}$, but in the inflation-

[26] The particle horizon is the maximum distance from which particles could have travelled to an observer at a particular time over the entire history of the Universe until the observation time.

[27] Technically speaking, there is only a true event horizon if the equation of state w continues to stay below $-1/3$.

ary paradigm t_i is *before* inflation. This ratio is now

$$\frac{\mathcal{H}_0^{-1}}{\mathcal{H}_\mathrm{i}^{-1}} = \frac{a_\mathrm{i}}{a_\mathrm{e}}\frac{a_\mathrm{e}}{a_0}\frac{H_\mathrm{i}}{H_0} = e^{-N}\frac{a_\mathrm{e}}{a_0}\frac{H_\mathrm{i}}{H_0} \sim e^{-N}10^{28}\frac{T_\mathrm{e}}{m_\mathrm{p}}, \tag{3.3}$$

where a_e is the scale factor at the end of inflation and we have parameterized the duration of inflation by the number of e-folds $N \equiv \ln(a_\mathrm{e}/a_\mathrm{i})$.[28] In the last step in Eq. (3.3), we have assumed that H is approximately constant throughout inflation. If inflation occurs near the Planck scale, then the requirement of $\mathcal{H}_0^{-1} \leq \mathcal{H}_\mathrm{i}^{-1}$ necessitates at least $N \sim 64$ e-folds of inflation.[29] It is easy to see that the requirement of $\mathcal{H}_0^{-1} \leq \mathcal{H}_\mathrm{i}^{-1}$ solves both the horizon and flatness problems.

3.4 A Simple Model

From Eq. (2.3b), we see that having $-1 < w < -1/3$ results in an accelerating background. The next step is to find what substances are capable of driving an accelerating expansion. Here we examine the simple case of a single scalar field φ with potential $V(\varphi)$ [2, 3, 9]. Its Lagrangian is given by

$$\mathcal{L} = \frac{1}{2}\partial^\alpha\varphi\partial_\alpha\varphi - V(\varphi). \tag{3.4}$$

The stress-energy tensor $T^\mu_\nu = (\partial\mathcal{L}/\partial(\partial_\mu\varphi))\partial_\nu\varphi - \mathcal{L}\delta^\mu_\nu$ in this case is then

$$T^\mu_\nu = \partial^\mu\varphi\partial_\nu\varphi - \left[\frac{1}{2}\partial^\alpha\varphi\partial_\alpha\varphi - V(\varphi)\right]\delta^\mu_\nu. \tag{3.5}$$

At this point, we decompose our field as $\varphi(\mathbf{x},t) = \bar{\varphi}(t) + \delta\varphi(\mathbf{x},t)$, assuming the homogenous part $\bar{\varphi}$ is much larger than $\delta\varphi$, which is treated as a perturbation. Parameterizing the stress-energy tensor as in Section 2.1.1, we can identify the background energy density ρ and pressure P as

$$\rho = \frac{1}{2}a^{-2}\bar{\varphi}'^2 + V, \qquad P = \frac{1}{2}a^{-2}\bar{\varphi}'^2 - V, \tag{3.6}$$

and note that all off-diagonal terms in $\bar{T}^\mu_\nu$ are zero. The equation of state for the field is then

$$w = \frac{\frac{1}{2}a^{-2}\bar{\varphi}'^2 - V}{\frac{1}{2}a^{-2}\bar{\varphi}'^2 + V}. \tag{3.7}$$

[28] See Section 4.6.2 for a more rigorous expression for Eq. (3.3).
[29] See Section 4.6.2 for some caveats.

We can now see that $w \approx -1$ if we have $\bar{\varphi}' \ll V$. To formalize this approximation, we introduce the 'slow-roll' variables

$$\epsilon = -\frac{\dot{H}}{H^2}, \qquad \tau = -\frac{\ddot{\varphi}}{H\dot{\varphi}}. \tag{3.8}$$

When $\epsilon, \tau \ll 1$, the field φ is said to be in the slow-roll regime and inflation ends when one of these conditions is violated. When $\epsilon \ll 1$, the equation of state is approximately $w \approx -1 + \frac{2}{3}\epsilon$ and thus the background expands near the de Sitter solution. In this case, the scale factor and conformal Hubble rate approximately evolve as

$$a \propto (-\eta)^{-(1+\epsilon)}, \qquad \mathcal{H} = -\frac{1+\epsilon}{\eta}, \tag{3.9}$$

where the conformal time is bounded by $-\infty < \eta < 0$.

The evolution of the field φ can be found using the Klein-Gordon equation, which for the background field implies

$$\bar{\varphi}'' + 3\mathcal{H}\bar{\varphi}' + a^2 V_{,\varphi} = 0 \tag{3.10}$$

and from Eq. (2.3a), the Hubble rate in this case is given by [30]

$$H^2 = \frac{8\pi G}{3}\left[\frac{1}{2}a^{-2}\bar{\varphi}'^2 + V\right]. \tag{3.11}$$

Using these equations we can find the values of the slow-roll variables for a particular inflationary potential. For example, for power-law inflation where the scale factor evolves as $a \propto t^p$ and $p > 1$, we have $\epsilon = \tau = p^{-1}$.

3.5 End of Inflation and Reheating

At some point, inflation must end and produce the standard model particles that comprise the present-day visible Universe. In terms of the slow-roll variables introduced in the previous section, inflation ends when at least one of the slow-roll variables attains a value comparable to unity. After this point, standard model particles must be produced in some manner, usually through the decay of the inflating substance, and then thermalize in a process known as reheating. The standard hot big bang cosmological model then commences after the end of reheating. The thermalized temperature attained by the produced particles is referred to as the reheat temperature T_{RH}, which must be at least $T_{\mathrm{RH}} \gtrsim 5 - 10\,\mathrm{MeV}$ in order for BBN to proceed successfully [3].

[30] We have assumed a flat background here, which should be valid once inflation has lasted long enough to drive the background to a nearly flat state.

3.6 Generation of Perturbations

A key feature of inflation is that it generates the perturbations that, for example, source the anisotropies in the CMB and give the initial conditions for large-scale structure formation. While inflation wipes away pre-inflationary features, the accelerating background quantum-mechanically excites perturbations that get stretched to superhorizon scales.

3.6.1 Quantization

Following Ref. [9], our prescription for properly quantizing the perturbations that arise during inflation will be to write the inflationary action in the form of a harmonic oscillator so we can use the same well-known quantization procedure as used for the harmonic oscillator. We can write the total action as $S = S_{\rm gr} + S_{\rm m}$, where $S_{\rm gr} \propto \int \mathcal{R}\sqrt{-g}d^4x$ is the GR action, with g ithe determinant of the metric, and $S_{\rm m}$ is the action for the matter present in the inflationary spacetime. As we wish to recover equations of motion for our perturbations to linear order, we expand the action to second order in the perturbations variables. For definiteness, we continue to examine the inflationary scenario of a single scalar field.

As per usual, we examine the quantization of scalar and tensor perturbations separately and forgo examining the vector perturbations, as they decay very rapidly. We start with the scalar perturbations. Obtaining the second order scalar parts of $S_{\rm gr}$ and $S_{\rm m}$ is straightforward, yet very tedious, so we will simply quote the result for the second order terms in the total action $\delta_2 S$, which is [9]

$$\delta_2 S = \frac{1}{2}\int \left[u'^2 - u_{,i}^2 - m_{\rm eff,S}^2 u^2\right] d^4x, \tag{3.12}$$

where δ_2 denotes the second order terms. The gauge-invariant variable u, commonly known as the Mukhanov-Sasaki variable, is given by $u = a(\delta\varphi + (\bar{\varphi}'/\mathcal{H})\psi)$ and will be our canonical variable for the quantization procedure. The action in Eq. (3.12) was written in a suggestive way by defining an effective mass term, given by

$$m_{\rm eff,S}^2 = -\frac{z''}{z}, \tag{3.13}$$

where $z = a\bar{\varphi}'/\mathcal{H}$, as to make the analogy with a harmonic oscillator more apparent. In general, the effective mass will be a function of time.

Turning now to the tensor perturbations, the Ricci scalar $\mathcal{R}$ can be found for the tensor perturbations $(h^{\mathrm{T}})^i_j$ without much difficulty, yielding

$$\delta_2 S_{\mathrm{gr}} = \frac{1}{24l^2}\int a^2\left[(h^{\mathrm{T}})^{i\prime}_j(h^{\mathrm{T}})^{j\prime}_i - (h^{\mathrm{T}})^i_{j,l}(h^{\mathrm{T}})^{j,l}_i\right]d^4x \tag{3.14}$$

for the gravitation part of the action, where $l = \sqrt{8\pi G/3}$ is the Planck length. The second order tensor part of the action for a single scalar field vanishes. By expressing h^i_j in terms of the individual polarization states h_p, where $(h^{\mathrm{T}})^i_j(h^{\mathrm{T}})^j_i = 2\sum_p (h^{\mathrm{T}}_p)^2$, the total tensor part of the action can be written as

$$\delta_2 S = \frac{1}{2}\int\left[U_p'^2 - U_{p,i}^2 - m^2_{\mathrm{eff,T}}U_p^2\right]d^4x, \tag{3.15}$$

where we have defined the canonical variable $U_p = ah_p/\sqrt{6l^2}$ for each polarization state. The effective mass for the tensors is given by

$$m^2_{\mathrm{eff,T}} = -\frac{a''}{a}. \tag{3.16}$$

We see that the actions for the scalar and tensor modes are very similar, with only the effective mass for each perturbation type being slightly different from one another. We continue our discussion using the notation for the scalar modes (except we will drop the subscript S on m^2_{eff}), with the tensor case found by a trivial change of variables.

Varying the action with respect to u yields the equation of motion

$$u'' - \nabla^2 u + m^2_{\mathrm{eff}}u = 0, \tag{3.17}$$

which can readily be identified as having the form the action for a harmonic oscillator with time-varying mass.

The next step is to identify the conjugate momentum π to u, found to be

$$\pi = \frac{\partial\mathcal{L}}{\partial u'} = u'. \tag{3.18}$$

We can now promote u and π to operators $\hat{u}$ and $\hat{\pi}$ and impose the standard commutation relations

$$\begin{aligned}[\hat{u}(\mathbf{x},\eta),\hat{u}(\tilde{\mathbf{x}},\eta)] &= [\hat{\pi}(\mathbf{x},\eta),\hat{\pi}(\tilde{\mathbf{x}},\eta)] = 0\\ [\hat{u}(\mathbf{x},\eta),\hat{\pi}(\tilde{\mathbf{x}},\eta)] &= i\delta(\mathbf{x}-\tilde{\mathbf{x}})\end{aligned} \tag{3.19}$$

We can write $\hat{u}(\mathbf{x},\eta)$ in terms of the creation and annihilation operators $\hat{a}^\dagger_{\mathbf{k}}$ and $\hat{a}_{\mathbf{k}}$ for a mode $\mathbf{k}$ as [31]

$$\hat{u}(\mathbf{x},\eta) = \frac{1}{\sqrt{2}}\int \frac{d^3k}{(2\pi)^{3/2}}\left[\hat{a}_{\mathbf{k}}\chi^*_{\mathbf{k}}(\eta)e^{i\mathbf{k}\cdot\mathbf{x}} + \hat{a}^\dagger_{\mathbf{k}}\chi_{\mathbf{k}}(\eta)e^{-i\mathbf{k}\cdot\mathbf{x}}\right]. \tag{3.20}$$

The mode function $\chi_{\mathbf{k}}(\eta)$ obeys

$$\chi''_{\mathbf{k}} + \left[k^2 + m^2_{\text{eff}}\right]\chi_{\mathbf{k}} = 0 \tag{3.21}$$

and the commutation relations in Eq. (3.19) imply the normalization

$$\chi'_{\mathbf{k}}\chi^*_{\mathbf{k}} - \chi_{\mathbf{k}}\chi^{*\prime}_{\mathbf{k}} = 2i. \tag{3.22}$$

With this normalization condition, the second order differential equation for the mode function given in Eq. (3.21) has one remaining integration constant, the choice of which determines the effect of the creation and annihilation operators on a physical state, which we now explore in more detail. Before proceeding any further, we remark that when quantizing fields in framework of general relativity, the notion of a state, in particular the vacuum state, is dependent on one's frame of reference [40, 41, 42, 43]. Indeed, two different observers can each define a set of mode functions that defines the vacuum state within their frame of reference, but these will not necessarily coincide with the vacuum state defined in the other's frame of reference (see [44, 45] for more details).

In the present context of an expanding FLRW spacetime, this can be seen by defining a vacuum state $|0\rangle$ for a particular annihilation operator $\hat{a}_{\mathbf{k}}$ which yields the minimum energy eigenvalue of the Hamiltonian

$$\hat{H} = \frac{1}{2}\int d^3k\left[\hat{\pi}_{\mathbf{k}}\hat{\pi}_{\text{-}\mathbf{k}} + \omega^2_k\hat{u}_{\mathbf{k}}\hat{u}_{\text{-}\mathbf{k}}\right] \tag{3.23}$$

at a *particular* time, where $\omega^2_k(\eta) = k^2 + m^2_{\text{eff}}(\eta)$. The last point is significant in that it is not guaranteed that the mode function will evolve in such a way that this state will be the state of lowest energy at a later time. Acting with the Hamiltonian on this state yields the energy density [44]

$$\rho = \frac{1}{4}\int d^3k(|\chi'_{\mathbf{k}}|^2 + \omega^2_k|\chi_{\mathbf{k}}|^2). \tag{3.24}$$

At a particular time η_p, for $\omega^2_k(\eta_p) > 0$ the mode function

$$\chi_{\mathbf{k}}(\eta_p) = \frac{1}{\sqrt{\omega_k(\eta_p)}}e^{ik\eta_p}, \qquad \chi'_{\mathbf{k}}(\eta_p) = i\sqrt{\omega_k(\eta_p)}e^{ik\eta_p} \tag{3.25}$$

[31] Here we use the Fourier conventions $f(\mathbf{x}) = (2\pi)^{-3/2}\int d^3k f_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}}$.

minimizes Eq. (3.24) at $\eta = \eta_p$.[32] However, if at a later time we have $\omega_k^2(\eta_p) < 0$ then not only is this state no longer the minimum energy state, but there is not even a clearly defined minimum energy state for such a mode at this time. If the mode function evolves in this manner, then modes defined to be in the vacuum state when well within the horizon ($k \gg \mathcal{H}$) will be in an excited state when well outside the horizon ($k \ll \mathcal{H}$).

Since it is not clear how to use our quantization procedure if one cannot clearly define a vacuum state, we assume that all modes of interest are well within the horizon at some 'initial' time. We choose the mode function such that it selects the *instantaneous* vacuum state when the mode is well within the horizon. From Eq. (3.25), the mode function that minimizes the energy density when well within the horizon at time η_{i} is given by

$$\chi_{\mathbf{k}}(\eta_{\mathrm{i}}) \approx \frac{1}{\sqrt{k}} e^{ik\eta_{\mathrm{i}}}, \qquad \chi'_{\mathbf{k}}(\eta_{\mathrm{i}}) \approx i\sqrt{k} e^{ik\eta_{\mathrm{i}}}. \tag{3.26}$$

The state selected by this initial condition is known as the Bunch-Davis vacuum [46], and approximates the Minkowski vacuum when the mode is well within the horizon.

The final step for determining the mode function is to specify m_{eff}. For most cases of interest, m_{eff} will be proportional to η^{-2}. For example, for a single scalar field in the slow-roll regime characterized by the slow-roll variables in Eq. (3.8), the scalar effective massive term is $m_{\mathrm{eff,S}} \approx -(2 + 6\epsilon - 3\tau)/\eta^2$. For $m_{\mathrm{eff}} = (\frac{1}{4} - \nu^2)/\eta^2$ with ν constant, from Eq. (3.21) the mode function is found to be

$$\chi_{\mathbf{k}} = \sqrt{\frac{\pi|\eta|}{2}} \left[C_1 H_\nu^{(1)}(k|\eta|) + C_2 H_\nu^{(2)}(k|\eta|) \right], \tag{3.27}$$

where $H_\nu^{(1)}$ and $H_\nu^{(2)}$ are the Hankel functions of the first and second kind and C_1 and C_2 are integration constants. The Bunch-Davis vacuum is selected by choosing $C_1 = 0, C_2 = 1$, so that the mode function is

$$\chi_{\mathbf{k}} = \sqrt{\frac{\pi|\eta|}{2}} H_\nu^{(2)}(k|\eta|). \tag{3.28}$$

3.6.2 Beyond the Horizon

Now that we have the mode function, we can form the two-point correlation function $\langle 0|\hat{u}(\mathbf{x},\eta)\hat{u}(\tilde{\mathbf{x}},\eta)|0\rangle$ with the Bunch-Davis vacuum state, which can

[32] Any phase of $\chi_{\mathbf{k}}$ will minimize the energy density. We set the phase as $e^{ik\eta}$ for later convenience.

be written in terms of the mode function as

$$\langle 0|\hat{u}(\mathbf{x},\eta)\hat{u}(\tilde{\mathbf{x}},\eta)|0\rangle = \int dk \frac{k^2}{4\pi^2}|\chi_k(\eta)|^2 \frac{\sin(k|\mathbf{x}-\tilde{\mathbf{x}}|)}{k|\mathbf{x}-\tilde{\mathbf{x}}|}. \tag{3.29}$$

The power spectrum $\mathcal{P}_u(k) = \frac{k^3}{4\pi^2}|\chi_k|^2$ for u is then

$$\mathcal{P}_u(k) = \frac{k^3}{8\pi}|\eta|\left|H_\nu^{(2)}(k|\eta|)\right|^2. \tag{3.30}$$

In general, we are interested in modes that are on superhorizon scales at the end of inflation. Although the details of the end of inflation and reheating may be very complicated, we can circumvent the need to know these details by introducing a few new perturbation variables. We define the gauge-invariant variable R as

$$R \equiv \mathcal{H}v + \psi, \tag{3.31}$$

which is related to the scalar canonical variable by $u = zR$. This can be interpreted as the curvature perturbation in a comoving gauge or the velocity perturbation on uniform curvature hypersurfaces. Now introduce the gauge-invariant variable ζ, defined by

$$\zeta = \psi + \mathcal{H}\frac{\delta\rho}{\rho'}, \tag{3.32}$$

which can be interpreted as the either the curvature perturbation on uniform density hypersurfaces or the density perturbation on uniform curvature hypersurfaces. From the Einstein equation Eq. (2.28a), R and ζ can be related by

$$3\beta(R-\zeta) = \nabla^2\Psi. \tag{3.33}$$

In many situations, we can neglect the effect of spatial derivatives when a mode is on superhorizon scales, in which case $R_k \approx \zeta_k$ for a superhorizon mode k.[33] By using the Einstein equation Eq. (2.28b) and momentum conservation equation in Eq. (2.29b) in the comoving gauge, we can readily arrive at the equation

$$\frac{R'}{\mathcal{H}} = \frac{1}{1+w}\frac{\delta P_{\rm nad}}{\rho} + 3\frac{\partial P}{\partial\rho}(R-\zeta) + \frac{2}{3}\frac{w}{1+w}\nabla^2\Pi \tag{3.34}$$

for the evolution of R. We can now see that in the absence of nonadiabatic pressure and anisotropic stress, if we have $R_k \approx \zeta_k$ on superhorizon scales,

[33] See Section 4.8 for a case where this is not a valid approximation.

then R_k (and therefore ζ_k) will be approximately constant on superhorizon scales. Therefore, as long as these conditions hold, by keeping track of the superhorizon values of R (or equivalently ζ) during inflation, we do not require the details of reheating to track a mode into the radiation dominated era.

In this light, the power spectrum $\mathcal{P}_R$ for R for modes that are on superhorizon scales at the end of inflation is of particular interest. $\mathcal{P}_R$ is most often described by its value at the pivot scale $k_p = 0.002\,\mathrm{Mpc}^{-1}$ and by its spectral index $n_\mathrm{s} = 1 + d\ln\mathcal{P}_R/d\ln k$ evaluated at some scale (usually k_p). A scale-invariant scalar spectrum corresponds to $n_\mathrm{s} = 1$. When a mode is on superhorizon scales, we can use the small argument approximation for the Hankel function so that $\chi_k \propto k^{-\nu}$ and $n_\mathrm{s} = 4 - 2\nu$. Using the slow-roll variables, the scalar spectral index is expressed as $n_\mathrm{s} = 1 - 4\epsilon + 2\tau$. We can see that slow-roll inflation produces a nearly scale-invariant scalar spectrum (consistent with observations), which arises because the conditions when a mode leaves the horizon are approximately the same for all modes of interest.

By use of Eq. (2.30), it can easily be shown that for the same conditions as described above, the tensor perturbations $(h^\mathrm{T})^i_j$ remain approximately constant on superhorizon scales. The power spectrum $\mathcal{P}_\mathrm{T}$ for h^T can be defined in the same manner as done for the scalar perturbations, with a shape characterized by the spectral index $n_\mathrm{T} = d\ln\mathcal{P}_\mathrm{T}/d\ln k$. For inflation driven by a single scalar field in slow-roll, the tensor spectral index is given by $n_\mathrm{T} = -2\epsilon$. Of prime interest is the ratio between scalar and tensor modes $r = \mathcal{P}_\mathrm{T}/\mathcal{P}_R$.

We have finally arrived at the primordial spectrum for the scalar and tensor perturbations (at least for the simple model described in Section 3.4). With this information, we can continue to follow the perturbations into later times using the tools described in Chapter 2.

Chapter 4

Inflation with an Elastic Solid

4.1 Introduction

The inflationary paradigm, the existence of a brief period of accelerated expansion in the early Universe, provides an explanation for the observed homogeneity, isotropy and flatness of the Universe [37, 38, 39]. On large scales it successfully accounts for the distribution of fluctuations seen in the cosmic microwave background (CMB) and the large-scale structure of the Universe. Inflation is often modelled in terms of a scalar field slowly evolving in its potential. Yet, the physical model of inflation is not known and even within the context of scalar fields many models are compatible with current observations. It is worthwhile exploring whether or not other physical frameworks, more general than a scalar field, can successfully account for a period of inflation.

In this chapter, we build a model of inflation that describes the substance that drives inflation by a continuous medium that can be characterized by its macroscopic properties. The simplest model of a continuous medium in general relativity is a perfect fluid. To drive an accelerating expansion, the medium must have an equation of state $w \equiv P/\rho < -1/3$, where ρ and P are the energy density and pressure of the fluid, respectively. However, a perfect fluid with constant w has a sound speed for longitudinal (density) waves of $c_{\rm s} = \sqrt{w}$, so demanding that the fluid drive an accelerated expansion formally results in an imaginary sound speed and an instability to small perturbations.

One generalization of a perfect fluid is a relativistic elastic solid. Elastic solids have a rigidity, and so can support both longitudinal and transverse waves. An elastic solid (both relativistic and nonrelativistic) can be characterized by a bulk modulus κ that depends on the equation of state w, and shear modulus μ that determines how rigid the solid is [47]. As in the nonrelativistic case, the longitudinal sound speed $c_{\rm s}$ depends upon both κ

and μ, while the transverse sound speed $c_{\rm v}$ only depends upon μ. In the relativistic case, a sufficiently rigid elastic solid can result in a real longitudinal sound speed $c_{\rm s}$, even in cases where w is negative enough to drive acceleration.

In this chapter we describe a model of a homogeneous and isotropic elastic solid coupled to general relativity. This model has previously been considered as a potential model of dark energy [48, 49] and recently similarities between a relativistic elastic solid and massive gravity have been noted [50].

In this work, we discuss in detail how an elastic solid can drive an inflationary epoch in the early Universe.[34] Linear perturbations in an elastic solid satisfy the equations of motion found in Ref. [49], which uses the framework for describing a macroscopic relativistic medium developed in Ref. [54]. We develop the quadratic action for a generic elastic medium, quantize the linear modes that are excited during inflation, and determine the spectra of scalar and tensor modes produced by an inflationary stage driven by an elastic solid.

A novel feature of this model is that, in contrast to what typically occurs when the Universe is dominated by a single substance, the anisotropic stress of the solid causes modes to evolve on superhorizon scales. As such, the final spectrum of superhorizon modes is sensitive to the manner in which inflation ends. We show here that the case where the sound speeds and equation of state are perfectly constant results in a blue-tilted scalar power spectrum, but if these quantities vary slowly in time then a red-tilted scalar power

[34] The notion that a relativistic elastic solid could drive inflation was first discussed in Ref. [51] and more recently in Refs. [52, 53]. The present work includes an in-depth treatment of the quantization of the scalar and tensor linear perturbations and their superhorizon evolution. Furthermore, new states of the elastic solid are found in which its equation of state is far from the fiducial value of -1 that nonetheless produces nearly scale-invariant spectra for the linear perturbations. In comparison to the aforementioned references, our work uses a different approach and treats the problem starting directly from the quadratic action for an elastic solid and includes an extended treatment of superhorizon evolution and reheating. The work of Ref. [52] uses an effective field theory approach that involves the presence of three scalar fields, which if certain symmetries are applied gives the same physical behaviour as an elastic solid. Although the effective field theory approach used in Ref. [52] is distinct from the analysis given here, the same equations of motion are reached for both models for cases where w and $c_{\rm s}$ evolve slowly near -1 and 0, respectively. Specifically, the equations of motion in Ref. [52] can be recovered using those found in this chapter by setting $c_{\rm s0} = \epsilon_0 = 0$ and making the notational substitutions $c_{\rm s,1} \to c_{L,c}$, $\epsilon_1 \to \epsilon_c$, $\tau_{\rm s} \to s_c$, and $\tau_\epsilon \to \eta_c$. In addition, Ref. [52] includes a discussion of non-Gaussianities not included in this chapter. Ref. [53] has also recently examined the implications of anisotropic superhorizon evolution in an inflating elastic solid.

spectrum is possible. Interestingly, we find here that in models with slowly evolving material properties a scalar spectral index near $n_{\rm s} \lesssim 1$, compatible with current observational constraints, can be found for w relatively far from the nominal inflationary value $w \simeq -1$.

While we do not specify a particular microphysical model for or formation mechanism of the elastic solid, we note that relativistic elastic solids have been used to model a variety of physical systems including networks of topological defects [55, 56]. A frustrated network of topological defects [57, 58] could potentially form a system with a negative equation of state that is sufficiently stable to provide the ~ 60 e-folds of inflation needed. These systems lack a preferred length scale and as such may undergo vast stretching without fracturing [48].

This chapter is organized as follows: In Sections 4.2 and 4.3 we review the relevant Einstein equations and linearized perturbation equations for a relativistic elastic solid. In Section 4.4 we derive the action for the scalar and tensor linear perturbations of a relativistic elastic solid. The superhorizon evolution in this model is discussed in Section 4.5 and its application to a period of inflation in the early Universe is discussed in Sections 4.6 to 4.8.

4.2 Einstein Equations

We consider the flat Friedmann-Robertson-Walker (FRW) metric [35]

$$ds^2 = a^2(\eta)(\eta_{\alpha\beta} - h_{\alpha\beta})dx^\alpha dx^\beta, \tag{4.1}$$

where η is the conformal time, $\eta_{\alpha\beta}$ is the metric for Minkowski space and the tensor $h_{\alpha\beta}$ represents small perturbations to $\eta_{\alpha\beta}$. The energy density ρ and pressure P of the background in a flat universe can be expressed as

$$\rho = \frac{\mathcal{H}^2}{l^2 a^2}, \qquad P = -\frac{2\mathcal{H}' + \mathcal{H}^2}{3l^2 a^2}, \tag{4.2}$$

where a prime $'$ represents a derivative with respect to the conformal time η, $\mathcal{H} = a'/a = aH$, H is the Hubble parameter, and $l = \sqrt{8\pi G/3}$ is the Planck length. We assume that we can parameterize the pressure by $P = w\rho$, where w is referred to as the equation of state, so that with Eq. (4.2) we can form the differential equation for $\mathcal{H}$

$$\mathcal{H}' = -\frac{1+3w}{2}\mathcal{H}^2. \tag{4.3}$$

[35] We use signature (+,-,-,-) and units in which $c = \hbar = k_{\rm b} = 1$. Our notation largely follows Ref. [9].

Using the Friedmann equation

$$\rho' = -3\mathcal{H}\rho(1+w), \tag{4.4}$$

the relationship between $dP/d\rho$ and w is found to be

$$\frac{dP}{d\rho} = w - \frac{w'}{3\mathcal{H}(1+w)}. \tag{4.5}$$

We parameterize the metric in Eq. (4.1) as

$$ds^2 = a^2(\eta)\left[(1+2\phi)d\eta^2 - 2B_{,i}\,dx^i d\eta \right. \\ \left. -((1+h/3)\delta_{ij} + 2E_{ij})dx^i dx^j\right], \tag{4.6}$$

where h is the trace of the spatial part of the metric perturbation and E_{ij} is traceless. If we decompose the tensor E_{ij} into scalar, vector, and tensor parts, then the scalar component of the spatial part of $h_{\alpha\beta}$ is

$$h^{\mathrm{S}}_{ij} = \frac{h}{3}\delta_{ij} + 2(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2)E = -2\psi\delta_{ij} + 2E_{,ij}\,, \tag{4.7}$$

where $\psi \equiv -\frac{1}{6}h + \frac{1}{3}\nabla^2 E$ is the curvature perturbation and we denote the tensor part of $2E_{ij}$ by the conventional notation h^{T}_{ij}, which in addition to being traceless is transverse ($h^{\mathrm{Ti}}_{j,i} = 0$).

The stress-energy tensor is parameterized in the standard form as

$$\delta T^0_0 = \delta\rho, \tag{4.8a}$$

$$\delta T^i_0 = (\rho + P)v^i, \tag{4.8b}$$

$$\delta T^i_j = -(\delta P\delta^i_j + P\Pi^i_j), \tag{4.8c}$$

where $\delta\rho$ and δP are the energy density and pressure perturbations, respectively, v^i is the velocity perturbation, and Π^i_j is the anisotropic stress.

The gauge-invariant Einstein equations for the scalar perturbations are

$$\nabla^2\Psi - 3\mathcal{H}(\Psi' + \mathcal{H}\Phi) = \frac{3}{2}l^2a^2\delta\rho^{(\mathrm{gi})}, \tag{4.9a}$$

$$\Psi' + \mathcal{H}\Phi = \beta v^{(\mathrm{gi})}, \tag{4.9b}$$

$$\Psi'' + \mathcal{H}(\Phi' + 2\Psi') + (\mathcal{H}^2 + 2\mathcal{H}')\Phi + \frac{1}{3}\nabla^2(\Phi - \Psi) = \frac{3}{2}l^2a^2\delta P^{(\mathrm{gi})}, \tag{4.9c}$$

$$\Psi - \Phi = 3l^2a^2P\Pi, \tag{4.9d}$$

where $\beta = \mathcal{H}^2 - \mathcal{H}' = \frac{3}{2}l^2a^2(\rho+P)$, and the energy-momentum conservation equations are

$$\delta^{(\mathrm{gi})\prime} - (1+w)(\nabla^2 v^{(\mathrm{gi})} + 3\Psi') + 3\mathcal{H}\left(\frac{\delta P^{(\mathrm{gi})}}{\rho} - w\delta^{(\mathrm{gi})}\right) = 0, \tag{4.10a}$$

$$v^{(\mathrm{gi})\prime} + \mathcal{H}(1-3w)v^{(\mathrm{gi})} + \frac{w'}{1+w}v^{(\mathrm{gi})} - \frac{\delta P^{(\mathrm{gi})}}{\rho+P} - \Phi - \frac{2}{3}\frac{w}{1+w}\nabla^2\Pi = 0, \tag{4.10b}$$

where the gauge-invariant perturbation variables are defined as

$$\Phi = \phi + \mathcal{H}(B-E') + (B-E')', \tag{4.11a}$$

$$\Psi = \psi - \mathcal{H}(B-E'), \tag{4.11b}$$

$$\delta\rho^{(\mathrm{gi})} = \delta\rho + \rho'(B-E'), \tag{4.11c}$$

$$\delta P^{(\mathrm{gi})} = \delta P + P'(B-E'), \tag{4.11d}$$

$$v^{(\mathrm{gi})} = v + B - E', \tag{4.11e}$$

noting that Π is already a gauge-invariant quantity.

The sole Einstein equation for the tensor perturbations is

$$(h^{\mathrm{T}})^{i\prime\prime}_j + 2\mathcal{H}(h^{\mathrm{T}})^{i\prime}_j - \nabla^2(h^{\mathrm{T}})^i_j = 6l^2a^2P(\Pi^{\mathrm{T}})^i_j. \tag{4.12}$$

4.3 Elastic Solid

In this section, we briefly summarize the formalism in Ref. [54] for a continuous relativistic medium and the findings of Ref. [49] for the linear perturbations in an isotropic relativistic elastic solid.

As shown in Ref. [54], the behaviour of a continuous relativistic medium can be described by the use of two different manifolds: a three-dimensional manifold $\mathcal{F}$ used to characterize the internal state of the medium and a four-dimensional spacetime manifold $\mathcal{M}$ used to describe its relativistic evolution. A projection $\mathcal{P} : \mathcal{M} \to \mathcal{F}$ is used to project timelike lines in $\mathcal{M}$ onto points in the material space on $\mathcal{F}$. This can be interpreted as projecting the worldline of a 'particle' of the medium onto a single point in the material space. The internal properties of the medium are characterized through tensors defined on $\mathcal{F}$ that are then mapped onto $\mathcal{M}$ via the inverse image $\mathcal{P}^{-1}$. We use

uppercase Latin letters $A, B, \ldots$ and lowercase Greek letters $\mu, \nu, \ldots$ to label the indices of tensors defined on $\mathcal{F}$ and $\mathcal{M}$, respectively.

As the four-demensional projection tensor $\gamma^{\mu\nu} = g^{\mu\nu} - u^\mu u^\nu$ can be used to find the distance between adjacent particles in their local rest frame, $\gamma^{\mu\nu}$ and its material space counterpart γ^{AB} characterize the strain of the medium. Working in material space, we assume that the energy density ρ and pressure tensor P^{AB} can be expressed in terms of the strain tensor γ^{AB} and are related by

$$\partial(\sqrt{|\gamma|}\rho) = -\frac{\sqrt{|\gamma|}}{2} P^{AB} \partial\gamma_{AB}, \tag{4.13a}$$

$$\partial(\sqrt{|\gamma|}P^{AB}) = -\frac{\sqrt{|\gamma|}}{2} E^{ABCD} \partial\gamma_{CD}, \tag{4.13b}$$

in close analogy to the classical case, where $|\gamma|$ is the determinant of γ_{AB}. The elasticity tensor E^{ABCD} has been introduced in Eq. (4.13b) to relate stress and strain tensors, as in the classical case.

As shown in Ref. [49], specifying the pressure and elasticity tensors is sufficient for describing the behaviour of linear perturbations in a relativistic elastic solid. By relating derivatives in $\mathcal{F}$ and $\mathcal{M}$, the spacetime pressure and elasticity tensors for an isotropic elastic solid are

$$P^{\mu\nu} = P\gamma^{\mu\nu}, \tag{4.14a}$$

$$E^{\mu\nu\rho\sigma} = \Sigma^{\mu\nu\rho\sigma} + \left((\rho + P)\frac{dP}{d\rho} - P\right)\gamma^{\mu\nu}\gamma^{\rho\sigma} + 2P\gamma^{\mu(\rho}\gamma^{\sigma)\nu}, \tag{4.14b}$$

where P is the pressure scalar and $\Sigma^{\mu\nu\rho\sigma}$ is the shear tensor given by

$$\Sigma^{\mu\nu\rho\sigma} = 2\mu\left(\gamma^{\mu(\rho}\gamma^{\sigma)\nu} - \frac{1}{3}\gamma^{\mu\nu}\gamma^{\rho\sigma}\right), \tag{4.15}$$

with μ being the shear modulus. For a perfectly elastic medium, the stress-energy tensor is related to the pressure tensor by

$$T^{\mu\nu} = \rho u^\mu u^\nu + P^{\mu\nu}, \tag{4.16}$$

where u^μ are flow vectors tangent to worldlines.

An elastic solid has a resistance to compressive and shearing motions and thus can support both longitudinal and transverse waves, which travel at speeds c_s and c_v, respectively. In both the relativistic and nonrelativistic cases, c_s is dependent upon both the bulk modulus κ and the shear modulus

μ, while $c_{\rm v}$ is dependent only upon μ. In the nonrelativistic case, the sound speeds and bulk modulus are given by [47]

$$c_{\rm s}^2 = \frac{\kappa + \frac{4}{3}\mu}{\rho}, \qquad c_{\rm v}^2 = \frac{\mu}{\rho}, \qquad \kappa = \rho\frac{dP}{d\rho}, \tag{4.17}$$

where the energy density ρ is dominated by the mass contribution in the nonrelativistic limit. The sound speeds for the relativistic case can be found by making the substitution $\rho \to \rho + P$ so that [59]

$$c_{\rm s}^2 = \frac{dP}{d\rho} + \frac{4}{3}c_{\rm v}^2, \qquad c_{\rm v}^2 = \frac{\mu}{\rho + P}, \tag{4.18}$$

and the bulk modulus is now given by $\kappa = (\rho + P)dP/d\rho$. We can see that even in the case where $dP/d\rho$ is negative, a real value for longitudinal sound speed $c_{\rm s}$ can be obtained if the rigidity is sufficiently large.

We now examine the perturbations in the elastic solid. Perturbations in a continuous medium can be described by a shift vector $\xi^\alpha = \xi^\alpha(x_0^\beta)$, so that if a particle in a medium is at position x_0^β when no perturbations are present, then the particle would be at position $x^\alpha(x_0^\beta) = \xi^\alpha(x_0^\beta) + x_0^\alpha$ when perturbations are present. By use of the above equations, it can be shown that the linear perturbations that arise in the stress-energy tensor can be written in terms of the shift vector ξ^α as [49]

$$\delta T_0^0 = -(\rho + P)\left(\xi^k{}_{,k} + \frac{h}{2}\right), \tag{4.19a}$$

$$\delta T_0^i = (\rho + P)\xi^{i\prime}, \tag{4.19b}$$

$$\delta T_j^i = \frac{dP}{d\rho}(\rho + P)\left(\xi^k{}_{,k} + \frac{h}{2}\right)\delta_j^i + \mu\left[2\xi^{(i}{}_{,j)} + h_j^i - \frac{2}{3}\delta_j^i\left(\xi^k{}_{,k} + \frac{h}{2}\right)\right]. \tag{4.19c}$$

By comparing these equations to the standard parameterizations given in Eq. (4.8), we can make the following identifications:

$$\delta\rho = -(\rho + P)\left(\xi^k{}_{,k} + \frac{h}{2}\right), \tag{4.20a}$$

$$v^i = \xi^{i\prime}, \tag{4.20b}$$

$$\delta P = \frac{dP}{d\rho}\delta\rho, \tag{4.20c}$$

$$\Pi_j^i = -\frac{\mu}{P}\left(2\xi^{(i}{}_{,j)} + h_j^i - \frac{2}{3}\delta_j^i\left(\xi^k{}_{,k} + \frac{h}{2}\right)\right). \tag{4.20d}$$

We note that Eq. (4.20c) implies that entropy perturbations are not present in the solid in the sense that the pressure perturbation is fully specified by the energy density and not the entropy. Taking the scalar parts of these equations yields

$$\delta = -(1+w)(\nabla^2\xi^{\mathrm{S}} - 3\psi + \nabla^2 E), \tag{4.21a}$$

$$v = \xi^{\mathrm{S}\prime}, \tag{4.21b}$$

$$\nabla^2\Pi = 2c_{\mathrm{v}}^2(1+w^{-1})\left[-\nabla^2\xi^{\mathrm{S}} - \nabla^2 E\right], \tag{4.21c}$$

where ξ^{S} is the scalar part of the shift vector. Using Eq. (4.21a), we can rewrite the anisotropic stress as

$$\begin{aligned}\nabla^2\Pi &= 2c_{\mathrm{v}}^2(1+w^{-1})\left[\frac{\delta}{1+w} - 3\psi\right] \\ &= -6c_{\mathrm{v}}^2(1+w^{-1})\zeta,\end{aligned} \tag{4.22}$$

where we have identified the gauge-invariant variable ζ as

$$\zeta \equiv \psi + \mathcal{H}\frac{\delta\rho}{\rho'}, \tag{4.23}$$

which can be interpreted as the curvature perturbation on uniform density hypersurfaces or as the density perturbation on uniform curvature hypersurfaces.

The tensor perturbations are simple in comparison. Eq. (4.20) implies that the tensor part of the anisotropic stress is simply

$$(\Pi^T)^i_j = -\frac{\mu}{P}(h^{\mathrm{T}})^i_j. \tag{4.24}$$

Having characterized the general properties of our material, we can now begin to examine how perturbations are excited in an elastic solid.

4.4 Action

To quantize the linear perturbations in the elastic solid, we start with its action and perturb it to second order in the perturbation variables to yield linear equations of motion. We decompose the action as $S = S_{\mathrm{m}} + S_{\mathrm{gr}}$, where S_{m} and S_{gr} are the matter and gravitational parts of the action, respectively. The gravitational part of the action is given by

$$S_{\mathrm{gr}} = -\frac{1}{6l^2}\int \mathcal{R}\sqrt{-g}d^4x, \tag{4.25}$$

where $g \equiv \det(g_{\mu\nu})$ and $\mathcal{R}$ is the Ricci scalar. The matter part of the action for a continuous medium is given by [60]

$$S_{\rm m} = -\int \rho^{\rm tot}\sqrt{-g}d^4x. \tag{4.26}$$

In the above equation, $\rho^{\rm tot}$ is the total energy density, which we decompose as $\rho^{\rm tot} = \rho_f^{\rm tot} + \rho_e^{\rm tot}$, where $\rho_f^{\rm tot}$ is the energy density corresponding to a perfect fluid and $\rho_e^{\rm tot}$ is the additional energy density arising from shear stresses in the elastic solid. The perfect fluid part of the action taken to second order in the perturbation variables can be expressed as [9]

$$\delta_2 S_f = -\int \left[\rho\frac{\delta_2\sqrt{-g}}{\sqrt{-g_0}} + (\rho+P)\left(\frac{\delta_1 n}{n_0}\frac{\delta_1\sqrt{-g}}{\sqrt{-g_0}} + \frac{\delta_2 n}{n_0}\right) + \frac{1}{2}\frac{dP}{d\rho}(\rho+P)\left(\frac{\delta_1 n}{n_0}\right)^2\right]\sqrt{-g_0}d^4x, \tag{4.27}$$

where here the subscript 0 indicates the background value, δ_1 and δ_2 denote the terms in a variable containing first and second order perturbations, respectively, and n is the number density. For a relativistic isotropic elastic solid, $\rho_e^{\rm tot}$ is given by [48]

$$\rho_e^{\rm tot} = \frac{P^2}{4\mu}\Pi_{ij}\Pi^{ij}, \tag{4.28}$$

so that the action for the elastic part perturbed to second order is

$$\delta_2 S_e = -\int \frac{a^4 P^2}{4\mu}\Pi_{ij}\Pi^{ij}d^4x. \tag{4.29}$$

4.4.1 Quantization of Scalar Modes

Using the expressions for the action as described above, the scalar part of the perfect fluid action, including the gravitational part, perturbed to second order can be found to be [9]

$$\delta_2 S_f + \delta_2 S_{\rm gr} = \frac{1}{6l^2}\int a^2\left[-6\left(\psi'^2 + 2\mathcal{H}\phi\psi' + \left(\mathcal{H}^2 - \frac{\beta}{3\frac{dP}{d\rho}}\right)\phi^2\right) -4(\psi' + \mathcal{H}\phi)\nabla^2(B - E') - 2\psi_{,i}\,(2\phi_{,i} - \psi_{,i}\,) + 2\beta(v^{,i} + B_{,i}\,)(v^{,i} + B_{,i}\,) -2\beta\frac{dP}{d\rho}\left(3\psi - \nabla^2 E - \nabla^2\xi^{\rm S} + \frac{\phi}{\frac{dP}{d\rho}}\right)^2\right]d^4x.$$

We now wish to put the action in canonical form. We will work in the comoving gauge where $v = B = 0$. The main reasons for this choice of gauge are that the action above is simplified greatly in this gauge and that the gauge-invariant variable R, defined by

$$R \equiv \mathcal{H}v + \psi, \tag{4.30}$$

in the comoving gauge is simply related to the metric perturbation ψ by $R = \psi$ and so represents the curvature perturbation in this gauge. If we are able to form an expression solely in terms of ψ in this gauge, the gauge-invariant expression can then be trivially found by substituting R for ψ. In the comoving gauge, the Einstein equations (4.9a) and (4.9b) and the momentum conservation equation (4.10b) are

$$\nabla^2(\psi + \mathcal{H}E') - 3\mathcal{H}(\psi' + \mathcal{H}\phi) - \frac{3}{2}\mathcal{H}^2\delta = 0, \tag{4.31a}$$

$$\psi' + \mathcal{H}\phi = 0, \tag{4.31b}$$

$$\frac{\frac{dP}{d\rho}}{1+w}\delta + \phi + \frac{2}{3}\frac{w}{1+w}\nabla^2\Pi = 0. \tag{4.31c}$$

Using Eqs. (4.31b) and (4.31c), the fluid part of the action in the comoving gauge becomes

$$\delta_2 S_f + \delta_2 S_{\rm gr} = \frac{1}{3l^2}\int\left[\frac{3(1+w)}{\frac{dP}{d\rho}}\psi'^2 - 3(1+w)\psi_{,i}^2 - \frac{4}{3}\frac{w^2}{1+w}\frac{\mathcal{H}^2}{\frac{dP}{d\rho}}(\nabla^2\Pi)^2\right]d^4x, \tag{4.32}$$

where a total derivative term has been dropped.

We now turn our attention to the additional part of the action for an elastic solid given in Eq. (4.29). For the scalar part of the anisotropic stress tensor, we have $(\Pi^S)_{ij}(\Pi^S)^{ij} = \frac{2}{3}(\nabla^2\Pi)^2 + $ (total derivative term), so the scalar part of the elastic part of the action is

$$\delta_2 S_e = -\frac{1}{6l^2}\int\frac{w^2a^2\mathcal{H}^2}{c_{\rm v}^2(1+w)}(\nabla^2\Pi)^2 d^4x, \tag{4.33}$$

where we have used the sounds speeds in Eq. (4.18). We can then write the

total action for the scalar perturbations as

$$\delta_2 S = \frac{1}{6l^2}\int a^2 \left[\frac{3(1+w)}{\frac{dP}{d\rho}}\psi'^2 - 3(1+w)\psi_{,i}^2 - \frac{c_\mathrm{s}^2 w^2 \mathcal{H}^2}{c_\mathrm{v}^2(1+w)\frac{dP}{d\rho}}(\nabla^2\Pi)^2\right] d^4x. \quad (4.34)$$

We can express $\nabla^2\Pi$ as a function of ψ by using Eqs. (4.22), (4.31b), and (4.31c), which yields

$$\nabla^2\Pi = 2\frac{c_\mathrm{v}^2}{c_\mathrm{s}^2}\frac{1+w}{w}\left[\frac{\psi'}{\mathcal{H}} - 3\frac{dP}{d\rho}\psi\right]. \quad (4.35)$$

Using this expression in the action above gives

$$\delta_2 S = \frac{1}{3l^2}\int z^2 \left[R'^2 - c_\mathrm{s}^2 R_{,i}^2 - 4c_\mathrm{v}^2\beta R^2 - 4c_\mathrm{v}^2\mathcal{H}\left(\frac{(c_\mathrm{v}^2)'}{c_\mathrm{v}^2} - \frac{(c_\mathrm{s}^2)'}{c_\mathrm{s}^2}\right)R^2\right] d^4x, \quad (4.36)$$

where we have cast the action into a gauge-invariant form and a total derivative term has been dropped. We have defined z by

$$z \equiv \frac{a\sqrt{\beta}}{c_\mathrm{s}\mathcal{H}} = \frac{a}{c_\mathrm{s}}\sqrt{\frac{3}{2}(1+w)}. \quad (4.37)$$

We can now define the canonical variable u as

$$u \equiv \sqrt{\frac{2}{3l^2}}zR, \quad (4.38)$$

so that the action becomes

$$\delta_2 S = \frac{1}{2}\int \left[u'^2 - c_\mathrm{s}^2 u_{,i}^2 - m_{\mathrm{eff,S}}^2(\eta)u^2\right] d^4x, \quad (4.39)$$

where another total derivative term has been dropped and the effective mass is

$$m_{\mathrm{eff,S}}^2(\eta) \equiv -\frac{z''}{z} + 4c_\mathrm{v}^2\beta + 4c_\mathrm{v}^2\mathcal{H}\left(\frac{(c_\mathrm{v}^2)'}{c_\mathrm{v}^2} - \frac{(c_\mathrm{s}^2)'}{c_\mathrm{s}^2}\right). \quad (4.40)$$

Varying the action with respect to u leads to the equation of motion

$$u'' - c_\mathrm{s}^2\nabla^2 u + m_{\mathrm{eff,S}}^2(\eta)u = 0. \quad (4.41)$$

It is clear that the action in Eq. (4.39) has the same form as the action for a harmonic oscillator with time-dependent mass, so we may use the same quantization procedure as is used to quantize a harmonic oscillator. The conjugate momentum π to u is

$$\pi = \frac{\partial \mathcal{L}}{\partial u'} = u'. \tag{4.42}$$

We now promote u and π to operators $\hat{u}$ and $\hat{\pi}$ and impose the commutation relations

$$\begin{aligned} &[\hat{u}(\mathbf{x},\eta), \hat{u}(\tilde{\mathbf{x}},\eta)] = [\hat{\pi}(\mathbf{x},\eta), \hat{\pi}(\tilde{\mathbf{x}},\eta)] = 0, \\ &[\hat{u}(\mathbf{x},\eta), \hat{\pi}(\tilde{\mathbf{x}},\eta)] = i\delta(\mathbf{x}-\tilde{\mathbf{x}}). \end{aligned} \tag{4.43}$$

Using the Fourier conventions

$$f(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^{3/2}} f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{4.44}$$

we can write $\hat{u}_{\mathbf{k}}$ in terms of the creation and annihilation operators $\hat{a}^\dagger_{\mathbf{k}}$ and $\hat{a}_{\mathbf{k}}$ as $\hat{u}_{\mathbf{k}}(\eta) = (\hat{a}_{\mathbf{k}}\chi^*_{\mathbf{k}}(\eta) + \hat{a}^\dagger_{-\mathbf{k}}\chi_{\mathbf{k}}(\eta))/\sqrt{2}$ so that

$$\hat{u}(\mathbf{x},\eta) = \frac{1}{\sqrt{2}} \int \frac{d^3k}{(2\pi)^{3/2}} \left[\hat{a}_{\mathbf{k}}\chi^*_{\mathbf{k}}(\eta)e^{i\mathbf{k}\cdot\mathbf{x}} + \hat{a}^\dagger_{\mathbf{k}}\chi_{\mathbf{k}}(\eta)e^{-i\mathbf{k}\cdot\mathbf{x}}\right], \tag{4.45}$$

and the mode function $\chi_{\mathbf{k}}(\eta)$ obeys

$$\chi''_{\mathbf{k}} + \left[c_{\mathrm{s}}^2 k^2 + m^2_{\mathrm{eff,S}}\right]\chi_{\mathbf{k}} = 0. \tag{4.46}$$

The commutation relations in Eq. (4.43) imply the normalization

$$\chi'_{\mathbf{k}}\chi^*_{\mathbf{k}} - \chi_{\mathbf{k}}\chi^{*\prime}_{\mathbf{k}} = 2i. \tag{4.47}$$

Once we solve for the mode function $\chi_{\mathbf{k}}$ from the differential equation in Eq. (4.46), subject to the the normalization condition above, we can calculate the power spectrum for the scalar perturbations $\mathcal{P}_R(k) = |R_k|^2 k^3/2\pi^2$. We will see in Section 4.5 that the perturbations associated with a single scalar mode with wavevector $\mathbf{k}$ is anisotropic, due to the presence of anisotropic stress. However, each mode has the same evolution (the solution for $\chi_{\mathbf{k}}$ in Eq. (4.46) is the same for each $\mathbf{k}$ with the same magnitude k). As discussed in Section 4.5, we will assume that expectation values are isotropic, so that $\langle u_{\mathbf{k}} u_{\tilde{\mathbf{k}}}\rangle = |u_k|^2\delta(\mathbf{k}+\tilde{\mathbf{k}})$. As a result, the integrand of the integral over d^3k in the two-point correlation function for the operator

$\hat{u}(\mathbf{x},\eta)$ with the vacuum state will depend only on k, so we can trivially integrate over the solid angle $d\Omega_k$, so that

$$\langle 0|\hat{u}(\mathbf{x},\eta)\hat{u}(\tilde{\mathbf{x}},\eta)|0\rangle = \int dk \frac{k^2}{4\pi^2}|\chi_k(\eta)|^2 \frac{\sin(k|\mathbf{x}-\tilde{\mathbf{x}}|)}{k|\mathbf{x}-\tilde{\mathbf{x}}|}. \tag{4.48}$$

We can now identify the power spectrum for u as $\mathcal{P}_u(k) = \frac{k^3}{4\pi^2}|\chi_k|^2$ and using Eq. (4.38) the power spectrum for R will be

$$\mathcal{P}_R(k) = \frac{3l^2k^3}{8\pi^2 z^2}|\chi_k|^2. \tag{4.49}$$

4.4.2 Quantization of Tensor Modes

We now turn our attention to the tensor modes. Using the tensor part of the metric in Eq. (4.6) to calculate the tensor part of the Ricci scalar $\mathcal{R}$, the gravitational part of the action can be found to be

$$\delta_2 S_{\rm gr} = \frac{1}{24l^2}\int a^2 \left[(h^{\rm T})^{i\prime}_j (h^{\rm T})^{j\prime}_i - (h^{\rm T})^i_{j,l}(h^{\rm T})^{j,l}_i\right] d^4x. \tag{4.50}$$

The only contribution to the matter part of the action is from the tensor part of the anisotropic stress, given by Eq. (4.24), which using the elastic part of the action in Eq. (4.29) is

$$\delta_2 S_e = -\int \frac{a^4\mu}{4P}(h^{\rm T})^i_j (h^{\rm T})^j_i d^4x. \tag{4.51}$$

With the transverse sound speed given in Eq. (4.18), the total action becomes

$$\delta_2 S = \frac{1}{24l^2}\int a^2 \Big[(h^{\rm T})^{i\prime}_j (h^{\rm T})^{j\prime}_i - (h^{\rm T})^i_{j,l}(h^{\rm T})^{j,l}_i \\ -4c_{\rm v}^2\beta (h^{\rm T})^i_j (h^{\rm T})^j_i\Big] d^4x. \tag{4.52}$$

It is convenient to express $(h^{\rm T})^i_j$ in terms of the individual polarization states $h^{\rm T}_p$, where $(h^{\rm T})^i_j(h^{\rm T})^j_i = 2\sum_p (h^{\rm T}_p)^2$, so that the action for each polarization state is

$$\delta_2 S = \frac{1}{12l^2}\int a^2 \left[(h^{\rm T}_p)'^2 - (h^{\rm T}_p)^2_{,i} - 4c_{\rm v}^2\beta (h^{\rm T}_p)^2\right] d^4x. \tag{4.53}$$

We can define the canonical variable U_p for the tensor perturbations as

$$U_p = \frac{a h^T_p}{\sqrt{6l^2}}, \tag{4.54}$$

so the action becomes

$$\delta_2 S = \frac{1}{2}\int \left[U_p'^2 - U_{p,i}^2 - m_{\mathrm{eff,T}}^2 U_p^2\right] d^4x, \tag{4.55}$$

where a total derivative has been dropped and the effective mass for the tensor modes is given by

$$m_{\mathrm{eff,T}}^2 = -\frac{a''}{a} + 4c_{\mathrm{v}}^2\beta. \tag{4.56}$$

As with the scalar perturbations, the action for the tensor perturbations has the same form as the action of a harmonic oscillator with time-dependent mass. We can therefore use the same quantization procedure as was used with the scalar perturbations. We promote the canonical variable U_p to an operator and write it in terms of creation and annihilation operators as

$$\hat{U}_p(\mathbf{x},\eta) = \frac{1}{\sqrt{2}}\int \frac{d^3k}{(2\pi)^{3/2}}\left[\hat{a}_{\mathbf{k}} X_{\mathbf{k}}^*(\eta) e^{i\mathbf{k}\cdot\mathbf{x}} + \hat{a}_{\mathbf{k}}^\dagger X_{\mathbf{k}}(\eta) e^{-i\mathbf{k}\cdot\mathbf{x}}\right], \tag{4.57}$$

and so the equation of motion for the mode function $X_{\mathbf{k}}$ is

$$X_{\mathbf{k}}'' + \left[k^2 + m_{\mathrm{eff,T}}^2\right] X_{\mathbf{k}} = 0. \tag{4.58}$$

The commutation relations analogous to Eq. (4.43) for the tensor case yields the normalization condition

$$X_{\mathbf{k}}' X_{\mathbf{k}}^* - X_{\mathbf{k}} X_{\mathbf{k}}^{*\prime} = 2i. \tag{4.59}$$

Accounting for both polarization states, the two-point function for the tensor perturbations with the vacuum state is then

$$\langle 0|(\hat{h}^T)_j^i(\mathbf{x},\eta)(\hat{h}^T)_i^j(\tilde{\mathbf{x}},\eta)|0\rangle = \int dk \frac{6l^2}{\pi^2 a^2} k^2 |X_k(\eta)|^2 \times \frac{\sin(k|\mathbf{x}-\tilde{\mathbf{x}}|)}{k|\mathbf{x}-\tilde{\mathbf{x}}|}, \tag{4.60}$$

with which we can identify the tensor power spectrum $\mathcal{P}_{\mathrm{T}}$ as

$$\mathcal{P}_{\mathrm{T}}(k) = \frac{6l^2k^3}{\pi^2a^2}|X_k|^2. \tag{4.61}$$

4.5 Superhorizon Evolution

An interesting phenomenon in this model is that both ζ and R evolve on superhorizon scales, even when the elastic solid is the only substance present

in the Universe. Typically, this type of superhorizon evolution only arises in the presence of a nonadiabatic pressure $\delta P_{\rm nad} = \delta P - (dP/d\rho)\delta\rho$. From Section 4.3, we saw that $\delta P/\delta\rho = dP/d\rho$ for an elastic solid, so the nonadiabatic pressure vanishes. However, the addition of the anisotropic stress in the elastic solid adds another type of stress to the system, which causes superhorizon evolution in a similar manner to cases when nonadiabatic pressures are present.

In the standard case when only adiabatic and isotropic pressures are present, both ζ and R remain approximately constant on superhorizon scales because once smoothed on a scale much larger than the horizon, each patch of the Universe smaller than the smoothing scale evolves approximately like a separate unperturbed FRW universe. This idea is known as the 'separate universe approach' [61]. The locally defined expansion $\tilde{\theta}(\mathbf{x}, t)$ with respect to coordinate time t is given by

$$\tilde{\theta}(\mathbf{x}, t) = 3H - 3\dot{\psi}(\mathbf{x}, t) + \nabla^2\sigma(\mathbf{x}, t), \tag{4.62}$$

where an overdot denotes a derivative with respect to coordinate time and $\sigma = \dot{E} - B$ is the (local) shear. Considering a flat slicing ($\psi = 0$), the local expansion will be equal to the background value if we can safely neglect the effects of the shear on large scales. In this case, since the (total) energy density evolves according to the local energy conservation equation, which to linear order is

$$\dot{\rho}^{\rm tot}(\mathbf{x}, t) = -[\tilde{\theta}(\mathbf{x}, t) + \nabla^2 v(\mathbf{x}, t)][\rho^{\rm tot}(\mathbf{x}, t) + P^{\rm tot}(\mathbf{x}, t)]. \tag{4.63}$$

Then after smoothing on superhorizon scales, the local energy density at each location will (approximately) follow the same unperturbed FRW evolution. Therefore, the difference between the energy density perturbations at different locations will be kept approximately constant in time and since ζ is proportional to the energy density perturbation in a flat slicing, ζ will be approximately constant on superhorizon scales.[36]

As was shown in Ref. [62], when the anisotropic stress is neglected, the shear is in fact negligible on large scales. However, the anisotropic stress acts as a source term for the shear (see Eq. (31) of Ref. [63]), causing the shear to be non-negligible on superhorizon scales in the case of an elastic solid. If the shear cannot be neglected, then different locations in the Universe after smoothing on superhorizon scales will not evolve as an unperturbed FRW

[36] A similar argument can be made for R, which is proportional to v in a spatially flat slicing, by using the local conservation of momentum.

universe owing to the fact that a FRW spacetime is shear free and in general the local expansion will be position dependent in a flat slicing.

Working in the gauge where $\psi = B = 0$ so that the shear is $\sigma = \dot{E}$, the trace-free part of the spatial components of the Einstein equations in Fourier space is

$$\dot{\sigma}_{\mathbf{k}} + 3H\sigma_{\mathbf{k}} - \frac{k^2}{a^2}\phi_{\mathbf{k}} = 3wH^2\Pi_{\mathbf{k}}. \tag{4.64}$$

As we can see, the shear is indeed sourced by the anisotropic stress and will evolve on superhorizon scales unless the anisotropic stress is negligible on these scales. It is easily seen from Eq. (4.22) that for an elastic solid the anisotropic stress is significant (i.e. comparable to the energy density and pressure perturbations) on superhorizon scales since in a spatially flat slicing $\Pi_{\mathbf{k}} = -2(c_{\rm v}^2/w)\delta_{\mathbf{k}}$. Therefore, we will have $|\Pi_{\mathbf{k}}| \sim |\delta_{\mathbf{k}}|$ in this slicing for a sufficiently rigid solid and the shear will evolve as

$$\dot{\sigma}_{\mathbf{k}} + 3H\sigma_{\mathbf{k}} - \frac{k^2}{a^2}\phi_{\mathbf{k}} = -6c_{\rm v}^2H^2\delta_{\mathbf{k}}, \tag{4.65}$$

so that the shear is sourced by the (non-negligible) density perturbations on superhorizon scales when viewed in this gauge.

If we consider a mode with wavevector $\mathbf{k} = (0, 0, k)$ then the scalar part of the anisotropic stress tensor in Fourier space, given by $(\Pi^S_{\mathbf{k}})_{ij} = (-k_ik_j/k^2 + \delta_{ij}/3)\Pi_{\mathbf{k}}$, is $(\Pi^S_{\mathbf{k}})_{ij} = \mathrm{diag}(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3})\Pi_{\mathbf{k}}$. From Eq. (4.8), the scalar perturbation to the spatial part of the stress tensor in a spatially flat slicing will be $(\delta T^S_{\mathbf{k}})_{ij} = -\mathrm{diag}(c_{\rm s}^2 - 2c_{\rm v}^2, c_{\rm s}^2 - 2c_{\rm v}^2, c_{\rm s}^2)\delta\rho_{\mathbf{k}}$, which is inherently anisotropic, having a different pressure in directions parallel and perpendicular to the direction of propagation.

Instead of considering perturbations about a FRW spacetime, we now examine the behaviour of an unperturbed Bianchi spacetime, which has the defining properties of being homogeneous and in general anisotropic. We concentrate on the Bianchi type I spacetime that has the metric

$$ds^2 = dt^2 - a_x(t)^2dx^2 - a_y(t)^2dy^2 - a_z(t)^2dz^2, \tag{4.66}$$

where a_x, a_y, and a_z are directional scale factors. The properties of nearly isotropic Bianchi spacetimes were detailed in Ref. [64], which treated their departure from isotropy as a linear perturbation. After smoothing on superhorizon scales, the metric perturbation E_{ij} in Eq. (4.6) from perturbing about a flat FRW spacetime is simply the symmetric trace-free tensor characterizing the anisotropy of a nearly isotropic Bianchi I spacetime to linear

order.[37] For example, consider the mode $\mathbf{k} = (0, 0, k)$. After smoothing, E_{ij} will be approximately uniform in a local patch of the Universe. In a spatially flat gauge where $h = 2\nabla^2 E$, if we denote the average value of h in this patch as $\bar{h}$, then a scalar mode with this wavevector will evolve approximately as a Bianchi I spacetime with $a_x = a_y = a$ and $a_z = a + \bar{h}$. A tensor mode with 'plus' polarization, with an average value of $\bar{h}_+$ in this patch, would evolve with directional scale factors $a_x = a + \bar{h}_+$, $a_y = a - \bar{h}_+$, and $a_z = a$.

For a single mode, we can absorb the shear on superhorizon scales into the background spacetime by perturbing about a Bianchi I spacetime instead of a flat FRW (which is the isotropic special case of Bianchi I). In this case, after smoothing on superhorizon scales, perturbations would again evolve according to an unperturbed metric, but in general would be of type Bianchi I, not FRW. In the standard inflationary scenario, the shear is negligible on superhorizon scales, so E_{ij} is approximately constant on these scales and can be removed from the metric by a simple coordinate redefinition, leaving the isotropic special case of our spacetime.

Although a single mode is formally anisotropic, in the case of inflation, modes are excited on a wide range of scales in all directions. We assume that initial perturbations are drawn from an isotropic Gaussian distribution and that expectation values will be isotropic and therefore continue to examine the perturbations in the metric in Eq. (4.1) in which perturbations are taken about an isotropic FRW spacetime.[38]

4.6 Inflation

We now apply the results of the previous sections to the case where inflation is driven by an elastic solid. We divide the analysis into two parts: the simple case with constant sound speeds and equation of state and the case where they are varying in time. We then consider the more specialized case where the sound speeds and equation of state slowly vary with time.

[37] See Eq. (59) of Ref. [64]. Also note that on superhorizon scales, by making the substitution $\sigma \to \dot{E}$, Eq. (4.64) is approximately Eq. (39) in Ref. [64].

[38] Although we take the expectation values over a single realization to be isotropic, in principle, a residual net anisotropy might persist. The persistence of anisotropic geometries in this context was recently studied in Ref. [53]. We set aside here questions pertaining to the precise size and impact of sustained anisotropies and only assume that corrections to the evolution of linear perturbations in an isotropic background appear at higher order.

4.6.1 Inflation with Constant Sound Speeds and Equation of State

With constant equation of state, $\mathcal{H}$ can easily be solved from Eq. (4.3) with an appropriate integration constant as

$$\mathcal{H} = \frac{2}{(1+3w)\eta}. \tag{4.67}$$

From Eq. (4.40), we see that in this case the effective mass for the scalar modes becomes

$$m^2_{\mathrm{eff,S}}(\eta) = -\frac{2-6w-24c_{\mathrm{v}}^2(1+w)}{(1+3w)^2\eta^2}. \tag{4.68}$$

The general solution for the mode function $\chi_{\mathbf{k}}(\eta)$ can now easily be found from Eq. (4.46) as

$$\chi_k = \sqrt{\frac{\pi|\eta|}{2}}\left[C_1 H_\nu^{(1)}(c_{\mathrm{s}}k|\eta|) + C_2 H_\nu^{(2)}(c_{\mathrm{s}}k|\eta|)\right], \tag{4.69}$$

where $H_\nu^{(1)}$ and $H_\nu^{(2)}$ are the Hankel functions of the first and second kind, C_1 and C_2 are integration constants and the index ν is

$$\nu = \frac{1}{2}\sqrt{1+4\left(\frac{2-6w-24c_{\mathrm{v}}^2(1+w)}{(1+3w)^2}\right)}. \tag{4.70}$$

When a mode is well within the horizon with $c_{\mathrm{s}}k|\eta| \gg 1$, the mode function can be approximated by

$$\chi_k \approx \frac{1}{\sqrt{c_{\mathrm{s}}k}}\left(C_1 e^{ic_{\mathrm{s}}k\eta} + C_2 e^{-ic_{\mathrm{s}}k\eta}\right), \tag{4.71}$$

which is the solution for Minkowski space. We initialize the mode by assuming that it is in its lowest energy state when it is well within the horizon, with mode function

$$\chi_k \approx \frac{1}{\sqrt{c_{\mathrm{s}}k}} e^{ic_{\mathrm{s}}k\eta}. \tag{4.72}$$

With these constants of integration, the mode function becomes

$$\chi_k = \sqrt{\frac{\pi|\eta|}{2}} H_\nu^{(2)}(c_{\mathrm{s}}k|\eta|). \tag{4.73}$$

When the mode is far outside the horizon with $c_{\rm s}k|\eta| \ll 1$, the mode function can be approximated as

$$\chi_k \approx \sqrt{\frac{\pi|\eta|}{2}}\frac{i\Gamma(\nu)}{\pi}\left(\frac{c_{\rm s}k|\eta|}{2}\right)^{-\nu}, \tag{4.74}$$

where $\Gamma(\nu)$ is the gamma function. With the evolution of the mode function $\chi_{\mathbf{k}}$ for modes well outside the horizon, we find the power spectrum for R to be

$$\mathcal{P}_R(k) \approx \frac{c_{\rm s}^{2(1-\nu)}\Gamma^2(\nu)4^\nu}{8\pi^3(1+w)}\frac{l^2k^{3-2\nu}|\eta|^{1-2\nu}}{a^2}. \tag{4.75}$$

Since for constant equation of state the scale factor evolves as $a \propto |\eta|^{\frac{2}{1+3w}}$, we do indeed see that R_k evolves with time when the mode is on superhorizon scales if $c_{\rm v}$ is nonzero. Using Eq. (4.67), we can write the power spectrum in terms of the Hubble parameter as

$$\mathcal{P}_R(k) \approx \frac{c_{\rm s}^{2(1-\nu)}\Gamma^2(\nu)}{4\pi^3(1+w)|1+3w|^{1-2\nu}}l^2(k/a)^{3-2\nu}H^{-1+2\nu}. \tag{4.76}$$

Although modes evolve on superhorizon scales, all modes well outside the horizon share the same time evolution. In other words, the presence of a superhorizon evolution will not affect the relative scale dependence of modes on superhorizon scales. Thus, we can calculate quantities like the scalar spectral index $n_{\rm s} = 1 + d\ln\mathcal{P}_R/d\ln k$ and its running using the same methods that are used in the case where the superhorizon evolution is small.

For constant sound speeds and equation of state, the scalar spectral index is $n_{\rm s} = 4 - 2\nu$. The necessary restrictions of $-1 < w < -1/3$, $0 \le c_{\rm v}^2 \le 1$, and $0 < c_{\rm s}^2 \le 1$ imply that $n_{\rm s}$ is bound from below by one. Thus, the scalar power spectrum for this case can only have a blue tilt, which has been ruled out to a high degree of likelihood [25]. If w is near -1, we can see from Eq. (4.68) that the effect of the shear stress on $m^2_{\rm eff,S}$ will be small and a nearly scale-invariant spectrum will be produced, as is the case in many models of inflation.

It is interesting to note that $c_{\rm s}$ near zero (w near $-\frac{4}{3}c_{\rm v}^2$) also produces a nearly scale-invariant two-point spectrum. In this case, the w dependence of the $-z''/z$ and $4c_{\rm v}^2\beta$ terms in $m^2_{\rm eff,S}$ cancels with one another so that a nearly scale-invariant spectrum can be produced for values of w far from -1 (but still bounded by $-1 < w < -1/3$). This result is not possible in standard inflationary models, since the $4c_{\rm v}^2\beta$ term is absent in these cases, so the w dependence of $m^2_{\rm eff,S}$ remains important.

As discussed in Ref. [66], inflationary scenarios that produce a nearly scale-invariant two-point function with a background in the far-from-de Sitter regime generically do not have nearly scale-invariant higher-point correlations, provided the perturbations are adiabatic in the sense of Ref. [67] — which requires the anisotropic stress to be negligible on large scales. However, the elastic solid model we describe here requires non-negligible anisotropic stress on large scales for a consistent description of linear perturbations. We leave the interesting question of higher-point correlation functions in far-from-de Sitter accelerating elastic solid models for future work.

Additionally, when w is far from -1, one must take care that an adequate number of e-folds of inflation can occur as the energy density and horizon size may evolve significantly during inflation. This can alter the minimum number of e-folds required to solve the 'horizon problem'. An upper bound on the number of e-folds of inflation will be set by putting bounds on the energy density, set on the lower end by the reheat temperature and on the higher end by a high-energy limit (see Section 4.6.2 for further details). Requiring inflation to start below the Planck scale and end with temperatures above $\sim$10s of MeV puts an upper bound on the equation of state of $w \lesssim -2/5$ when a nearly scale-invariant spectrum is achieved. However, as will be discussed in Section 4.8, a power spectrum amplitude compatible with current observations requires either very small values of $c_{\rm s}$ or super-Planckian densities for values of w extremely far from -1.

Returning to the discussion of Section 4.5, for constant sound speeds and equation of state with $c_{\rm v} \neq 0$, the superhorizon modes of h in the $\psi = B = 0$ gauge (as well as E since $h = 2\nabla^2 E$ in this gauge) evolve as

$$h'_{\mathbf{k}} \propto A_{\mathbf{k}} k^{-\nu} |\eta|^{-\frac{5+3w}{2+6w}-\nu}, \tag{4.77}$$

where the factor $A_{\mathbf{k}}$ determines the initial amplitude of $h'_{\mathbf{k}}$ for a particular mode. If we scale the wavevector of the mode as $\mathbf{k} \to \alpha\mathbf{k}$ for some constant α, the same late-time evolution of the superhorizon mode can remain unchanged by simultaneously scaling $A_{\mathbf{k}} \to \alpha^{\nu} A_{\mathbf{k}}$, an example of which can be seen in Fig. 4.1. As such, we cannot determine which particular modes a superhorizon sized anisotropy originated from.

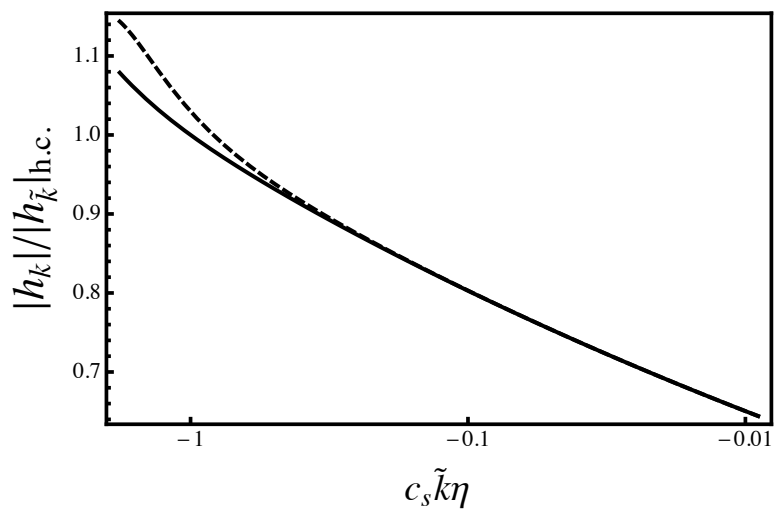


Figure 4.1: Evolution of h modes in the $\psi = B = 0$ gauge for $w = -0.9$ and $c_{\text{v}}^2 = 0.8$ (note that a plot of E_k would look identical as $h_{\mathbf{k}} = -2E_{\mathbf{k}}$ in this gauge). The solid and dashed lines show the evolution for modes with $|\mathbf{k}| = \tilde{k}$ and $|\mathbf{k}| = 2\tilde{k}$, respectively, and the subscript h.c. denotes horizon crossing. Initial amplitudes of the perturbations are chosen so that the modes coincide when both are on superhorizon scales.

4.6.2 The 'Horizon Problem' Revisited

An interesting feature of this model of inflation is that it allows the possibility of far from de Sitter backgrounds that nevertheless produce a nearly scale-invariant two-point correlation function. In such a case, the horizon may change significantly during inflation, thus altering the 'horizon problem', in which we expect to be able to fit the present-day horizon size into the horizon at the beginning of inflation expanded to today. Labelling quantities evaluated at the beginning of inflation, reheating, and the present-day by the subscripts i, RH, and 0, respectively, we require

$$\begin{aligned} H_0^{-1} &\leq \frac{a_0}{a_i} H_i^{-1} \\ &= \frac{a_0}{a_{\text{RH}}} \frac{a_{\text{RH}}}{a_i} H_i^{-1} \\ &\approx \left(\frac{g_{s\text{RH}}}{g_{s0}}\right)^{1/3} \frac{T_{\text{RH}}}{T_0} e^N H_i^{-1}, \end{aligned} \tag{4.78}$$

where N is the number of e-folds of inflation and g_s is the effective number of relativistic degrees of freedom contributing to the entropy density.

If the equation of state during inflation is w, then assuming w is approximately constant, the horizon size changes as $\propto a^{3(1+w)/2}$ during inflation, so that $H_i^{-1} \approx e^{-3(1+w)N/2} H_{\rm RH}^{-1}$. The minimum number of e-folds of inflation to solve the 'horizon problem' then becomes

$$\begin{aligned} N &\geq \frac{1}{1-\epsilon}\left[\ln\left(\frac{T_0}{H_0}\right) + \ln\left(\frac{H_{\rm RH}}{T_{\rm RH}}\right) + \frac{1}{3}\ln\left(\frac{g_{s0}}{g_{s\rm RH}}\right)\right] \\ &\approx \frac{1}{1-\epsilon}\left[54 + \ln\left(\frac{T_{\rm RH}}{10^{13}\,{\rm GeV}}\right) + \frac{1}{3}\ln\left(\frac{106.75}{g_{s\rm RH}}\right)\right], \end{aligned} \tag{4.79}$$

where $\epsilon = 3(1+w)/2$ and we have taken $g_{s0} = 43/11$. Thus, the minimum number of e-folds required when w is far from -1 (ϵ is large) may be substantially larger than that for the standard $w \simeq -1$ case.

If Λ is a high-energy cut-off scale that is an upper bound for the initial energy density of inflation (presumably the Planck scale), then for $\epsilon \neq 0$, N is bounded from above by

$$N \lesssim \frac{1}{2\epsilon}\left[172 + 4\left[\ln\left(\frac{\Lambda}{m_{\rm p}}\right) - \ln\left(\frac{T_{\rm RH}}{\rm GeV}\right)\right] - \ln\left(\frac{g_{\rm RH}}{106.75}\right)\right], \tag{4.80}$$

where $m_{\rm p} = \sqrt{8\pi/3}\, l^{-1}$ is the Planck mass and $g_{\rm RH}$ is the effective number of relativistic degrees of freedom contributing to the energy density. If the bound from Eq. (4.79) provides the strongest constraint on the minimum number of e-folds of inflation, which will likely be the case if w is far from -1, then the maximum reheat temperature such that N is appropriately bounded is

$$\begin{aligned} \log_{10}\left(\frac{T_{\rm RH}}{\rm GeV}\right) \approx \frac{1}{1-\epsilon/2}\Bigg(&19 - 24\epsilon \\ &+0.44(1-\epsilon)\ln\left(\frac{\Lambda}{m_{\rm p}}\right) + (0.18\epsilon - 0.11)\ln\left(\frac{g_{\rm RH}}{106.75}\right)\Bigg), \end{aligned} \tag{4.81}$$

where we have assumed $g_{\rm RH} = g_{s\rm RH}$.

4.6.3 Non-Constant Sound Speeds and Equation of State

Since having the sound speeds and equation of state perfectly constant cannot result in a red-tilted scalar spectrum, we would like to examine if adding a time dependence can result in a red-tilted scalar spectrum. It will prove useful to introduce an alternative time variable q, defined by

$$q(\eta) \equiv -\int^{\eta} c_{\rm s}(\tilde{\eta})d\tilde{\eta}, \tag{4.82}$$

and a new field $y_{\mathbf{k}} \equiv \sqrt{c_{\rm s}}\chi_{\mathbf{k}}$, so that the equation of motion for the mode function in Eq. (4.46) becomes

$$y_{\mathbf{k},qq} + \left[k^2 + \tilde{m}^2_{\rm eff,S}\right] y_{\mathbf{k}} = 0, \tag{4.83}$$

where $,_q = d/dq$ and

$$\tilde{m}^2_{\rm eff,S} \equiv \frac{m^2_{\rm eff,S}}{c_{\rm s}^2} - \frac{(\sqrt{c_{\rm s}})_{,qq}}{\sqrt{c_{\rm s}}}. \tag{4.84}$$

The advantage of this change of variables is that the squared sound speed $c_{\rm s}^2$ does not appear in front of the k^2 term in Eq. (4.83), as it does in Eq. (4.46), so that the same methods used for solving for the equation of motion of $\chi_{\mathbf{k}}$ in the case where $c_{\rm s}$ is constant can be used to solve for the new mode function $y_{\mathbf{k}}$ as a function of the new time variable q.

In the previous section, the mode function was easily solved because $m^2_{\rm eff,S}$ was inversely proportional to η^2. Accordingly, we reparameterize $\tilde{m}^2_{{\rm eff},S}$ by $\mathcal{B}_{\rm S} \equiv -q^2\tilde{m}^2_{{\rm eff},S}$, so that the equation of motion for $y_{\mathbf{k}}$ becomes

$$y_{\mathbf{k},qq} + \left[k^2 - \frac{\mathcal{B}_{\rm S}}{q^2}\right] y_{\mathbf{k}} = 0. \tag{4.85}$$

To obtain a solution where the running of $n_{\rm s}$ is small, we consider solutions where $\mathcal{B}_{\rm S}$ is nearly constant, in which case Eq. (4.85) would have solution

$$y_k(q) \approx \sqrt{\frac{\pi q}{2}}\left[C_1 H^{(1)}_{\gamma_{\rm S}}(kq) + C_2 H^{(2)}_{\gamma_{\rm S}}(kq)\right], \tag{4.86}$$

where the index $\gamma_{\rm S}$ is

$$\gamma_{\rm S} \equiv \frac{1}{2}\sqrt{1 + 4\mathcal{B}_{\rm S}}. \tag{4.87}$$

As with the previous case, we choose the integration constants C_1 and C_2 to select the lowest energy state when $kq \gg 1$. This coincides with the asymptotic solution in Eq. (4.72) if the change of $c_{\rm s}$ with time is small at some early time when all modes of interest are well within the horizon. Writing the normalization condition in Eq. (4.47) in terms of the mode function $y_{\mathbf{k}}$ and the time variable q gives

$$(y_{\mathbf{k}})_{,q}\, y^*_{\mathbf{k}} - y_{\mathbf{k}}(y_{\mathbf{k}})^*{}_{,q} = -2i. \tag{4.88}$$

With these choices, the mode function $y_{\mathbf{k}}$ evolves as

$$y_k(q) = \sqrt{\frac{\pi q}{2}} H^{(1)}_{\gamma_{\rm S}}(kq). \tag{4.89}$$

When dealing with the time variable q, $k|q| \sim 1$ does not necessarily imply $c_{\rm s}k|\eta| \sim 1$. However, in the case when $c_{\rm s}$ varies slowly in time, as is considered in Section 4.6.4, then $k|q| \sim 1$ when $c_{\rm s}k|\eta| \sim 1$, in which case there will be no confusion about what is meant by a mode crossing the horizon.

Analogously with the previous section, on superhorizon scales, the mode function y_k is approximately equal to

$$y_k(q) \approx -\sqrt{\frac{\pi q}{2}}\frac{i\Gamma(\gamma_{\rm S})}{\pi}\left(\frac{kq}{2}\right)^{-\gamma_{\rm S}}, \tag{4.90}$$

at which time the power spectrum for R becomes

$$\mathcal{P}_R(k) \approx \frac{c_{\rm s}\Gamma^2(\gamma_{\rm S})4^{\gamma_{\rm S}}}{8\pi^3(1+w)}\frac{l^2k^{3-2\gamma_{\rm S}}q^{1-2\gamma_{\rm S}}}{a^2}, \tag{4.91}$$

which implies that the scalar spectral index $n_{\rm s}$ is now given by

$$n_{\rm s} = 4 - 2\gamma_{\rm S}. \tag{4.92}$$

It is trivial to check that in the case that the sound speeds and equation of state are constant, the above equations simplify to those given in the previous section. However, if we can find time-varying sound speeds and/or equation of state such that $\mathcal{B}_{\rm S}$ is approximately constant, the bounds on the index $\gamma_{\rm S}$ may be extended to include red-tilted scalar spectra.

4.6.4 Slowly Varying Sound Speeds and Equation of State

From the above considerations, we wish to find a parameterization of the sound speeds and equation of state that allows for a small variation in time in such a way that results in $\mathcal{B}_{\rm S}$ being approximately constant and allows the scalar spectral index to be less than one.

We will use the variable ϵ, which coincides with the slow-roll variable from standard inflationary scenarios, defined by

$$\epsilon \equiv -\frac{\dot{H}}{H^2} = 1 - \frac{\mathcal{H}'}{\mathcal{H}^2}, \tag{4.93}$$

so that $w = -1 + \frac{2}{3}\epsilon$. In this context, ϵ simply parameterizes the departure of w from -1. We parameterize ϵ as $\epsilon(\eta) = \epsilon_0 + f_\epsilon(\eta)$ for some slowly varying function $f_\epsilon(\eta)$. We write the time dependence of f_ϵ as

$$\frac{d\ln f_\epsilon}{d\ln\eta} = -\tau_\epsilon. \tag{4.94}$$

If we assume that $|\tau_\epsilon| \ll 1$, then $\epsilon(\eta)$ will be given by

$$\epsilon(\eta) \approx \epsilon_0 + \epsilon_1(\eta/\eta_*)^{-\tau_\epsilon} \tag{4.95}$$

for some reference time η_*. With this time dependence, w slowly varies near $-1 + \frac{2}{3}\epsilon_0$ for some time, but at some later time, it will evolve at a more rapid pace. We will choose the reference time η_* to be the end of inflation to ensure that the time dependence of w is small during inflation. We also allow c_s to vary in time and use a parameterization analogous to the one used for ϵ, so that

$$c_\mathrm{s}(\eta) \approx c_{\mathrm{s}0} + c_{\mathrm{s}1}(\eta/\eta_*)^{-\tau_\mathrm{s}}, \tag{4.96}$$

where $|\tau_\mathrm{s}| \ll 1$. With this parameterization, our new time variable q is

$$q(\eta) = -\eta\left[c_{\mathrm{s}0} + \frac{c_{\mathrm{s}1}(\eta/\eta_*)^{-\tau_\mathrm{s}}}{1-\tau_\mathrm{s}}\right]. \tag{4.97}$$

The requirement that $\mathcal{B}_\mathrm{S}$ be approximately constant for $\eta \leq \eta_*$ is met so long as both τ_ϵ and τ_s are sufficiently small. Note that solutions where w departs significantly from -1 are valid since ϵ_0 is not required to be small. If we desire w to stay close to -1, we would add the restrictions $|\epsilon_0| \ll 1$ and $|\epsilon_1| \ll 1$. To obtain a slightly red-tilted scalar spectrum, we will want the sound speeds and equation of state to evolve near constant values that result in a nearly scale-invariant (blue-tilted) scalar spectrum. Therefore, from Section 4.6.1, we will want to consider solutions where at least one of ϵ_0 and $c_{\mathrm{s}0}$ are small.

For the rest of this chapter, we restrict ourselves to cases where $|\epsilon_1| \ll 1$, in which case Eq. (4.3) can be used to solve for $\mathcal{H}$ and subsequently the scale factor a, which to linear order in τ_ϵ and ϵ_1 is

$$\mathcal{H} \approx -\frac{1-\epsilon_0+\epsilon_1}{\eta(1-\epsilon_0)^2}, \qquad a \approx a_*(\eta/\eta_*)^{-\frac{1-\epsilon_0+\epsilon_1}{(1-\epsilon_0)^2}}, \tag{4.98}$$

where a_* is the value of the scale factor at $\eta = \eta_*$. In the case where $\epsilon_0 = c_{\mathrm{s}0} = 0$, $\mathcal{B}_\mathrm{S}$ to first order in our small parameters is found to be

$$\mathcal{B}_\mathrm{S} \approx 2 + \frac{15}{2}\tau_\mathrm{s} + \frac{3}{2}\tau_\epsilon - 3c_{\mathrm{s}1}^2\epsilon_1, \tag{4.99}$$

and the scalar spectral index n_s becomes

$$n_\mathrm{s} \approx 1 - 5\tau_\mathrm{s} - \tau_\epsilon + 2c_{\mathrm{s}1}^2\epsilon_1, \tag{4.100}$$

from which we see can yield a red-tilted scalar spectrum (see the first three rows of Table 4.1 for examples). Note that although $c_s \to 0$ if $c_{\mathrm{s}0} = 0$

in the limit where $\eta \to -\infty$, $c_{\rm s}$ at the beginning of inflation will not be significantly different from its value at the end of inflation ($c_{\rm s}(\eta_*) = c_{\rm s1}$) as long as $|\tau_{\rm s}| \ll 1$ and the number of e-folds of inflation is modest (i.e. 100's of e-folds).

As previously stated, we can still estimate the running of $n_{\rm s}$ by conventional means where relevant quantities are evaluated at horizon crossing. As horizon crossing occurs when $c_{\rm s}k|\eta| \sim 1$, at horizon crossing $d\ln k \sim -(\eta^{-1} + c_{\rm s}'/c_{\rm s})^{-1} d\eta$, so that

$$\frac{dn_{\rm s}}{d\ln k} \sim \left(\frac{1}{\eta} + \frac{c_{\rm s}'}{c_{\rm s}}\right)^{-1} \frac{d}{d\eta}\sqrt{1+4\mathcal{B}_{\rm S}}\Big|_{c_{\rm s}k|\eta|\sim 1}. \tag{4.101}$$

For $\epsilon_0 = c_{\rm s0} = 0$, the running to second order in our small parameters is

$$\frac{dn_{\rm s}}{d\ln k} \sim 2c_{\rm s1}^2\epsilon_1(\tau_\epsilon + 2\tau_{\rm s}), \tag{4.102}$$

from which we see that the running of $n_{\rm s}$ vanishes at first order.

As in the constant sound speeds and equation of state case, we find that we are not restricted to very small values of ϵ_0 and $c_{\rm s0}$, although for brevity we will not explicitly write out $n_{\rm s}$ and its running and instead illustrate through numerical examples. In fact, we can formally find solutions with w varying slowly near values up to $-1/3$, corresponding to values of ϵ_0 just below 1, that result in a slightly red-tilted scalar spectrum with a small running of its spectral index, although, as previously mentioned, achieving the necessary number of e-folds of inflation becomes increasingly challenging for values of w very far from -1.

As an example, choosing $\epsilon_0 = 1/4$ so that w varies close to $-5/6$ and taking the other parameters to have the values listed in the fourth row of Table 4.1, we obtain a scalar spectral index $n_{\rm s} \approx 0.96$ and running $dn_{\rm s}/d\ln k \sim -10^{-5}$. Another example where w varies near $-2/3$ is given in the fifth row of Table 4.1. Note that the reheat temperature used in this example is significantly lower compared to the other examples listed in the table to allow for the required number of e-folds of inflation. Choosing $c_{\rm s0}$ to be nonzero instead of ϵ_0, with $c_{\rm s0} = 0.15$ and the values in the last row of Table 4.1 yields a scalar spectral index of $n_{\rm s} \approx 0.96$ and running $dn_{\rm s}/d\ln k \sim -10^{-3}$.

We conclude this section by writing the power spectrum for R at the end

of inflation for superhorizon modes, which from Eq. (4.91), is

$$\mathcal{P}_R(k) \approx \frac{3\Gamma^2(\gamma_{\rm S})(c_{\rm s0}+c_{\rm s1})}{8\pi^3(\epsilon_0+\epsilon_1)} \times \left[\frac{c_{\rm s0}(1-\epsilon_0+\epsilon_1)+c_{\rm s1}(1+\epsilon_1+\tau_{\rm s}-\epsilon_0(1+\tau_{\rm s}))}{2(1-\epsilon_0)^2}\right]^{1-2\gamma_{\rm S}} \times l^2(k/a_*)^{3-2\gamma_{\rm S}}H_*^{-1+2\gamma_{\rm S}}, \quad (4.103)$$

where the subscript $*$ denotes evaluation at $\eta=\eta_*$. Since modes can evolve on superhorizon scales, $\mathcal{P}_R(k)$ at the end of reheating may be different from the expression given above. This concern will be addressed in Section 4.8.

4.7 Gravitational Waves

To find the amplitude of gravitational waves produced during inflation, we need solve for the equation of motion of the mode function $X_{\mathbf{k}}$ in Eq. (4.58), which can be solved in an analogous manner to $\chi_{\mathbf{k}}$ in the scalar case. This task will be relatively easy compared to the scalar perturbations, since the tensor perturbations travel with a sound speed equal to unity so there is no need to switch time variables in order to solve the differential equation for $X_{\mathbf{k}}$. Analogous to the scalar case, for the tensor modes we can define

$$\mathcal{B}_{\rm T} \equiv -\eta^2 m_{\rm eff,T}^2. \quad (4.104)$$

Using the parameterizations of the sound speeds and equation of state in Section 4.6.4, $\mathcal{B}_{\rm T}$ to first order in $\tau_{\rm s}$, τ_ϵ, and ϵ_1 with $\epsilon_0=c_{\rm s0}=0$ is

$$\mathcal{B}_{\rm T} \approx 2-3c_{\rm s1}^2\epsilon_1. \quad (4.105)$$

Since $\mathcal{B}_{\rm T}$ is constant to linear order in our small parameters, the tensor mode function in Eq. (4.58) has solution

$$X_k \approx \sqrt{\frac{\pi|\eta|}{2}}\left[C_1H^{(1)}_{\gamma_{\rm T}}(k|\eta|)+C_2H^{(2)}_{\gamma_{\rm T}}(k|\eta|)\right], \quad (4.106)$$

where C_1 and C_2 are new constants of integration and the index $\gamma_{\rm T}$ is

$$\gamma_{\rm T} = \frac{1}{2}\sqrt{1+4\mathcal{B}_{\rm T}}. \quad (4.107)$$

The choice of $C_1=0$ and $C_2=1$ satisfies the normalization condition of Eq. (4.59) and selects the lowest energy state when the mode is well within

c_{s0}	c_{s1}	ϵ_0	ϵ_1	τ_s	τ_ϵ	T_{RH}(GeV)	n_s	$10^9 A_\zeta$	$r(k_p)$	n_T	$dn_s/d\ln k$
0	0.01	0	0.01	0.005	0.015	1.55×10^{13}	0.96	2.43	$\sim 10^{-12}$	-3.1×10^{-4}	$\sim -10^{-5}$
0	0.8	0	0.01	0.0086	0.01	3.69×10^{15}	0.96	2.43	0.002	0.013	$\sim 10^{-4}$
0	0.3	0	0.001	0.005	0.015	6.06×10^{14}	0.96	2.43	$\sim 10^{-6}$	1.5×10^{-4}	$\sim 10^{-6}$
0	0.001	0.25	0.01	0.0078	0.0078	3.05×10^{12}	0.96	2.43	$\sim 10^{-15}$	-0.0002	$\sim -10^{-5}$
0	10^{-8}	0.5	0.01	0.008	0.001	3.84×10^{6}	0.96	2.43	$\sim 10^{-40}$	-4.1×10^{-5}	$\sim -10^{-7}$
0.15	0.05	0	0.01	0.024	0.01	1.79×10^{14}	0.96	2.43	$\sim 10^{-6}$	6.1×10^{-4}	$\sim -10^{-3}$

Table 4.1: Examples of choice of parameters for slowly varying sound speeds and equation of state. For cases where $c_{s0} = \epsilon_0 = 0$, n_s, n_T, and r are calculated directly from Eqs. (4.100), (4.110), and (4.131), respectively. In examples where either $\epsilon_0 \neq 0$ or $c_{s0} \neq 0$, n_s and n_T are calculated by first computing $\mathcal{B}_S$ or $\mathcal{B}_T$, as defined in Sections 4.6.3 and 4.7, respectively. The power spectrum for ζ has been parameterized as $P_\zeta(k) = A_\zeta(k/k_p)^{n_s-1}$ where $k_p = 0.002\,\mathrm{Mpc}^{-1}$. In calculating A_ζ and r, the model of a rapid decay into radiation was assumed to end inflation and reheat the Universe. For cases with $c_{s0} = \epsilon_0 = 0$, A_ζ and r are calculated using Eqs. (4.126) and (4.131), respectively. For examples with either $\epsilon_0 \neq 0$ or $c_{s0} \neq 0$, A_ζ and r are calculated using the power spectra given in Eqs. (4.103) and (4.111), respectively, along with the relevant equations found in Section 4.8 to relate R, h^T just before the end of inflation to ζ, h^T following the end of inflation. All examples use $g_{RH} = 106.75$.

the horizon. When modes are well outside the horizon, the mode function can be approximated as

$$X_k \approx \sqrt{\frac{\pi|\eta|}{2}} \frac{i\Gamma(\gamma_{\rm T})}{\pi} \left(\frac{k|\eta|}{2}\right)^{-\gamma_{\rm T}}. \tag{4.108}$$

Using this approximation for the mode function with Eq. (4.61) yields the power spectrum for gravitational waves on superhorizon scales

$$\mathcal{P}_{\rm T}(k) \approx \frac{6 \times 2^{2\gamma_{\rm T}-1} l^2 \Gamma^2(\gamma_{\rm T})}{\pi^3} \frac{k^{3-2\gamma_{\rm T}} |\eta|^{1-2\gamma_{\rm T}}}{a^2}. \tag{4.109}$$

When $\epsilon_0 = c_{\rm s0} = 0$, the tensor spectral index $n_{\rm T} = d\ln\mathcal{P}_{\rm T}/d\ln k$ is

$$n_{\rm T} \approx 2c_{\rm s1}^2 \epsilon_1. \tag{4.110}$$

At the end of inflation, the tensor power spectrum can be written as

$$\mathcal{P}_{\rm T}(k) \approx \frac{6\Gamma^2(\gamma_{\rm T})}{\pi^3} \left(\frac{1-\epsilon_0+\epsilon_1}{2(1-\epsilon_0)^2}\right)^{1-2\gamma_{\rm T}} \times l^2 (k/a_*)^{3-2\gamma_{\rm T}} H_*^{-1+2\gamma_{\rm T}}. \tag{4.111}$$

4.8 End of Inflation and Reheating

Unlike in cases where the superhorizon evolution of modes is small, the details of the end of inflation will affect the amplitude of modes on superhorizon scales, although the time evolution of all superhorizon modes will be the same. There are many possibilities for ending inflation and reheating within this model. One possibility is once all modes of interest are on superhorizon scales having w increase rapidly so that it surpasses $-1/3$ and the Universe stops inflating. At some later point, the solid can lose its rigidity, at which point the superhorizon evolution will be small and the details of reheating will not affect modes on superhorizon scales. Another possibility, which we will examine in more detail, is that inflation ends with the decay of the elastic solid.

In general, ζ_k and R_k will not be equal during inflation on superhorizon scales. We can easily find the relationship between ζ and R by using Eqs. (4.22) and (4.35), which is

$$\zeta = \frac{dP/d\rho}{c_{\rm s}^2} R - \frac{1}{3c_{\rm s}^2 \mathcal{H}} R', \tag{4.112}$$

from which we see that in general, even on superhorizon scales, $\zeta_k \neq R_k$. After inflation ends and the rigidity vanishes, the superhorizon evolution

will be small and $\zeta_k \approx R_k$, but the change in ζ_k and R_k must be tracked through the transition that ends inflation.

We now consider the case where the elastic solid rapidly decays into a perfect fluid to end inflation. Following Ref. [65], we can write the total stress-energy tensor as

$$T^{\mu\nu} = \sum_\alpha T^{\mu\nu}_{(\alpha)}, \tag{4.113}$$

where $T^{\mu\nu}_{(\alpha)}$ is the stress-energy tensor of component $\alpha = \{e, f\}$, with e and f denoting the elastic solid and perfect fluid decay product, respectively. For this analysis, it will be more convenient to use the coordinate time t instead of the conformal time. While the local energy-momentum transfer 4-vector $Q^\nu_{(\alpha)}$ for each species can be nonzero, so that

$$\nabla_\mu T^{\mu\nu}_{(\alpha)} = Q^\nu_{(\alpha)}, \tag{4.114}$$

we must have $\sum_\alpha Q^\nu_{(\alpha)} = 0$ so that the total stress-energy tensor is covariantly conserved.

For the scalar perturbations, we can define a ζ_α variable for each substance α, defined by

$$\zeta_\alpha \equiv \psi + H\frac{\delta\rho_\alpha}{\dot\rho_\alpha} \tag{4.115}$$

that is related to the total ζ by

$$\zeta = \sum_\alpha \frac{\dot\rho_\alpha}{\dot\rho}\zeta_\alpha. \tag{4.116}$$

Similarly, the variable R_α for each substance, defined by

$$R_\alpha = \mathcal{H}v_\alpha + \psi, \tag{4.117}$$

is related to the total R by

$$R = \sum_\alpha \frac{\rho_\alpha + P_\alpha}{\rho + P} R_\alpha. \tag{4.118}$$

The total energy density ρ and pressure P are given by $\rho = \sum_\alpha \rho_\alpha$ and $P = \sum_\alpha P_\alpha$, while the entropy perturbation between substances α and β is

$$\mathcal{S}_{\alpha\beta} \equiv 3(\zeta_\alpha - \zeta_\beta) \tag{4.119}$$

and its relative velocity perturbation is given by

$$v_{\alpha\beta} \equiv v_\alpha - v_\beta = \frac{R_\alpha - R_\beta}{\mathcal{H}}. \tag{4.120}$$

For the decay of the elastic solid to a perfect fluid, the energy-momentum transfer is given by

$$Q_e^\nu = -Q_f^\nu = -\Gamma g^{\nu\beta} u_\beta \rho_e (1 + w_e), \tag{4.121}$$

where u_β is the total velocity 4-vector of the elastic solid and the perfect fluid and Γ is the decay rate of the elastic solid into the fluid (not to be confused with the gamma function used previously).

In the current case, where a single substance is decaying into another single substance, we expect that entropy perturbations will not be generated in the decay. In general, the evolution of the entropy perturbation between the elastic solid and the fluid is given by

$$\begin{aligned}(\dot{\mathcal{S}}_{ef})_k = &\left[\frac{\dot{Q}_e}{\dot{\rho}_f} + \frac{Q_e}{2\rho}\left(\frac{\dot{\rho}_f}{\dot{\rho}_e} - \frac{\dot{\rho}_e}{\dot{\rho}_f}\right)\right](\mathcal{S}_{ef})_k \\ &+ \frac{k^2}{a^2 H}\left[\left(1 - \frac{Q_f}{\dot{\rho}_f}\right)(R_f)_k - \left(1 - \frac{Q_e}{\dot{\rho}_e}\right)(R_e)_k\right],\end{aligned} \tag{4.122}$$

where $Q_\alpha = Q^0_{(\alpha)}$ is the background value of the time component of the energy-momentum-transfer 4-vector. We refer to Appendix A.2 for the explicit form of the background and perturbation equations used to derive this relation. We can see that if on superhorizon scales the entropy perturbation vanishes at some time, the entropy perturbation will stay approximately constant past this time. If the decay is rapid ($\Gamma \gg H$), then any preexisting entropy perturbations will quickly be driven to zero at the very beginning of the decay. Therefore, for times of interest, we set the entropy perturbation to zero. With no entropy perturbation, $\zeta_e = \zeta_f$ and ζ will be continuous across the decay, when it changes from $\zeta = \zeta_e$ before the decay to $\zeta = \zeta_f$ after the decay. On the other hand, we do not expect the relative velocity perturbation to be zero during the decay, so in general R will change rapidly during the decay as it goes from $R = R_e$ before the decay to $R = R_f \approx \zeta_f$ on superhorizon scales after the decay.[39] In this light, we will follow ζ instead of R from the end of inflation into radiation domination. The point where the elastic solid decays presumably occurs when a macroscopic quantity such as the energy density or pressure reaches a critical value. Although, due to the inhomogeneities in these fields, the decay many not occur at the same value of the scale factor at every location, since these perturbations are small and the superhorizon evolution is not drastic when scale-invariant spectra are

[39] The assertions that ζ is continuous and R is discontinuous through the decay were verified numerically.

produced, the approximation of the decay occurring uniformly at $a = a_*$ should be a reasonable assumption.

To find the postinflationary scalar power spectrum, we match ζ and its first derivative at the time of the decay and will assume that the decay product is radiation. During radiation domination, on large scales $\zeta_{\mathbf{k}}$ evolves as

$$\zeta_{\mathbf{k}}'' + 2\mathcal{H}\zeta_{\mathbf{k}}' \approx 0, \tag{4.123}$$

which has solution

$$\zeta_{\mathbf{k}}(a \geq a_*) \approx \zeta_{\mathbf{k}*} + \frac{\zeta_{\mathbf{k}*}'}{\mathcal{H}_*}\left(1 - \frac{a_*}{a}\right), \tag{4.124}$$

where integration constants have been chosen so that $\zeta_{\mathbf{k}}$ and $\zeta_{\mathbf{k}}'$ are continuous over the decay. From Eq. (4.124), we see that $\zeta_{\mathbf{k}}$ has both a constant and decaying mode during radiation domination. Within a few e-folds after the decay, the decaying mode becomes negligible, as seen in Fig. 4.2.

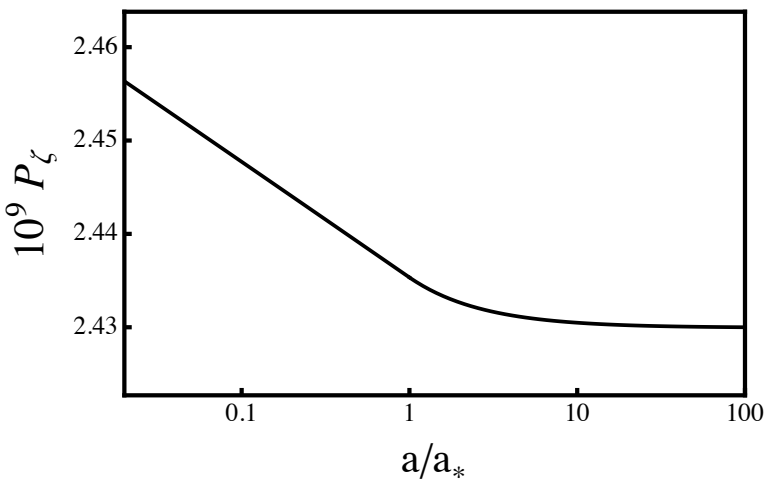


Figure 4.2: The power spectrum of ζ during the decay of the elastic solid to radiation for a superhorizon mode. During inflation, sound speed and equation of state parameters are chosen to be those listed in the third row of Table 4.1.

Using Eq. (4.112) to calculate ζ from R before the decay, for the case where $\epsilon_0 = c_{\rm s0} = 0$, we find that the relationship between $\mathcal{P}_\zeta(k, a \gg a_*)$ and $\mathcal{P}_R(k, a_*)$ right before the decay, given in Eq. (4.103), is

$$\frac{P_\zeta(k, a \gg a_*)}{\mathcal{P}_R(k, a \to^- a_*)} = \left|\frac{3 - 4(3 + 2c_{\rm s1}^2)\epsilon_1 + 4\tau_{\rm s} + 2\tau_\epsilon}{3c_{\rm s1}^4}\right|. \tag{4.125}$$

As $c_{\rm s1} < 1$ (but not necessarily $c_{\rm s1} \ll 1$), typically ζ_k will be larger than R_k, in which case if ζ_k is in the linear regime, so will R_k.

We can express the power spectrum of ζ in terms of the pivot scale $k_p = 0.002\,\mathrm{Mpc}^{-1}$ as $P_\zeta(k) = A_\zeta (k/k_p)^{n_{\rm s}-1}$. Assuming rapid reheating, for the $c_{\rm s0} = \epsilon_0 = 0$ case A_ζ can be written as

$$A_\zeta = 10^{-22}\Delta \left| \frac{c_{\rm s1}^{n_{\rm s}-6}}{\epsilon_1} \right| \left(\frac{g_{\rm RH}}{106.75} \right)^{\frac{7-n_{\rm s}}{6}} \left(\frac{T_{\rm RH}}{10^{13}\,\mathrm{GeV}} \right)^{5-n_{\rm s}}, \tag{4.126}$$

where $T_{\rm RH}$ is the reheat temperature and $g_{\rm RH}$ is the effective number of relativistic degrees of freedom contributing to the energy density at reheating.[40] In the above expression, Δ is a constant that is given in detail in Appendix A.3. For nearly scale-invariant scalar spectra, Δ is of order unity. When the sound speeds and equation of state are constant, A_ζ is given by the same expression except with the substitutions $c_{\rm s1} \to c_{\rm s}$ and $\epsilon_1 \to \epsilon$.

Tracking the effect of a rapid decay to radiation on the tensor modes is straightforward, since as seen in Eq. (4.12) the tensor modes will only be affected by the change in anisotropic stress. Accordingly, we match the tensor perturbations and their first derivatives across the decay as the anisotropic stress vanishes. From Appendix A.1, the equation of motion for the tensor perturbations is given by [41]

$$(h_{\mathbf{k}}^{\rm T})'' + 2\mathcal{H}(h_{\mathbf{k}}^{\rm T})' + (k^2 + 4c_{\rm v}^2\beta)h_{\mathbf{k}}^{\rm T} = 0. \tag{4.127}$$

Since there is negligible rigidity in the radiation fluid, the transverse sound speed will vanish after the decay. During radiation domination, superhorizon tensor modes have the same approximate evolution equation as the scalar modes in Eq. (4.123); therefore, matching the tensor modes and their first derivatives across the decay will have the same form as the scalar mode solution in Eq. (4.124), so that

$$h_{\mathbf{k}}^T(a \geq a_*) \approx h_{\mathbf{k}*}^T + \frac{(h_{\mathbf{k}*}^T)'}{\mathcal{H}_*}\left(1 - \frac{a_*}{a}\right). \tag{4.128}$$

Using the above equation with Eq. (4.109), the postinflationary tensor power spectrum of superhorizon modes is related to its value at the end of inflation

[40] In Eq. (4.126), we have assumed that all relativistic species are in thermodynamic equilibrium at $T_{\rm RH}$ so that $g_{\rm RH} = g_{s\rm RH}$, where g_s is the effective number of relativistic degrees of freedom contributing to the entropy density, and we make this assumption throughout this chapter.

[41] In this section, $h^{\rm T}$ will label a component of the tensor $h_{ij}^{\rm T}$.

by

$$\frac{\mathcal{P}_{\mathrm{T}}(k, a \gg a_*)}{\mathcal{P}_{\mathrm{T}}(k, a_*)} = \frac{(1-2\gamma_{\mathrm{T}})^2(\epsilon_0-1)^4}{4(1-\epsilon_0+\epsilon_1)^2}, \tag{4.129}$$

where the tensor power spectrum at the end of inflation in given in Eq. (4.111).

The tensor-to-scalar ratio $r = \mathcal{P}_{\mathrm{T}}/\mathcal{P}_\zeta$ after the decaying modes mentioned above are negligible for the $\epsilon_0 = c_{\mathrm{s}0} = 0$ case can now be found to be

$$r \approx 16 c_{\mathrm{s}1}^5 \epsilon_1 (a_* H_*/k)^{n_{\mathrm{s}}-1-n_{\mathrm{T}}}. \tag{4.130}$$

The tensor-to-scalar ratio is mildly dependent on the (physical) wavenumber and Hubble rate at the end of inflation, since in general the scalar and tensor modes have different tilts. At the pivot scale k_p, the tensor-to-scalar ratio is

$$r(k_p) \approx 1.94 \times 10^{22.9(n_{\mathrm{s}}-0.96-n_{\mathrm{T}})} c_{\mathrm{s}1}^5 \epsilon_1 \left(\left(\frac{g_{\mathrm{RH}}}{106.75} \right)^{1/6} \frac{T_{\mathrm{RH}}}{10^{13}\,\mathrm{GeV}} \right)^{n_{\mathrm{s}}-1-n_{\mathrm{T}}}. \tag{4.131}$$

For this case, the tensor-to-scalar ratio is suppressed by the $c_{\mathrm{s}1}^5$ term, so for small values of $c_{\mathrm{s}1}$ the tensor-to-scalar ratio will be highly suppressed. For example, with the parameter values in the first row of Table 4.1 with $c_{\mathrm{s}1} = 0.01$, the tensor-to-scalar ratio is $r \sim 10^{-12}$. However, if $c_{\mathrm{s}1}$ assumes a higher value, then this suppression is more moderate, illustrated by the values in the second row of Table 4.1, which with $c_{\mathrm{s}1} = 0.8$ yields a tensor-to-scalar ratio of $r = 0.002$.

As in Section 4.6.4, we do not write out an explicit expression for the scalar amplitude or tensor-to-scalar ratio for cases when either $c_{\mathrm{s}0}$ or ϵ_0 are nonzero and will soon illustrate with numerical examples instead. But before doing this, we can gain some insight by examining the case when the sound speeds and equation of state are constant. As previously mentioned, a nearly scale-invariant scalar spectrum can be produced when w is far from -1 if c_{s} is sufficiently small. However, in this case there are added considerations as the energy density changes significantly during the course of inflation. If inflation lasts just long enough to solve the 'horizon problem' and Λ is the energy scale at the beginning of inflation, A_ζ will be given by

$$A_\zeta \approx 10^{\frac{-109-n_{\mathrm{s}}(13.1+10.4\epsilon)+30\epsilon}{2-\epsilon}} \Delta \left| \frac{c_{\mathrm{s}}^{n_{\mathrm{s}}-6}}{\epsilon} \right| \times \left(\frac{g_{\mathrm{RH}}}{106.75} \right)^{\frac{n_{\mathrm{s}}-1+18\epsilon-4n_{\mathrm{s}}\epsilon}{6(2-\epsilon)}} \left(\frac{\Lambda}{m_{\mathrm{p}}} \right)^{\frac{5-n_{\mathrm{s}}}{1+\epsilon/(2(1-\epsilon))}} \tag{4.132}$$

for c_s, ϵ constant, where m_p is the Planck mass. For solutions with $n_s \sim 1$, if ϵ is raised to higher values, A_ζ may drop significantly. If Λ is bounded by the Planck scale, for large values of ϵ (w far from -1), to keep $A_\zeta \sim 10^{-9}$, c_s may have to be fine-tuned to a very small value. Alternatively, this fine-tuning may be averted if one is comfortable having inflation start at super-Planckian scales.

This issue is demonstrated for the slow-varying sound speeds case in the fifth row in Table 4.1, in which w varies close to $-2/3$. To attain the same scalar amplitude as was used in the other examples in Table 4.1 and have inflation start at or below the Planck scale and last for a sufficiently long duration, c_s had to assume the extremely small value of $\sim 10^{-8}$. In the fourth row of Table 4.1, $w \simeq -5/6$, so while the departure from -1 is still significant, we find c_s can assume much larger values near 10^{-3} — and smaller values of w may have correspondingly larger values of c_s.

However, even if we do not tune c_s to be very small, it is theoretically interesting that we can achieve a scale-invariant spectrum far from $w \simeq -1$ even if the amplitude of perturbations are not large enough to match observations. For instance, the parameter values $w = -2/3$ and $c_s = 1/10$ produce a slightly blue-tilted spectrum for both scalar and tensor perturbations (with $n_s \approx 1.04$ and $n_T \approx 0.04$). Understanding the physical origin of this near scale invariance is an extremely interesting question that might give new insight into the physics of horizons.

Lastly, we compute the value of the tensor-to-scalar ratio for our examples where either c_{s0} or ϵ_0 are nonzero. Using the values in the fourth row of Table 4.1, where w varies slowly near $-5/6$, gives a tensor-to-scalar ratio at the pivot scale of $r \sim 10^{-15}$. We again see that the tensor-to-scalar ratio is highly suppressed by c_{s1}. With the values listed in the last row of Table 4.1 with $c_{s0} = 0.15$ yields $r \approx 10^{-6}$ and the suppression of r is more moderate.

4.9 Conclusion

By having a sufficiently rigid structure, a relativistic elastic solid is capable of driving an inflationary stage in the early Universe. In the case of constant sound speeds c_s and c_v and equation of state w, a blue-tilted scalar power spectrum is produced. Allowing the sound speeds and equation of state to vary slowly in time can result in a red-tilted scalar power spectrum with small running. When c_s is small, the tensor-to-scalar ratio will be highly suppressed, but can attain larger values for higher values of c_s.

An interesting feature of this model is that perturbations evolve on su-

perhorizon scales, even in the absence of nonadiabatic pressure. The superhorizon evolution results from the shear stresses in the solid, where the propagation of a single perturbative mode causes an anisotropic pressure. Because of this anisotropy, when smoothed on a superhorizon scale, different locations in the Universe will not share the same FRW evolution, as they do when both shearing stresses and nonadiabatic pressures are absent. As a result, the perturbations do not 'freeze-out' soon after horizon crossing and consequently, the details of the end of inflation can impact both scalar and tensor power spectra for modes that are on superhorizon scales when inflation ends. The case of a rapid decay of the elastic solid into radiation was explored as a specific example.

Finally and intriguingly, we find this model allows for w to vary slowly near values that are significantly different from -1 and can find cases where this produces nearly scale-independent scalar and tensor power spectra despite being far from the de Sitter regime. This is surprising and unexpected and it would be interesting to determine the underlying physical reason for this phenomena.

Chapter 5

The Physics of 21-cm Radiation

5.1 Introduction

The remaining chapters of this thesis will deal with the cosmic 21-cm signal emitted by neutral hydrogen. The nature of the cosmic 21-cm signal changes drastically throughout the evolution of the Universe, most notably during the reionization of hydrogen at redshifts $z \sim 6-10$. In this chapter, we review some of the basic properties of the 21-cm signal and give a rough description of its evolution. In addition, we will conclude this chapter with a brief introduction to measurement techniques used with interferometric radio telescopes.

5.2 Properties of 21-cm Radiation

5.2.1 The Brightness Temperature

As the first step for describing the basic properties of the 21-cm signal, we write the equation for radiative transfer [68]

$$\frac{dI_\nu}{ds} = \frac{h\nu}{4\pi}\phi(\nu)[n_1 A_{10} - (n_0 B_{01} - n_1 B_{10})I_\nu], \tag{5.1}$$

where I_ν is the specific intensity, $\phi(\nu)$ is the line profile (normalized by $\int \phi(\nu)d\nu = 1$), ds is a proper length element, and n_0 and n_1 are the number densities for the unexcited and excited hyperfine states with degeneracies g_0 and g_1, respectively (in the present case $g_0 = 1$ and $g_1 = 3$). A_{10}, B_{10}, and B_{01} are the Einstein coefficients for spontaneous emission, stimulated emission, and absorption, respectively. The Einstein coefficients are related to one another by $B_{10}/A_{10} = c^2/2h\nu^3$ and $B_{10}/B_{01} = g_0/g_1$, where $A_{10} = 2.85 \times 10^{-15}\,\mathrm{s}^{-1}$. The differential equation in Eq. (5.1) is easily solved and can be expressed in the Rayleigh-Jeans limit (so that I_ν can be written in

terms of a brightness temperature $T_b(\nu) = c^2 I_\nu / 2k_\mathrm{b}\nu^2)$ as

$$T_b'(\nu) = T_\mathrm{S}(1 - e^{-\tau_\nu}) + T_\mathrm{R}'(\nu)e^{-\tau_\nu}, \tag{5.2}$$

where T_R' is the background radiation brightness, $\tau_\nu = \int ds\alpha_\nu$ is the optical depth, and α_ν is the absorption coefficient given by

$$\alpha_\nu = \frac{h\nu}{4\pi}\phi(\nu)(n_0 B_{01} - n_1 B_{10}). \tag{5.3}$$

The excitation temperature T_S for the 21-cm transition, known as the spin temperature, specifies the relative number density of excited to unexcited states by

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{-T_*/T_\mathrm{S}}, \tag{5.4}$$

where the energy difference for the hyperfine transition is $E_{01} = 5.9\times10^{-6}\,\mathrm{eV}$ with equivalent temperature $T_* = E_{10}/k_\mathrm{B} = 68\,\mathrm{mK}$ [19, 69]. In all situations that we will consider we will have $T_\mathrm{S} \gg T_*$ so that $n_1/n_0 \approx 3$ and thus stimulated emission will be an important process.

We can now write the optical depth as

$$\tau_\nu = \int ds\sigma_{10}\phi(\nu)n_0(1 - e^{-T_*/T_\mathrm{S}}), \tag{5.5}$$

where $\sigma_{01} = 3c^2 A_{10}/8\pi\nu^2$. The integral in Eq. (5.5) can be evaluated using $ds = (c/aH)da$, so the brightness temperature of the 21-cm signal measured against the cosmic microwave background (CMB) at redshift z is given by [19]

$$\begin{aligned}\delta T_\mathrm{b}(z) &= \frac{T_\mathrm{S} - T_\gamma}{1+z}(1 - e^{-\tau_{\nu_0}}) \\ &\approx 27 x_\mathrm{HI}(1+\delta)\left(1 - \frac{T_\gamma}{T_\mathrm{S}}\right)\left(\frac{1+z}{10}\frac{0.15}{\Omega_\mathrm{m}h^2}\right)^{1/2} \\ &\times \left(\frac{\Omega_\mathrm{b}h^2}{0.023}\right)\left(\frac{H}{H + dv_\parallel/dr_\parallel}\right)\,\mathrm{mK},\end{aligned} \tag{5.6}$$

where τ_{ν_0} is the optical depth at the 21-cm frequency ν_0, T_γ is the CMB temperature, and $dv_\parallel/dr_\parallel$ is the comoving velocity gradient along the line of sight. We have expressed $\delta T_\mathrm{b}(z)$ in terms of the neutral hydrogen fraction $x_\mathrm{HI} = n_\mathrm{HI}/n_\mathrm{H}$, where $n_\mathrm{H} = \bar{n}_\mathrm{H}(1+\delta)$ and δ is the overdensity.

An important feature of the 21-cm brightness temperature is that it saturates in emission when $T_\mathrm{S} \gg T_\gamma$, a state which is expected from the

time of reionization onwards. In this case we can safely drop the T_γ/T_S term in Eq. (5.6), which makes δT_b independent of the spin temperature. This simplifies the situation greatly as the spin temperature is often difficult to calculate.

5.2.2 The Spin Temperature

For the pre-reionization 21-cm signal, we require the value of the spin temperature. In cosmological contexts, there are three main sources that can affect the spin temperature: the CMB temperature from the absorption of or stimulated emission from CMB photons, the kinetic temperature of the surrounding gas via collisions, and ultraviolet (UV) fields via the Wouthuysen-Field (WF) mechanism [70, 71]. In the absence of the latter two effects, the spin temperature reaches thermal equilibrium with the CMB temperature on a timescale much shorter than those relevant for cosmology. In this case, $T_\mathrm{S} \approx T_\gamma$ and no 21-cm signal can be observed. To observe the 21-cm signal, the spin temperature must depart from the CMB temperature by means of collisions or the WF mechanism.

The manner in which the spin temperature couples to the kinetic temperature of the surrounding gas by collisional excitations is a familiar process in physics, while the WF mechanism is somewhat more obscure. The WF mechanism describes the process where hyperfine states are mixed by way of the absorption and subsequent emission of a Lyman-α photon. Selection rules allow transitions between 1S and 2P hyperfine levels where the electron when returned to the 1S state is in a different hyperfine level then it was before the absorption of the Lyman-α photon.[42] This process is illustrated in Fig 5.1.

[42]Similar transitions are possible with higher Lyman levels, although the effect of such transitions is negligible compared to the Lyman-α transition. However, transitions to higher Lyman levels may be important from cascades through the 2P levels [19].

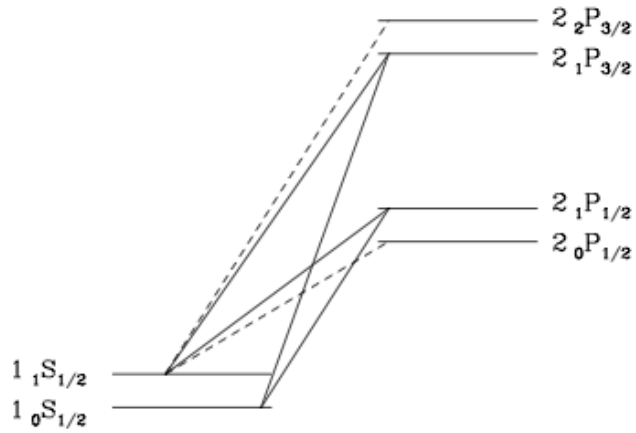


Figure 5.1: Hyperfine levels relevant for the WF mechanism. Transitions using solid lines change the $1S$ hyperfine levels while the dashed lines do not. Figure from Ref. [72].

As the time scales of the aforementioned processes are all much shorter than cosmological time scales, we can safely assume equilibrium and thus have the balanced equation

$$n_0(C_{01} + P_{01} + B_{01}I_{\rm CMB}) = n_1(C_{10} + P_{10} + A_{10} + B_{10}I_{\rm CMB}), \tag{5.7}$$

where C_{01}, C_{10} and P_{01}, P_{10} are the excitation and de-excitation rates via collisions and the WF mechanism, respectively, and $I_{\rm CMB}$ is the specific intensity of the CMB. The ratio between excitation and de-excitation rates by means of collisions is given by the kinetic temperature of the surrounding gas $T_{\rm K}$ as

$$\frac{C_{01}}{C_{10}} = \frac{g_1}{g_0} e^{-T_*/T_{\rm K}} \approx 3\left(1 - \frac{T_*}{T_{\rm K}}\right), \tag{5.8}$$

where again we have assumed that we will have $T_{\rm K} \gg T_*$ for all situations under consideration. It will prove to be convenient to keep track of the ratio between excitation and de-excitation rates via the WF mechanism in an analogous manner by use of the colour temperature T_α defined by

$$\frac{P_{01}}{P_{10}} = 3\left(1 - \frac{T_*}{T_\alpha}\right). \tag{5.9}$$

With these definitions, Eq. (5.7) in the Rayleigh-Jeans limit becomes

$$T_{\rm S}^{-1} = \frac{T_\gamma^{-1} + x_\alpha T_\alpha^{-1} + x_c T_{\rm K}^{-1}}{1 + x_\alpha + x_c}, \tag{5.10}$$

where x_c and x_α are the collisional and WF coupling coefficients, respectively, given by

$$x_c = \frac{C_{10}}{A_{10}}\frac{T_*}{T_\gamma} \qquad\qquad x_\alpha = \frac{P_{10}}{A_{10}}\frac{T_*}{T_\gamma}. \tag{5.11}$$

When examining collisional coupling, the important collisions are that of HI with other hydrogen atoms, and free electrons and protons (see Ref. [19] for an in-depth analysis).

In order to calculate the WF coupling coefficient x_α, we must first find the WF de-excitation rate P_{10}, which will depend on the scatter rate of Lyman-α photons. By careful examination of the Lyman-α transition between different hyperfine levels, one finds that if the background radiation field is constant over the different hyperfine levels, then P_{10} is related to the total scattering rate of Lyman-α photons P_α by $P_{10} = (4/27)P_\alpha$ [70]. P_α in turn is related to the angle-averaged specific intensity J_ν of the background radiation at frequency ν by

$$P_\alpha = 4\pi \int d\nu \sigma_\nu J_\nu, \tag{5.12}$$

where σ_ν is the local absorption cross section at frequency ν. We can write x_α as a function of J_ν evaluated at the Lyman-α line centre (denoted by J_α) as

$$x_\alpha = S_\alpha \frac{J_\alpha}{J_\nu^c}, \tag{5.13}$$

where $J_\nu^c = 5.825 \times 10^{-12}(1+z)\,\mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{Hz}^{-1}\mathrm{sr}^{-1}$ and S_α is a correction term to account for the variation in the background radiation field near the Lyman-α line centre (see Ref. [73] for a calculation of S_α).

The colour temperature T_α determines the relative rate of excitations to de-exciations of the $1S$ hyperfine levels via the absorption and re-emission of Lyman-α photons. As such, the colour temperature will depend on the relative occupation number n_ν of photons with frequency ν in the vicinity of the Lyman-α frequency, as each transition will require absorption or emission of Lyman-α photons with slightly different frequencies. Since this difference in energy is relatively small, we can approximate the ratio P_{01}/P_{10} as

$$\frac{P_{01}}{P_{10}} \approx \frac{g_1}{g_0}\left(1 + \nu_0 \frac{d\ln n_\nu}{d\nu}\right), \tag{5.14}$$

and by comparing to Eq. (5.9) we can write the colour temperature as

$$T_\alpha \approx \frac{h}{k_\mathrm{b}}\left(-\frac{d\ln n_\nu}{d\nu}\right)^{-1}. \tag{5.15}$$

If the scattering rate of Lyman-α photons is very high, as is the case for the high-redshift Universe, through the exchange of energy through atomic recoils the photon spectrum near the Lyman-α frequency will be given approximately by a blackbody spectrum of temperature $T_{\rm K}$, in which case Eq. (5.15) implies $T_\alpha \sim T_{\rm K}$ [74]. A more precise expression for T_α as a function of $T_{\rm K}$ can be found in Ref. [73].

5.3 History of the 21-cm Signal

Currently, the exact history of the pre-reionization 21-cm signal is not precisely known. However, we anticipate the presence of a few likely general events in the 21-cm signal's evolution. In this section we give a brief description of some likely generic features in the evolution of the 21-cm signal.

After recombination, although most electrons are found within atoms, a small fraction ($\sim \mathcal{O}(10^{-3})$) of residual free electrons remain and couple to the CMB through Compton scattering. This coupling remains strong well after recombination until the residual free electrons become extremely diffuse. While strongly coupled to the CMB temperature, the gas has a kinetic temperature $T_{\rm K} \approx T_\gamma$. At these early times, the gas is dense enough so that collisional coupling is strong, so at this time we have $T_{\rm S} \approx T_{\rm K} \approx T_\gamma$ and thus no 21-cm signal can be observed at this point. Numerical calculations using the RECFAST[43] code [75] predict that the decoupling of the kinetic temperature of the gas from the CMB temperature occurs around $z \sim 150$, presumed to be well into the dark ages before significant astrophysical structure formation occurs.

After decoupling from the CMB, the gas cools adiabatically with a cooling rate which is faster than that of the CMB. Therefore, an absorption signal may be present after decoupling. As the gas continues to cool, the collisional coupling becomes less efficient, driving $T_{\rm S}$ back up to the CMB temperature. At this point, the 21-cm signal may disappear again and remains at zero unless the spin temperature deviates from the CMB temperature again due to the presence of astrophysical sources. Photons emitted from astrophysical sources may affect the spin temperature through the WF process as well as may heat the the gas, raising the spin temperature. A signal in emission is generally predicted, followed by reionization which drives the 21-cm signal emitted from the intergalactic medium (IGM) to zero once again. After reionization is complete, 21-cm emissions are confined to overdense regions.

[43] http://www.astro.ubc.ca/people/scott/recfast.html

5.4 Radio Interferometry and Detection of 21-cm Signal

Now that we have a basic understanding of the 21-cm signal and a general idea of its evolution throughout cosmic history, we now focus our attention on the detection of the 21-cm signal by means of radio interferometric telescopes.

Interferometric telescopes consist of a array of antennas, each of which measure the electromagnetic field at a particular location. We label the signal from the i-th feed as F_i. The correlation between feeds i and j, known as the visibility, is given by

$$V_{ij} = \langle F_i^* F_j \rangle = \frac{1}{\sqrt{\Omega_i \Omega_j}} \int d^2\hat{\mathbf{n}} A_i^*(\hat{\mathbf{n}}) A_j(\hat{\mathbf{n}}) e^{2\pi i \hat{\mathbf{n}} \cdot \mathbf{u}_{ij}} T(\hat{\mathbf{n}}), \tag{5.16}$$

where $T(\hat{\mathbf{n}})$ is the brightness temperature of the sky (related to the intensity I by $T = (\lambda^2/2k_\mathrm{b})I$) coming from direction $\hat{\mathbf{n}}$ and $\mathbf{u}_{ij}$ is the spatial separation between the two feeds in units of wavelength, which we refer to as a baseline. $A_i(\hat{\mathbf{n}})$ is the antenna response which has a solid angle of $\Omega_i = \int d^2\hat{\mathbf{n}} |A_i(\hat{\mathbf{n}})|^2$. The term $\hat{\mathbf{n}} \cdot \mathbf{u}_{ij}$ describes the lag in the arrival of the signal at one feed compared to the other and the exponential term describes the interference between the two signals. Although we have not made any implicit mention of polarization at this point, the feed response F_i, the beam $A_i(\hat{\mathbf{n}})$, and the sky intensity $T(\hat{\mathbf{n}})$ all implicitly refer to either a particular polarization or combination of polarizations. For example, the sky signal may be decomposed into separate components for each of the Stoke's parameters. However, such details are not necessary for the purposes of this section.

It is apparent from Eq. (5.16) that the visibilities measure Fourier modes (or more appropriately spherical harmonics) on the sky modulated by the beam response and that a map of the sky can be attained by sampling the visibility in the $\mathbf{u}$ plane and then Fourier transforming. We can think of each visibility as being sensitive to a limited number of modes on the sky. As such, the angular resolution of the telescope is set by the largest baseline. We can estimate the angular resolution of an array at wavelength λ by $\Delta\theta \sim \lambda/L$, where L is the length of the longest baseline. This angle corresponds to the comoving distance $D_A(1+z)\lambda/L$ in the direction perpendicular to the line of sight. For an array with a longest baseline of $L = 100\,\mathrm{m}$, $\Delta\theta$ runs between $\sim 0.2^\circ - 0.5^\circ$ in the frequency range $400-800$ MHz, which corresponds to

comoving distances roughly between $\sim 10-50\,\mathrm{Mpc}$.[44] Such a resolution should be adequate for measuring the baryon acoustic oscillations (BAO), which has a comoving length scale of $\sim 150\,\mathrm{Mpc}$.

The power p measured by the autocorrelation of a particular feed can be represented by $p = g^2 k_\mathrm{b} \Delta\nu T_\mathrm{sys}$, where g is the gain of the feed and $\Delta\nu$ is the size of the frequency bin. The system temperature T_sys is a sum over all sources of power, including both sources on the sky as well as instrumental noise.

By forming the four-point function and with use of Wick's theorem, the variance of the output of the correlator forming the visibilities can be found. If the noise between feeds i and j are uncorrelated, this variance is given by $2T_\mathrm{sys,i}T_\mathrm{sys,j}$. By averaging over time, the variance can be reduced by a factor of $N = 2t_\mathrm{int}\Delta\nu$, representing the number of independent measurements possible within the integration time t_int. The variance on a visibility then becomes

$$\sigma_{ij}^2 = \frac{T_\mathrm{sys,i}T_\mathrm{sys,j}}{t_\mathrm{int}\Delta\nu}. \tag{5.17}$$

This highlights the key factors in reducing uncertainty in measurements of radio signals: Having long integration times and a system that adds as little as possible additional noise to the system temperature can both improve the precision of our instrument.

[44] Assuming a cosmology with $\Omega_\mathrm{m} = 0.27$, $\Omega_\Lambda = 0.73$, and $h = 0.7$.

Chapter 6

The Imprint of Warm Dark Matter on the Cosmological 21-cm Signal

6.1 Introduction

Hierarchical structure formation[45] within the ΛCDM model has been exceptionally accurate in describing the large-scale Universe within the range $\sim 10\,\mathrm{Mpc}-1\,\mathrm{Gpc}$, as demonstrated from studies of the cosmic microwave background (CMB) and the clustering of galaxies. However, for over a decade concerns have been raised over whether the standard assumption of cold dark matter (CDM) provides an adequate fit to data on smaller, sub-Mpc scales. These include predictions from N-body simulations that yield an overabundance of galactic satellites around our galaxy and in the field [76, 77, 78], as well as in voids [79], and produce overly-dense galactic centres with 'cuspy' density profiles [80, 81, 82] and are inconsistent with observations of the kinetic properties of bright Milky Way satellites [83, 84].

One possible explanation lies with baryonic feedback processes [85, 86, 87, 88, 89], although accurately modelling these mechanisms is often challenging and difficulties may persist in matching to observations.

Another possible explanation is to change the properties of dark matter so it is warm (WDM).[46] This may alleviate these small-scale problems due to the higher velocities of the dark matter. In this case, structures are smoothed on scales below the dark matter's free-streaming length. Non-relativistic residual velocities can delay halo collapse and star formation. These effects may reduce the number of sub-halos and low-mass galaxies that are formed as well as flatten out galactic centres.

[45] Hierarchical structure formation describes the general formation procedure of smaller objects forming first and then merging to form larger structures.

[46] Other possible alterations to the standard CDM model that may resolve these small-scale problems include self-interacting dark matter [90, 91, 92] and atomic dark matter or other models with acoustic damping of dark matter fluctuations [93, 94].

The two most popular WDM candidates in the literature motivated by particle physics have been the sterile neutrino [95, 96, 97] and the gravitino [98, 99]. While WDM may be produced in a number of different ways, it is most often described as a thermal relic that decouples while relativistic, but is non-relativistic by matter-radiation equality as to preserve structure beyond the Mpc scale. In this case, the WDM would have a particle mass $m_{\rm X}$ of the order of a keV. Although for our purposes the free-streaming scale of the dark matter is a more fundamental quantity, we use the standard convention of discussing the WDM mass of a thermal relic instead. We caution that for other WDM production mechanisms the correspondence between free-streaming length and mass will be different. We also remark that the results presented in this chapter can be applicable to models other than WDM that have similar cut-off scales in their power spectrum (see, e.g. Ref. [93]).

As WDM suppresses growth of small structures, which form first in the hierarchical structure formation of CDM, early star formation is delayed in WDM models. Detection of signals emitted from high-redshift objects either directly, such as from gamma-ray bursts (GRBs) [100] or strongly lensed galaxies [101], or indirectly through the redshift of reionization [102], can place constraints on $m_{\rm X}$. Recently, Ref. [103] using GRB catalogues placed a constraint of $m_{\rm X} > 1.6 - 1.8\,\mathrm{keV}$ at 95% CL. Requiring WDM models to be able to reproduce both the stellar mass function and Tully-Fisher relation places a lower bound of $m_{\rm X} \geq 0.75\,\mathrm{keV}$ [104]. The Lyman-α forest can probe scales down to $\sim 1\,\mathrm{Mpc}$ and can provide strict limits on $m_{\rm X}$ [105, 106, 107, 108], with the most recent and stringent constraint of $m_{\rm X} > 3.3\,\mathrm{keV}$ at 2σ [109]. Although it has been claimed that the less dense galactic cores formed in WDM models may provide a better fit to the kinematic data of bright Milky Way satellites [110], there is an ongoing debate as to whether WDM with a mass above current lower bounds can create a large enough galactic core as needed to solve the 'cusp-core' problem [111, 112]; though see Ref. [113].

Highly-redshifted 21-cm radiation emitted from the hyperfine spin-flip of neutral hydrogen is a promising new tool to probe the high-redshift Universe [19, 114, 115, 116, 117]. If WDM is present in sufficient quantities to significantly delay structure formation, it could potentially leave a trace within the 21-cm radiation signal. Light emitted by the first astrophysical sources can couple the spin temperature of neutral hydrogen to the kinetic temperature of the IGM through the Wouthuysen-Field (WF) mechanism [70, 71], as well as heat and ionize the IGM. Thus, a delay in the appearance of these early sources can alter the 21-cm signal and delay milestones in the

signal. In this chapter, we will examine the effects of WDM on the pre-reionization 21-cm signal. This era may be especially useful for examining WDM since WDM inhibits the formation of low-mass halos that form first in CDM models and thus differences between the halo populations in CDM and WDM increase with redshift. As astrophysics is very poorly known at high-redshifts ($z \geq 6$), *we will focus on characterizing degeneracies between the unknown astrophysics and the presence of WDM.*

The outline of this chapter is as follows: In Section 6.3, we review the effects of the free-streaming of the WDM on the linear power spectrum and its residual velocities on halo collapse. The basic properties of the 21-cm signal are outlined in Section 6.4 and its simulation is described in Section 6.5, with the simulation results discussed in Section 6.6. Throughout this chapter, we assume cosmological parameter values of $\Omega_\Lambda = 0.73$, $\Omega_\mathrm{m} = 0.27$, $\Omega_\mathrm{b} = 0.046$, $h = 0.7$, $\sigma_8 = 0.82$, $n_\mathrm{s} = 0.96$. We quote all quantities in comoving units, unless stated otherwise.

6.2 Thermal Relic

As previously mentioned, WDM is most often described as a thermal relic [4, 118, 119]. Here we derive some basic results that relate fundamental properties of the WDM to its free-streaming length.

We begin by examining the distribution function for a WDM thermal relic. Since the physical momentum $\mathbf{p}$ scales as $|\mathbf{p}| \propto a^{-1}$ for both relativistic and nonrelativistic noninteracting particles, after decoupling the comoving momentum $\mathbf{q} = a\mathbf{p}$ of the WDM remains constant with time (assuming any interactions are negligible at this time). Consequently, after decoupling the distribution function f of the WDM written as a function of the comoving momentum is constant in time. Therefore, the distribution function as a function of physical momentum evolves as $f(\mathbf{p}, t) = f_i((a(t)/a_i)\mathbf{p})$, where the subscript i denotes evaluation at some initial time. Since the WDM decouples while relativistic, the initial distribution function f_i for the WDM is a function of p/T_i and at any later time (both when the WDM is relativistic or nonrelativistic) the distribution function will be given by $f(\mathbf{p}, t) = f_i((a(t)/a_i)p/T_i)$. We can define an effective temperature $T_\mathrm{X}(t) = (a_i/a(t))T_i$ for the WDM that specifies its distribution function while both relativistic and nonrelativistic as $f(\mathbf{p}, t) = f_i(p/T_\mathrm{X}(t))$, which will coincide with its physical temperature while relativistic.

Assuming all species to be in thermal equilibrium at early enough times, we can use the conservation of entropy given in Eq. (2.17) with $T_\mathrm{X} \propto a^{-1}$

to relate T_{X} to the photon temperature T

$$\frac{T_{\mathrm{X}}}{T} = \left(\frac{g_{s0}}{g_{s*}}\right)^{1/3}, \tag{6.1}$$

where g_{s0} and g_{s*} are the number of relativistic degrees of freedom contributing to the entropy at present and decoupling, respectively. As the WDM decouples while relativistic, this calculation is identical to the decoupling of neutrinos, where we recover the well-known result for the fraction between the neutrino and photon temperatures by setting g_{s*} to its value at neutrino decoupling ($g_{s*} = 10.75$) [1]. Since $g_{s0} \leq g_{s*}$, T_{X} will be lower than the photon temperature soon after the WDM decouples, as the WDM misses out on the entropy release from the annihilation of other species after its decoupling.

The present day ratio between the WDM and photon number densities can be found from Eq. (2.15c) as

$$\frac{n_{\mathrm{X}0}}{n_{\gamma 0}} = \left(\frac{T_{\mathrm{X}}}{T}\right)^3 \frac{g_{n\mathrm{X}}}{g_{n\gamma}} = \frac{g_{s0}}{g_{s*}} \frac{g_{n\mathrm{X}}}{2}, \tag{6.2}$$

where $g_{n\gamma} = 2$ and $g_{n\mathrm{X}}$ are the number of relativistic degrees of freedom contributing to the number density of the photons and WDM, respectively. Most often the WDM is assumed to be a spin-$\frac{1}{2}$ particle, in which case $g_{n\mathrm{X}} = 3/2$. Since at the present day the WDM is nonrelativistic, its energy density can be found by $\rho_{\mathrm{X}} = m_{\mathrm{X}} n_{\mathrm{X}}$, which results in the present-day fractional energy density of the WDM of

$$\Omega_{\mathrm{X}} h^2 \approx \frac{115}{g_{s*}} \frac{g_{n\mathrm{X}}}{1.5} \frac{m_{\mathrm{X}}}{\mathrm{keV}}, \tag{6.3}$$

where we have used $g_{s0} = 43/11$ [3]. Therefore, the required value of g_{s*} to obtain the observed present-day dark matter density is

$$g_{s*} \approx 767 \left(\frac{\Omega_{\mathrm{X}} h^2}{0.15}\right)^{-1} \frac{g_{n\mathrm{X}}}{1.5} \frac{m_{\mathrm{X}}}{\mathrm{keV}}. \tag{6.4}$$

On first observation, we see that a keV scale WDM thermal relic would have had to decouple at a time when g_{s*} is much larger than that in the standard model while all species are relativistic, requiring physics beyond the standard model to add an array of new particles. However, as was noted in Ref. [120], other scenarios, such as a production of entropy after WDM decoupling, have the net effect of increasing g_{s*}, relaxing the requirement of having to add variety of new particles.

We now return to the distribution function of the WDM. At late times when the WDM is nonrelativistic, we can use $p_{\rm X} = m_{\rm X} v$ to write its distribution function as a function v/v_0 where v is the velocity of the WDM and $v_0 = T_{\rm X}/m_{\rm X}$, which using Eqs. (6.1) and (6.4) is given by

$$v_0(z) = 0.0121(1+z)\left(\frac{\Omega_{\rm X} h^2}{0.15}\right)^{1/3}\left(\frac{g_{n\rm X}}{1.5}\right)^{-1/3}\left(\frac{m_{\rm X}}{\rm keV}\right)^{-4/3}\,{\rm km\,s^{-1}}. \quad (6.5)$$

For the remainder of this chapter, we assume our WDM thermal relic to be a fermion, although adapting the results for a boson is straightforward. With this, the distribution function of the WDM is $f(v) = [\exp(v/v_0)+1]^{-1}$ with a root-mean-squared velocity $v_{\rm rms} = 3.597 v_0$.

6.3 Effect of WDM on structure formation

6.3.1 Free-streaming

The free-streaming of WDM particles smears out perturbations on small scales, as WDM particles stream out of over-dense regions and into under-dense regions. Perturbations are suppressed on scales below that corresponding to the WDM particle horizon.

The effect of free-streaming on the spectrum of linear perturbations can be included by use of a transfer function $T_{\rm X}(k)$ that dampens small-scale fluctuations as compared to those in CDM. This transfer function can be found by fitting the results of a Boltzmann code that utilizes the WDM velocity found in Section 6.2, which we take as

$$T_{\rm X}(k) = (1 + (\epsilon k R_{\rm c}^0)^{2\nu})^{-\eta/\nu}, \quad (6.6)$$

where $\epsilon = 0.361$, $\eta = 5$, and $\nu = 1.2$ [120]. $R_{\rm c}^0$ is the comoving cutoff scale, at which the power in $k = 1/R_{\rm c}^0$ is reduced by half compared to that in CDM, and is given by

$$R_{\rm c}^0 = 0.201\left(\frac{\Omega_{\rm X} h^2}{0.15}\right)^{0.15}\left(\frac{g_{n\rm X}}{1.5}\right)^{-0.29}\left(\frac{m_{\rm X}}{\rm keV}\right)^{-1.15}{\rm Mpc}, \quad (6.7)$$

where $g_{n\rm X}$ is the number of effective degrees of freedom contributing to number density, with bosons contributing unity to $g_{n\rm X}$ and fermions contributing 3/4. We will use the standard assumption that the WDM is a spin-$\frac{1}{2}$ fermion, so that $g_{n\rm X} = 3/2$. $\Omega_{\rm X}$ is the energy density parameter contributed by the WDM, which we set to $\Omega_{\rm X} = \Omega_{\rm m} - \Omega_{\rm b}$ as we will only

be considering models where WDM constitutes the whole of the dark matter. The transfer function in Eq. (6.6) serves to suppress small-scale linear perturbations in the power spectrum, which we generate using the transfer function of Ref. [28].

6.3.2 Residual velocities

In addition, the residual velocity dispersion of the WDM delays the growth of non-linear perturbations and consequently collapse into virialized halos. This can be thought of as an 'effective pressure'. Ref. [120] modelled the collapse in WDM by studying collapse in an analogous system comprised of a monoatomic adiabatic gas at temperature $T \propto v_{\rm rms}^2$. The gas temperature evolves as $T \propto 1/a^2$ so its root-mean-square velocity evolves as $v_{\rm rms} \propto 1/a$, as the case with WDM as seen in Eq. (6.5). The initial temperature of the gas is set such that it shares the same $v_{\rm rms}$ with the WDM as found in Section 6.2.

Using the gas analogue in a spherically symmetric hydrodynamics simulation, Ref. [120] computed the linear collapse threshold $\delta_c(M, z)$, finding that the collapse threshold rises sharply near the Jeans mass $M_{\rm J}$, the mass in which a gas cloud's internal pressure can no longer support it against gravitational collapse, for the analogue gas. From Eq. (6.5), the Jeans mass of the gas analogue is proportional to $M_{\rm J} \propto T^{3/2}/\rho^{1/2} \propto (\Omega_{\rm X} h^2)^{1/2} g_{n\rm X}^{-1} m_{\rm X}^{-4}$. The results of Ref. [103] showed that using the extended Press-Schechter (EPS) formalism to compute the collapse fraction with a sharp minimum mass cutoff at $M_{\rm J}$ and the collapse threshold for spherical collapse in CDM ($\delta_c \approx 1.69$) is in good agreement with the full random-walk procedure with the WDM modified collapse threshold as used in Ref. [120]. To achieve this close agreement, a factor of 60 was added to the expression for $M_{\rm J}$ originally found in Ref. [120], so that $M_{\rm J}$ is given by

$$M_{\rm J} \approx 1.5 \times 10^{10} \left(\frac{\Omega_{\rm X} h^2}{0.15}\right)^{1/2} \left(\frac{g_{n\rm X}}{1.5}\right)^{-1} \left(\frac{m_{\rm X}}{\rm keV}\right)^{-4} {\rm M}_\odot. \tag{6.8}$$

As using the sharp cutoff at $M_{\rm J}$ is much less computationally intensive and easily integrable within the EPS formalism, we employ this method instead of the full random-walk procedure.

6.3.3 Halo Abundances

The production rate of photons that are capable of heating or ionizing the IGM, or coupling the spin temperature to the colour temperature via the

WF mechanism, is modelled as being proportional to the collapse fraction $f_{\rm coll}(z, M_{\rm min})$ of halos with sufficient mass ($\geq M_{\rm min}$) to host star-forming galaxies. To compute the mean collapse fraction, we use the Sheth-Tormen mass function found in Eq. (2.48), giving the comoving number density of halos with mass between M and $M + dM$ as

$$\frac{dn_{\rm ST}}{dM} = -A\sqrt{\frac{2}{\pi}}\frac{\bar{\rho}_{\rm m}}{M}\frac{d\ln\sigma}{dM}\hat{\nu}(1+\hat{\nu}^{-2p})e^{-\hat{\nu}^2/2}, \tag{6.9}$$

where $\hat{\nu} = \sqrt{a}\delta_c(M,z)/\sigma(M)$, $\bar{\rho}_{\rm m}$ is the mean matter energy density, $\sigma(M)$ is the rms of density fluctuations smoothed on a scale that encompasses a mass M. A, a, and p are fit parameters taken as $A = 0.353$, $a = 0.73$, and $p = 0.175$ [121]. The mean collapse fraction is computed as

$$f_{\rm coll}(> M_{\rm min}, z) = \frac{1}{\rho_{\rm m}}\int_{M_{\rm min}}^{\infty} M\frac{dn_{\rm ST}}{dM}dM, \tag{6.10}$$

where $M_{\rm min} = \max(M_{\rm J}, M_{\rm sf})$ and $M_{\rm sf}$ is the minimum halo mass where star-formation can occur. $M_{\rm J}$ is assigned a value of zero in the case of CDM. It will be convenient to express $M_{\rm sf}$ in terms of the corresponding virialized halo temperature $T_{\rm vir}$ as (see Eq. (2.52))

$$\begin{aligned} M_{\rm sf} = 9.37\times10^7 &\left(\frac{\mu}{0.6}\right)^{-3/2}\left(\frac{h}{0.7}\right)^{-1}\left(\frac{\Omega_{\rm m}}{0.3}\right)^{-1/2} \\ &\times\left(\frac{1}{\Omega_{\rm m}^z}\frac{\Delta_{\rm c}}{18\pi^2}\right)^{-1/2}\left(\frac{1+z}{10}\right)^{-3/2}\left(\frac{T_{\rm vir}}{10^4\,{\rm K}}\right)^{3/2}{\rm M}_\odot. \end{aligned} \tag{6.11}$$

The mean collapse fraction in CDM and WDM models can be seen in Fig. 6.1. At high redshifts, small halos begin to collapse in CDM, while no or few such halos collapse in WDM, resulting in a large relative difference between the collapse fractions in these models. However, this difference becomes smaller with lower redshifts as objects on scales larger than that inhibited by WDM start to collapse in both models. At late times, in the CDM scenario the mass within halos of sizes suppressed by WDM only represents a small fraction of the total mass within all collapsed structures, so the relative difference between the mean collapse fraction in CDM and WDM models is small at those times. Therefore, while structure formation is delayed in WDM models, the mean collapse fraction raises more rapidly as compared to CDM.

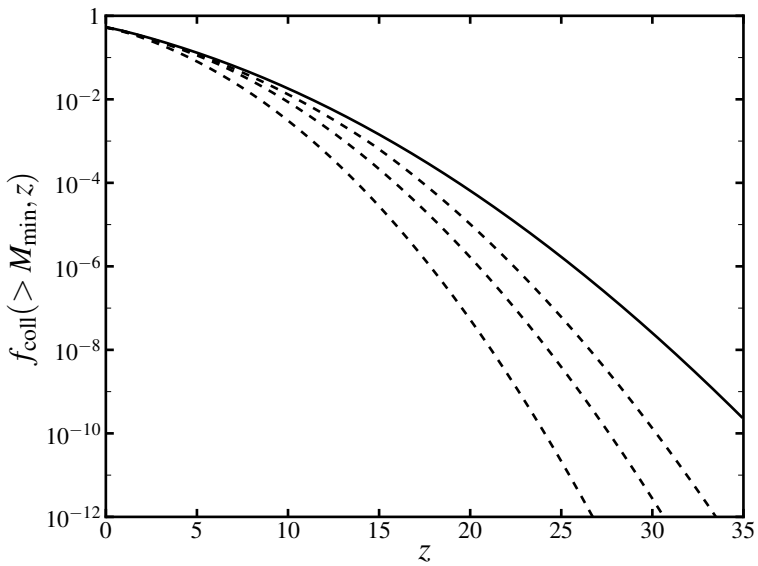


Figure 6.1: Mean collapse fraction for CDM (solid) and WDM (dashed) models. The WDM curves in ascending order are for $m_{\rm X} = 2, 3, 4\,{\rm keV}$. The collapse fraction is calculated using Eq. (6.10) with $M_{\rm sf}$ set by $T_{\rm vir} = 10^4\,{\rm K}$.

6.4 Cosmic 21-cm signal

The brightness temperature of the 21-cm signal measured against the CMB at redshift z is given by

$$\begin{aligned}\delta T_{\rm b}(z) &= \frac{T_{\rm S} - T_\gamma}{1+z}(1 - e^{-\tau_{\nu_0}}) \\ &\approx 27 x_{\rm HI}(1+\delta)\left(1 - \frac{T_\gamma}{T_{\rm S}}\right)\left(\frac{1+z}{10}\frac{0.15}{\Omega_{\rm m}h^2}\right)^{1/2} \\ &\times \left(\frac{\Omega_{\rm b}h^2}{0.023}\right)\left(\frac{H}{H + dv_\parallel/dr_\parallel}\right)\,{\rm mK},\end{aligned} \tag{6.12}$$

where τ_{ν_0} is the optical depth at the 21-cm frequency ν_0, $T_{\rm S}$ and T_γ are the spin and CMB temperatures, respectively, $x_{\rm HI}$ is the neutral fraction of hydrogen, δ is the overdensity, H is the Hubble parameter and $dv_\parallel/dr_\parallel$ is the comoving velocity gradient along the line of sight. The spin temperature can be represented by

$$T_{\rm S}^{-1} = \frac{T_\gamma^{-1} + x_\alpha T_\alpha^{-1} + x_c T_K^{-1}}{1 + x_\alpha + x_c}, \tag{6.13}$$

where $T_{\rm K}$ and T_α are the kinetic and colour temperatures, respectively, and x_c and x_α are the collisional and WF coupling coefficients, respectively.

The earliest possible measurable cosmic 21-cm signal would be emitted during the 'dark ages' before significant star formation occurs. At these early times, the gas is dense enough so that collisional coupling is strong and $T_{\rm S} \approx T_{\rm K}$. Before $z \sim 150$, residual free electrons strongly couple the gas kinetic temperature to the CMB through Compton scattering, so $T_{\rm S} \approx T_{\rm K} \approx T_\gamma$ and no 21-cm signal can be observed at this time. After this point, any remaining free electrons are so defuse that the gas is decoupled from the CMB and cools adiabatically as $T_{\rm K} \propto (1+z)^2$. Since the CMB temperature decreases at the slower pace of $T_\gamma \propto (1+z)$, a 21-cm signal in absorption may be observed (at least in principle) at this time [122, 123, 124, 125]. As the gas continues to cool, the collisional coupling becomes less efficient, driving $T_{\rm S}$ back up to the CMB temperature. As this scenario is relatively unaffected by structure formation, we do not expect the presence of WDM to significantly affect this era of the 21-cm signal and will restrict our attention to later times with redshifts below $z \sim 35$.[47]

It will be important to keep in mind that the kinetic temperature of the gas will be lower than the CMB temperature when WF coupling first becomes effective. As the Lyman-α background grows, the increasing strength of the WF coupling will drive $T_{\rm S}$ from a value near the CMB temperature to the lower kinetic temperature of the gas, thus producing another absorption signal. As WDM delays structure formation, the production of significant UV and X-ray backgrounds will be delayed, which in turn modifies the WF coupling, X-ray heating, and reionization. We therefore focus our attention to the astrophysical epochs in the 21-cm signal.

6.5 Simulation of 21-cm signal

The 21-cm signal is simulated using the publicly available 21CMFAST code. This is a semi-numerical simulation that generates density, velocity, ionization and spin temperature fields in a 3D box with length size ∼Gpc. In this section we briefly summarize the code. See Refs. [128, 129] and references within for further details.

An initial linear density field is generated as a Gaussian random field described by a power spectrum. The initial linear density field is then evolved using the Zeldovich approximation.

[47] On the other hand, these early epochs may be affected by dark matter decay or annihilation [126, 127].

Since we will be examining high-redshift eras, it will be necessary to compute the spin temperature and consequently the colour and kinetic temperatures and their associated coupling coefficients. The WF coupling coefficient x_α is given by

$$x_\alpha = S_\alpha \frac{J_\alpha}{J_\nu^c}, \tag{6.14}$$

where J_α is the angle-averaged Lyman-α background flux, S_α is a quantum correction term and $J_\nu^c = 5.825 \times 10^{-12}(1+z)\,\mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{Hz}^{-1}\mathrm{sr}^{-1}$. S_α and the colour temperature T_α are computed according to Ref. [73]. The kinetic temperature T_K is calculated by solving the set of (local) coupled differential equations for T_K and the ionized fraction x_e in the neutral IGM, given by

$$\frac{dx_\mathrm{e}(\mathbf{x}, z)}{dz} = \frac{dt}{dz}\left(\Lambda_\mathrm{ion} - \alpha_\mathrm{A} C x_\mathrm{e}^2 n_\mathrm{b} f_\mathrm{H}\right), \tag{6.15a}$$

$$\frac{dT_\mathrm{K}(\mathbf{x}, z)}{dz} = \frac{2}{3k_\mathrm{b}(1+x_\mathrm{e})}\frac{dt}{dz}\sum_p \epsilon_p + \frac{2T_\mathrm{K}}{3n_\mathrm{b}}\frac{dn_\mathrm{b}}{dz} - \frac{T_\mathrm{K}}{1+x_\mathrm{e}}\frac{dx_\mathrm{e}}{dz}, \tag{6.15b}$$

where Λ_ion is the ionization rate per baryon, α_A is the case-A recombination coefficient, C is the clumping factor, n_b is the total baryon number density, f_H is the hydrogen number fraction, and ϵ_p is the heating rate for process p. The heating processes considered are X-ray heating ϵ_X and Compton heating ϵ_comp.

It is necessary to estimate the emission rate of photons at a particular frequency to compute ϵ_X, Λ_ion, and J_α. The primary sources of X-ray photons are expected to be high-mass X-ray binaries and the inverse-Compton scattering off of relativistic electrons accelerated in supernovae [19]. The UV background is expected to be sourced from the collisional excitation of neutral hydrogen by electrons ionized by X-rays as well as by direct stellar emission [129]. We will make the conventional assumption that the emission rate of photons over the frequencies of interest can be approximated as being proportional to the star-formation rate, which is a reasonable assumption given the local correlation between star formation rate and X-ray luminosity [130]. The star formation rate in turn is approximated and readily computed by using the growth of the collapse fraction. The comoving emissivity e at frequency ν is then

$$e(\nu) = f_* \rho_\mathrm{b} N(\nu) \frac{df_\mathrm{coll}}{dt}, \tag{6.16}$$

where f_* is the fraction of baryons that are incorporated into stars, $\rho_\mathrm{b} = \bar{\rho}_\mathrm{b}(1+\delta_\mathrm{nl})$ is the total baryon density including the non-linear overdensity δ_nl, and $N(\nu)$ is the number of photons with frequency ν per solar mass in

stars. The local collapse fraction is computed using the hybrid prescription of Ref. [132], where the biased EPS method is used to compute relative local halo abundances whose mean is then normalized to fit the mean collapse fraction given by the Sheth-Tormen mass function in Eq. (6.10).

Ionization fields are generated by assuming that a region is ionized if it contains more ionizing photons than neutral hydrogen atoms (multiplied by $1 + \bar{n}_{\rm rec}$, where $\bar{n}_{\rm rec}$ is the mean number of recombinations per baryon). The excursion-set formalism is used with the condition that $\zeta f_{\rm coll}(\mathbf{x}, z, R) \geq 1 - x_{\rm e}(\mathbf{x}, z, R)$ for a cell centred at location $\mathbf{x}$ to be fully ionized, where $f_{\rm coll}(\mathbf{x}, z, R)$ is the collapse fraction smoothed on scale R, ζ is the ionization efficiency, and $1 - x_{\rm e}(\mathbf{x}, z, R)$ is the remaining fraction of neutral hydrogen within R. This criterion is evaluated at deceasing scales R and if the cell is not marked as fully ionized as the scale of the pixel length is reached, the cell's ionization fraction is marked as $\zeta f_{\rm coll}(\mathbf{x}, z, R_{\rm cell}) + x_{\rm e}(\mathbf{x}, z)$. Lastly, we note that the ionization efficiency can be decomposed as $\zeta = A_{\rm He} f_* f_{\rm esc} N_{\rm ion}/(1 + \bar{n}_{\rm rec})$, where $f_{\rm esc}$ is the fraction of ionizing photons that escape their host galaxy, $N_{\rm ion}$ is the number of ionizing photons per baryon inside stars and $A_{\rm He}$ is a correction factor due to the presence of Helium.

6.6 Simulation Results

As much is unknown about astrophysical properties during high-redshift eras, we will examine possible degeneracies in the 21-cm signal between WDM and astrophysical quantities. As a first step, we will compare the delayed WDM 21-cm signal with that in CDM with a reduced photon-production efficiency. Specifically, we decrease the efficiency uniformly over frequency by decreasing f_*, but note that f_* is degenerate with other parameters used to calculate photon production efficiencies.

The box used in our simulation runs was 750 Mpc on a side and was comprised of 300^3 cells. The 21-cm signal was simulated in the redshift range $z = 5.6$ to 35. We set the minimum halo virial temperature that supports star formation to be $T_{\rm vir} = 10^4\,{\rm K}$ as to approximate the minimum temperature need to efficiently cool the halo gas through atomic cooling, neglecting possible feedback processes.[48]

Our fiducial model uses a f_* value of $f_{*\rm fid} = 10\%$. We set the number

[48] Although the very first stars were likely formed within smaller halos with $T_{\rm vir}$ on the order of $10^3\,{\rm K}$ that were molecularly cooled, star formation in such halos can easily be disrupted by feedback processes [133, 134] and we therefore neglect radiation from sources located in such halos.

of X-rays per solar mass in stars to $N_{\rm X}(\nu) = 2.2 \times 10^{56}\,{\rm M}_\odot^{-1} \times (\nu/7 \times 10^{16}\,{\rm Hz})^{-1.5}$, to roughly match X-ray luminosities at low redshifts [129] and the number of UV photons per solar mass in stars to $N_{\rm UV} = 2.5 \times 10^{60}\,{\rm M}_\odot$ to approximately coincide with that of Pop II stars [131].[49] The fiducial ionization efficiency is taken as $\zeta = 31.5$.

Examples of the mean spin and kinetic temperatures for CDM and WDM models are plotted in Fig. 6.2. As expected, for WDM $T_{\rm S}$ stays near T_γ for a longer time and the lowest point in the absorption trough, where the X-ray heating rate first surpasses the adiabatic cooling rate, occurs later. As mentioned in Section 6.3.3, although the mean collapse fraction is lower in WDM models, it grows more rapidly, which is reflected in the heating of the gas. In addition, Fig. 6.2 shows curves for CDM with the lower f_* value of $f_*/f_{*\rm fid} = 0.1$, which in our model happens to delay star formation such that the minimum value of $\bar{T}_{\rm S}$ occurs roughly at the same time as in the WDM example used. In this case, the X-ray heating rate increases at a much slower rate after the minimum in $\bar{T}_{\rm S}$ as compared to the two other cases shown, since lowering f_* reduces the photon production efficiency in stars of all masses. In both non-fiducial cases shown, $\bar{T}_{\rm S}$ and thus $\delta\bar{T}_{\rm b}$ reach a lower value in their absorption troughs since the gas undergoes further cooling in the extra time needed for the X-ray heating to become efficient.

The evolution of the mean brightness temperatures for WDM models with $m_{\rm X} = 2, 3, 4\,{\rm keV}$ are shown in Fig. 6.3. [50] It is readily seen that having WDM with a particle mass of a few keV can substantially change the mean 21-cm brightness temperature evolution. While lowering f_* within CDM models can delay the strong absorption signal, the resulting absorption trough is much wider than in WDM. For the same delay in the minimum of $\delta\bar{T}_{\rm b}$, the delay in reionization is greater for CDM than for WDM. Although reionization may be greatly delayed, well past $z = 6$, in models with low values of f_*, our primary focus is on the pre-reionization 21-cm signal. We caution against automatically discarding these models, as the star-formation efficiency may diverge from earlier values by reionization.

Examining the gradient of the global signal in Fig. 6.3b, we see the suppressing f_* in CDM models only shifts the mean signal to lower redshifts.

[49] Note that since f_* is degenerate with a frequency-independent value of N, only the ratio of $N_{\rm X}/N_{\rm UV}$ is relevant.

[50] We caution the reader that WDM models with $m_{\rm X} = 2, 3\,{\rm keV}$ are disfavoured by recent Lyman-α observations [109]. However, Lyman-α forest constraints are still susceptible to astrophysical (thermal and ionization history) and observational (sky and continuum subtraction) degeneracies. Therefore, it is still useful to confirm these constraints using the redshifted 21-cm signal.

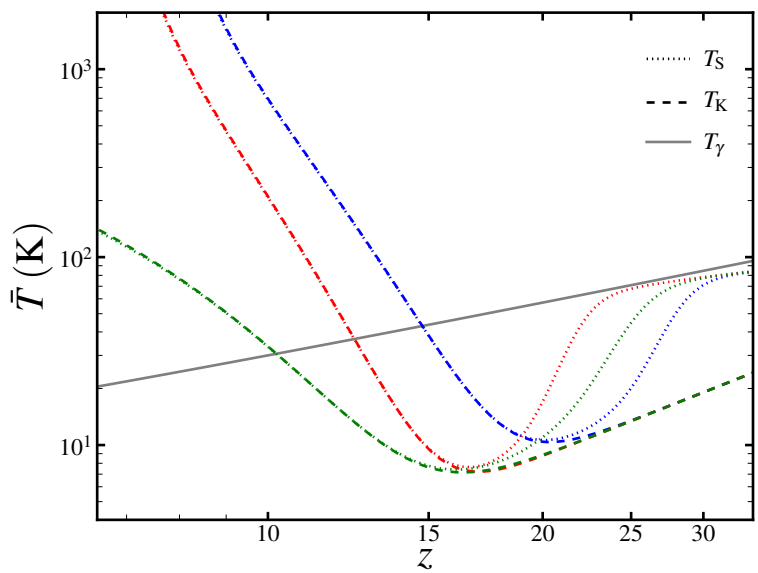


Figure 6.2: Mean spin temperatures $\bar{T}_S$ for CDM and WDM models. The dotted curves show $\bar{T}_S$ for our fiducial CDM model (blue), WDM with $m_X = 3\,\mathrm{keV}$ (red), and CDM with $f_*/f_{*\mathrm{fid}} = 0.1$ (green). In addition, the mean kinetic temperature $\bar{T}_K$ of each model is plotted with a dashed curve in the same colour used for $\bar{T}_S$. The grey solid line is the CMB temperature.

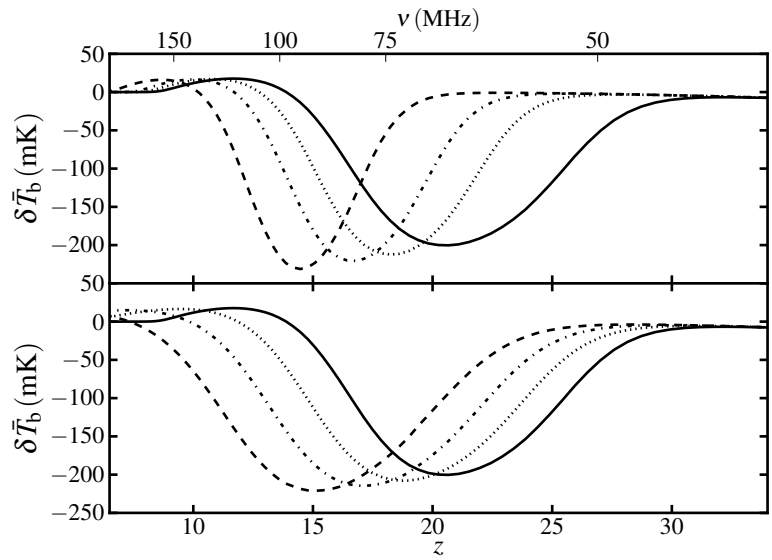
On the other hand, decreasing $m_{\rm X}$ in WDM models increases the gradients of the mean signal. In CDM models, $\partial\delta\bar{T}_{\rm b}/\partial z$ attains values near 33 mK (-45 mK) near its maximum (minimum) regardless of its f_* value. This can increase significantly in WDM models, for example to ~ 64 mK (~ -77 mK) at its maximum (minimum) for WDM with $m_{\rm X} = 2$ keV.

The effect of WDM on the global 21-cm signal can be tracked through different 'critical points' in the signal's evolution. We choose these points to be the redshift $z_{\rm min}$ at which $\delta\bar{T}_{\rm b}$ reaches its minimum value, the redshift $z_{\rm h}$ when the kinetic temperature of the gas is heated above the CMB temperature, and the redshift of reionization $z_{\rm r}$ taken to be the redshift where the mean ionized fraction is $\bar{x}_i(z_{\rm r}) = 0.5$. These points are plotted for both CDM and WDM in Fig. 6.4. The solid curves track the effect of lowering f_* on the redshifts of the critical points in CDM models (the values of f_* can be read from the upper horizontal axis). The dashed curves show the effect of WDM on these redshifts, where the value of $m_{\rm X}$ for each model can be read from the lower horizontal axis.

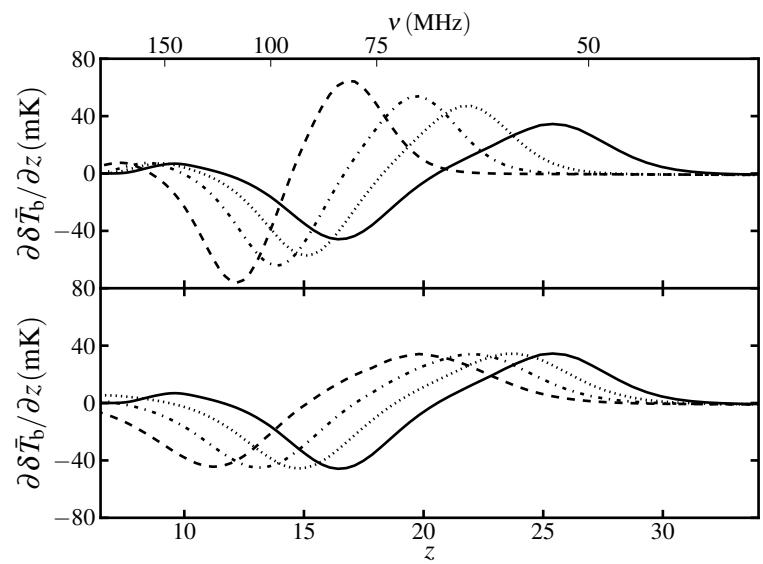
We begin to explore possible degeneracies between CDM and WDM cosmologies by finding the value of f_* required in CDM that would have a particular critical point occur at the same redshift as it would in WDM with a particular value of $m_{\rm X}$. In other words, for a particular event that occurs at redshift z_e, we would like to find the curve that satisfies $z_e(f_*|{\rm CDM}) = z_e(m_{\rm X}|{\rm WDM})$. These curves for $z_{\rm min}$, $z_{\rm h}$, and $z_{\rm r}$ can be seen in Fig. 6.5. We can see that if one uses the milestone $z_{\rm r}$ to distinguish between CDM and WDM with $m_{\rm X} = 2, 3, 4$ keV then f_* has to be known within a factor of 3.0, 1.8, and 1.4, respectively. Using $z_{\rm min}$ instead, f_* only has to be known within a factor of 50, 13, and 4.8 for $m_{\rm X} = 2, 3, 4$ keV, respectively, since the impact of WDM is larger at higher redshifts. Near $m_{\rm X} = 15$ keV, using $z_{\rm min}$ to distinguish WDM from CDM requires f_* to be known within a factor of 1.1 and drops to 1.01 by $m_{\rm X} \sim 20$ keV (although the astrophysical motivations for WDM as mentioned in the introduction loses much of its appeal past a few keV).

As the value of $m_{\rm X}$ is lowered, the curves in Fig. 6.5 diverge from one another, as the more rapid growth of structure in WDM changes the relative timing of the milestones. Therefore, if f_* is approximately constant throughout the epochs under consideration, adjusting the value of f_* in CDM so that a particular critical point occurs at the same redshift as it does in WDM will misalign other critical points and thus cannot reproduce the whole history of $\delta\bar{T}_{\rm b}$ in WDM models.

However, we can mimic the WDM mean brightness temperature evolution with CDM if we allow f_* to vary in time. To illustrate this, Fig. 6.6



(a)



(b)

Figure 6.3: Mean 21-cm brightness temperature $\delta\bar{T}_b$ (a) and its derivative with respect to redshift (b). In all plots, the solid curve is the fiducial CDM model. The upper plots show the results of WDM runs where the dashed, dotted-dashed, and dotted curves are for $m_X = 2, 3, 4\,\mathrm{keV}$, respectively. The lower plots show CDM runs where the dashed, dotted-dashed, and dotted curves are for CDM models with $f_*/f_{*\mathrm{fid}} = 0.03, 0.1, 0.5$, respectively.

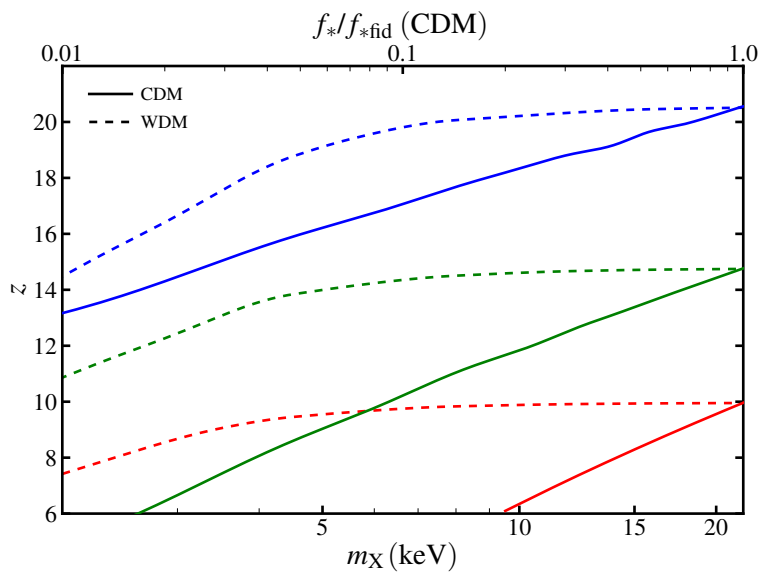


Figure 6.4: 'Critical points' in the mean 21-cm signal. Redshifts of critical points for CDM (solid curves) and WDM (dashed curves) models. For CDM curves, the redshifts of the critical points are plotted as a function of f_*, which can be read from the top horizontal axis. For WDM curves, the critical point redshifts are plotted as a function of m_X, the values of which can be read from the lower horizontal axis. In descending order from the right, the curves are the redshifts z_min (blue), z_h (green), and z_r (red) for each model.

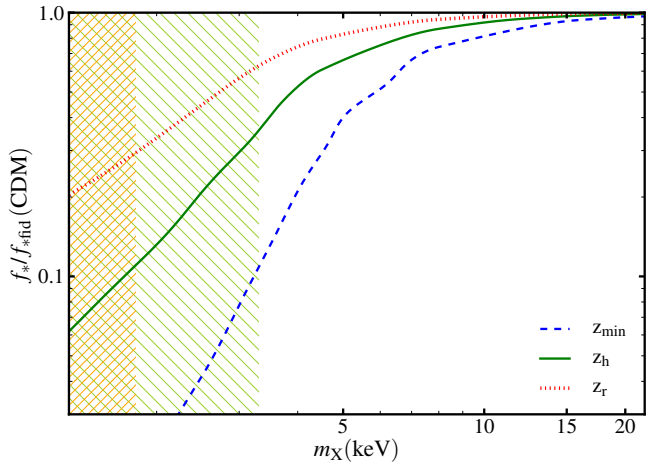


Figure 6.5: Parameter space curves $z_e(f_*|\mathrm{CDM}) = z_e(m_\mathrm{X}|\mathrm{WDM})$ for various critical points $z_e \in \{z_\mathrm{min}, z_\mathrm{h}, z_\mathrm{r}\}$. The orange (green) hatched region shows models disfavoured by observations of GRBs (the Lyman-α forest) from Ref. [103] (Ref. [109]).

shows the form of $f_*(z)$ needed to reproduce the mean 21-cm signal for WDM with $m_\mathrm{X} = 2, 4\,\mathrm{keV}$. At high redshifts ($z \gtrsim 15, 25$ for $m_\mathrm{X} = 2, 4\,\mathrm{keV}$), f_* is more than an order of magnitude smaller than its value at the end of reionization to compensate for the delay of structure formation in WDM. When more massive halos start to collapse (near $z = 10, 20$ for $m_\mathrm{X} = 2, 4\,\mathrm{keV}$), f_* rises quickly by roughly an order of magnitude to mimic the more rapid change of the collapse fraction in WDM and finally levels off during reionization. While this evolution of f_* may be possible, it seems contrived without an underlying model of such evolution.

Even in cases where f_* evolves in such a way as to mimic the mean brightness temperature in WDM, one can differentiate between WDM and CDM by examining the spectrum of perturbations in the 21-cm signal at certain points in its evolution. Perturbations in the UV and X-ray fields add power to the 21-cm power spectrum Δ^2_{21} on large scales. Since the bias of sources in WDM can be greater than that in CDM [135], more power is added on large scales in WDM than in CDM. This effect is most easily seen at times when inhomogeneities in x_α or T_K are at their maximum. Fig. 6.7 shows the evolution of the power spectrum for the modes $k = 0.08\,\mathrm{Mpc}^{-1}$ and $k = 0.18\,\mathrm{Mpc}^{-1}$, showing a three peak structure, where the peaks from

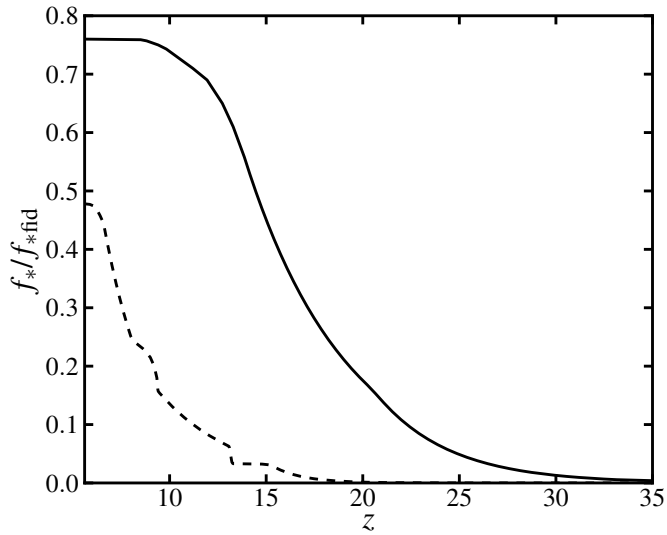


Figure 6.6: Evolution of $f_*(z)$ in CDM required to match the mean brightness temperature $\delta\bar{T}_\mathrm{b}$ in WDM with $m_\mathrm{X} = 2\,\mathrm{keV}$ (dashed) and $m_\mathrm{X} = 4\,\mathrm{keV}$ (solid). All other parameters are set to their values in the fiducial CDM model.

high to low redshift are associated with inhomogeneities in x_α, T_K, and x_HI, respectively. When inhomogeneities in T_K are at their maximum, the power at $k = 0.08, 0.18\,\mathrm{Mpc}^{-1}$ can be boosted in WDM by as much as a factor of $2.4, 2.0\,(1.3, 1.1)$ for $m_\mathrm{X} = 2\,\mathrm{keV}$ ($m_\mathrm{X} = 4\,\mathrm{keV}$). When inhomogeneous in x_α are near their height, the power at $k = 0.08\,\mathrm{Mpc}^{-1}$ can be increased by a factor of 1.5 (1.2) for WDM with $m_\mathrm{X} = 2\,\mathrm{keV}$ ($m_\mathrm{X} = 4\,\mathrm{keV}$).

Current and next generation interferometric radio telescopes may be used to detect the boost in power associated with WDM models. The dotted curves in Fig. 6.7 show forecasts for the 1σ power spectrum thermal noise levels for 2000 hours of observation time, computed by Ref. [115], for the Murchison Widefield Array (MWA), the Square Kilometre Array (SKA), and for the proposed Hydrogen Epoch of Reionization Array (HERA). This estimate is quite conservative in that it ignores the contribution of foreground-contaminated modes [136]. From these forecasts, we can see that the MWA may be able to at least marginally detect the boost in power for the $m_\mathrm{X} = 2\,\mathrm{keV}$ model at the reionization and X-ray heating peaks. In addition, these estimates indicate that next generation instruments will be able to easily measure the excess of power at these scales for $m_\mathrm{X} = 2, 4\,\mathrm{keV}$ models over

a wide range of redshifts.

The 21-cm power spectrum during a redshift near the time when $T_{\rm K}$ is at its most inhomogeneous state is plotted in Fig. 6.8 for WDM with $m_{\rm X} = 2, 4\,{\rm keV}$ and their CDM counterparts. One can see that the boost in power in WDM may continue to k values lower than those used in Fig. 6.7. In particular, the power near $k = 0.01\,{\rm Mpc}^{-1}$ in WDM models with $m_{\rm X} = 2\,{\rm keV}$ ($m_{\rm X} = 4\,{\rm keV}$) may be larger by a factor of 3 (1.3) as compared to in CDM models at these times.

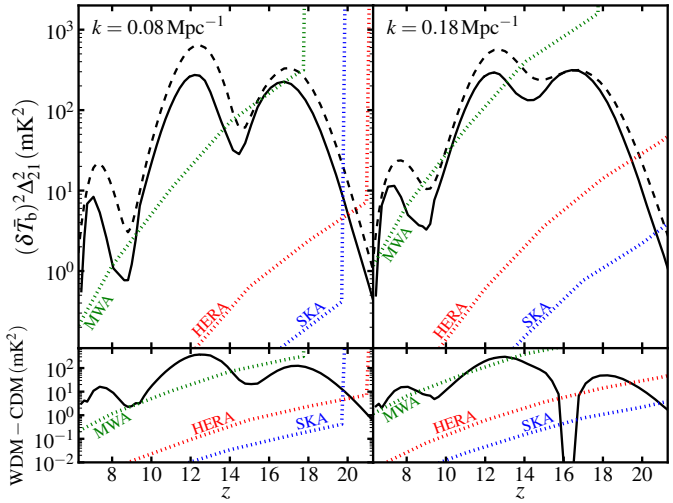
Finally, we mention that for simplicity we have chosen to vary only one astrophysical property. By allowing other astrophysical parameters to vary as a function of redshift, most notably $M_{\rm min}$, it might be possible to produce a 21-cm power spectrum degenerate with WDM throughout the redshifts under investigation and we leave this question for future work.

6.7 Conclusions

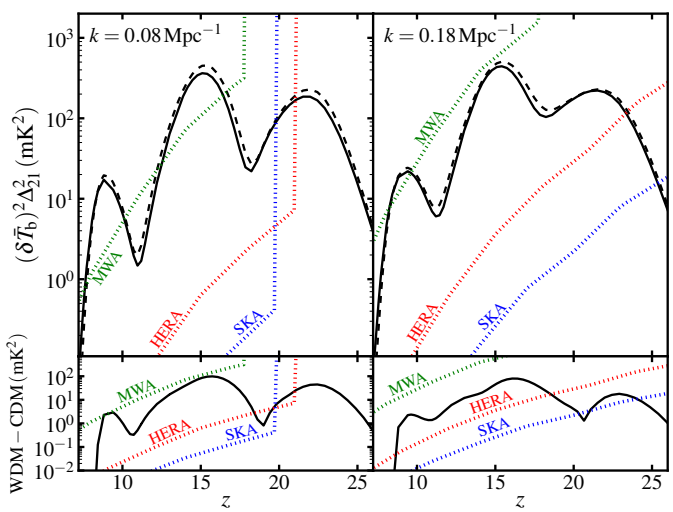
In warm dark matter models, the abundance of small halos is suppressed, which can leave a strong imprint at high redshifts. Since structure formation is delayed but more rapid in WDM, the mean 21-cm signal will follow suit, resulting in a delayed, deeper and more narrow absorption trough. These effects can easily be seen in the global 21-cm signal for WDM with free-streaming lengths above current observational bounds for thermal relic masses as high as $m_{\rm X} \sim 10{-}20\,{\rm keV}$ ($R_{\rm c}^0 \sim 6{-}13\,{\rm kpc}$).

Suppressing the photon-production efficiency of astrophysical sources can delay the 21-cm signal as well. As such, to discriminate between WDM and CDM models by measuring the redshift of reionization, the photon-production efficiency must be known to within a factor of 3.0, 1.8, and 1.4 for WDM with $m_{\rm X} = 2, 3, 4\,{\rm keV}$ ($R_{\rm c}^0 \approx 86, 54, 39\,{\rm kpc}$), respectively. Since the impact of WDM is larger at higher redshifts, if milestones in the mean 21-cm signal that occur at higher redshift are used to differentiate WDM and CDM models, the precision to which this efficiency must be known decreases. For example, if measuring the redshift of the minimum of the mean 21-cm signal (during the astrophysical epoch of the signal) the efficiency must only be known within a factor of 50, 13, and 4.8 for $m_{\rm X} = 2, 3, 4\,{\rm keV}$, respectively.

If the star-formation remains approximately constant over the range of redshifts under consideration, degeneracy between CDM and WDM models may be broken by examining the gradient of the mean 21-cm signal, which is larger in WDM due to its more rapid pace of structure formation. In



(a)



(b)

Figure 6.7: Evolution of the angle-averaged power spectrum of δT_b for WDM with (a) $m_\mathrm{X} = 2\,\mathrm{keV}$ and (b) $m_\mathrm{X} = 4\,\mathrm{keV}$. The top panels show power spectra at $k = 0.08, 0.18\,\mathrm{Mpc}^{-1}$ for WDM (dashed) and the CDM model (solid). CDM models have $f_*(z)$ chosen to reproduce the global 21-cm signal found for the respective WDM model. The bottom panels show the difference in the power spectrum between WDM and CDM models. Dotted curves show forecasts for the $1-\sigma$ power spectrum thermal noise as computed in Ref. [115] with 2000h of observation time. The dotted green, blue, and red curves are the forecasts for the MWA, SKA, and HERA, respectively.

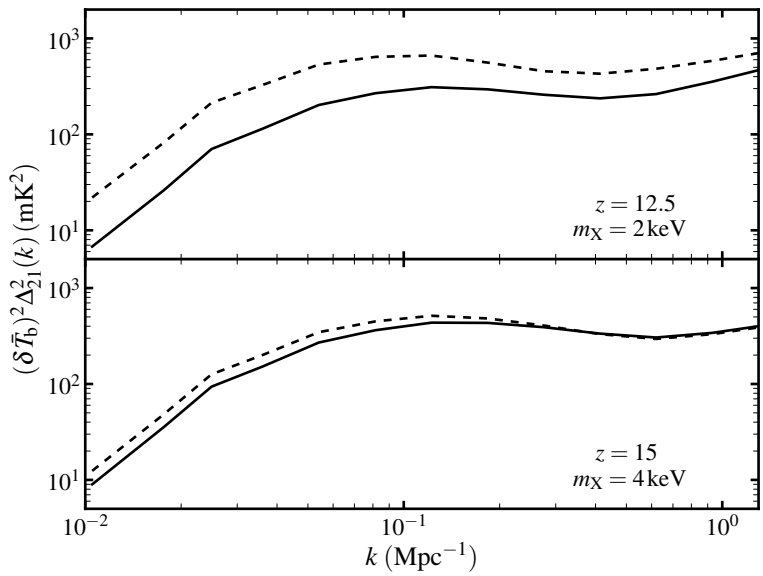


Figure 6.8: Power spectrum of the brightness temperature δT_b. The top panel shows the power spectrum at $z = 12.5$ for WDM with $m_X = 2\,\mathrm{keV}$ (dashed) and CDM (solid). In the CDM model, $f_*(z)$ evolves as shown in Fig. 6.6 such that it reproduces the global signal in the WDM model. Similarly, the bottom panel shows the power spectrum at $z = 15$ for WDM with $m_X = 4\,\mathrm{keV}$ (dashed) and CDM (solid) with $f_*(z)$ chosen to match the global signal in this WDM model. The power spectrum of each model is plotted at a redshift near where the X-ray background is at its most inhomogeneous state in its respective model.

addition, the spectrum of perturbations in the 21-cm signal may as well be used to break this degeneracy, as the 21-cm power spectrum in WDM has an excess of power on large scales owing to the stronger biasing of sources in WDM. This is true even if the photon-production efficiency evolves with redshift in such a way as to reproduce with CDM the global 21-cm signal in WDM models. For WDM with $m_{\rm X} = 2\,{\rm keV}$ ($m_{\rm X} = 4\,{\rm keV}$), the power in the 21-cm signal at $k = 0.08, 0.18\,{\rm Mpc}^{-1}$ can be increased by a factor as high as $2.4, 2.0\,(1.3, 1.1)$ as compared to that in CDM. Power spectrum measurements made by current interferometric telescopes, such as the MWA, should be able to discriminate between CDM and WDM models with $m_{\rm X} \lesssim 3\,{\rm keV}$, while next generation telescopes will easily be able differentiate between CDM and all relevant WDM models.

In this work, we assume that atomically-cooled halos drive the 21-cm signal. If instead smaller, molecularly-cooled halos, whose production is suppressed in WDM, play a significant role in producing the 21-cm signal in CDM, then the effects differentiating WDM from CDM described above would be even more pronounced. On the other hand, if star-formation was not efficient in halos with $T_{\rm vir} = 10^4\,{\rm K}$, the differences between CDM and WDM in the 21-cm signal would be diminished.

Chapter 7

Forecasting 21-cm BAO Experiments

7.1 Introduction

In the search for the underlying nature of dark energy, precise measurements of the expansion of the Universe are essential for constraining models of dark energy [5, 6, 137]. One such class of experiments are designed to measure the baryon acoustic oscillations (BAO) at different redshifts, from which an expansion history can be inferred. With many new experiments designed to measure the BAO on the horizon, forecasting their ability to measure the BAO and constrain dark energy parameters plays an important role for their design. These forecasts can be used to optimize the design and operation of the experiment and estimate the impact of noise and foregrounds on the measurement of the BAO and ultimately on the dark energy equation of state.

This chapter describes the development of software used to make such forecasts and some of the forecasts made for the CHIME[51] telescope. These forecasts estimate the ability of an experiment to measure the matter power spectrum, projecting these uncertainties onto measurements of the BAO, expansion parameters, and finally onto the dark energy equation of state.

For this analysis, we use the standard parameterization of the dark energy equation of state [137]

$$w_{\mathrm{DE}}(z) = w_0 + w_a[1 - a(z)] = w_0 + w_a \frac{z}{1+z}, ‘ \tag{7.1}$$

and thus our ultimate goal is to determine the precision in which the parameters w_0 and w_a can be measured. With this parameterization, the Hubble

[51] http://chime.phas.ubc.ca

rate is given by

$$H^2(z) = H_0^2\Big[\Omega_{\rm m}(1+z)^3 + \Omega_{\rm k}(1+z)^2 + \Omega_\Lambda(1+z)^{3(1+w_0+w_a)}\exp(-3w_a z/(1+z))\Big]. \quad (7.2)$$

We begin this chapter by describing the physics of the BAO and its potential for constraining $w_{\rm DE}$ followed by a brief discussion of other methods that can be used to constrain dark energy parameters, which can be used in conjunction with BAO experiments to provide more stringent constraints. The remainder of this chapter describes forecasting methods for BAO experiments and their resulting forecasts, where emphasis is given to cylinder transit telescopes such as CHIME.

7.2 Constraining Dark Energy Parameters

Each experiment designed to constrain the dark energy equation of state will have different systematic errors, may cover a different redshift range, and will produce a different contour in the $w_0 - w_a$ plane. As such, the constraints on dark energy parameters improves greatly when the results from different observational methods are combined. In particular, the report from the *Dark Energy Task Force* (DETF) [137] endorses pursuing multiple techniques for measuring the dark energy equation of state that includes measurements of BAO, type Ia supernovae (SN), galaxies clusters (CL), and weak lensing (WL).

The observables of a dark energy experiment are in some way affected by either $H(z)$ directly or a quantity dependent on it (i.e. $D_A(z)$, $G(z)$, etc...), which in turn is dependent on the dark energy parameters. As will be discussed in more detail in Section 7.3.2, the BAO can be measured in directions both parallel and perpendicular to the line of sight, so has the potential of measuring both $H(z)$ and $D_A(z)$ separately. SN Ia can act as standard candles as their absolute luminosity at peak brightness occurs nearly at the same point for every SN Ia. As standard candles, SN Ia can be used to measure the luminosity distance $D_L = (1+z)^2 D_A$. SN Ia provided some of the first definitive observational evidence for a late period of acceleration and remains an important tool for constraining the dark energy equation of state. CL abundances $dN/dMd\Omega dz$ observed in a region of solid angle $d\Omega$ and in a redshift bin dz can be compared to the mass function dn/dM calculated assuming a particular dark energy model. The

mass function may be calculated analytically using the methods described in Section 2.5 or more precisely using N-body simulations. Dark energy effects the mass function through both the growth function $G(z)$ via the rms of density fluctuations $\sigma_R^2(z) = (G(z)/G(z=0))^2\sigma_R^2(z=0)$ as well as through the combination $D_A^2(z)/H(z)$ needed to convert between a comoving and physical volume. WL measures the statistical distortion of the images of galaxies that pass by large masses. The level of distortion depends on the growth of the density fluctuations that distort the image, hence on the growth function $G(z)$ as well as on the distances between the lens, source, and observer and so is sensitive to expansion history as well.

One of the ultimate goals of these experiments is to place constraints on w_{DE} and its time evolution. In this light, the DEFT figure of merit (FOM), defined as the reciprocal of the area of the 95% confidence contour in the $w_0 - w_a$ plane (marginalizing over all over parameters), can be used to evaluate the effectiveness of an experiment to constrain w_{DE}, where a larger FOM indicates more constraining power.

7.3 Measuring the Acoustic Scale

7.3.1 The Sound Horizon

One of the most basic quantities characterizing the acoustic oscillations in the photon-baryon fluid is the sound horizon. The sound speed c_{s} of the fluid is given by

$$c_{\mathrm{s}}^2 = \frac{\delta P}{\delta \rho} \approx \frac{\delta\rho_\gamma/3}{\delta\rho_\gamma + \delta\rho_{\mathrm{b}}} = \frac{1}{3(1+R_{\mathrm{b}})}, \tag{7.3a}$$

$$R_{\mathrm{b}} = \frac{\delta\rho_{\mathrm{b}}}{\delta\rho_\gamma} = \frac{\dot\rho_{\mathrm{b}}}{\dot\rho_\gamma} = \frac{3}{4}\frac{\rho_{\mathrm{b}}}{\rho_\gamma} = \frac{3}{4}\frac{\Omega_{\mathrm{b}}}{\Omega_\gamma}a. \tag{7.3b}$$

The comoving sound horizon r_{s} is then

$$r_{\mathrm{s}}(z) = \int_0^{\eta(z)} d\tilde\eta c_{\mathrm{s}}(\tilde\eta) = H_0^{-1}\int_0^{a(z)} \frac{d\tilde a}{\tilde a^2 E(\tilde a)\sqrt{3(1+R_{\mathrm{b}}(\tilde a))}}, \tag{7.4}$$

where $E(a) = H(a)/H_0$. We will be evaluating the sound horizon during matter domination when $E(a) = \sqrt{\Omega_{\mathrm{m}}}\sqrt{a+a_{\mathrm{eq}}}/a^2$, where a_{eq} is the scale factor at matter-radiation equality. With this, the comoving sound horizon is then

$$r_{\mathrm{s}}(z) = \frac{4}{3}H_0^{-1}\sqrt{\frac{\Omega_\gamma}{\Omega_{\mathrm{m}}\Omega_{\mathrm{b}}}}\ln\left(\frac{\sqrt{1+R_{\mathrm{b}}(z)}+\sqrt{R_{\mathrm{b}}(z)+R_{\mathrm{eq}}}}{1+\sqrt{R_{\mathrm{eq}}}}\right), \tag{7.5}$$

where $R_{\rm eq} \equiv R_{\rm b}(a_{\rm eq})$.

We are interested in the final comoving sound horizon when decoupling occurs. We define the decoupling time of a species as when its optical depth drops to unity. The optical depth for the photons τ can be found by integrating $\dot{\tau} = n_e \sigma_{\rm T} a$, where n_e is the number density of free electrons and $\sigma_{\rm T}$ is the Thomson cross section. The optical depth for the baryons $\tau_{\rm d}$ can be found from $\dot{\tau}_{\rm d} = \dot{\tau}/R_{\rm b}$, where the factor of $R_{\rm b}$ accounts for the difference in population between the baryons and photons. Since there are more photons than baryons at this time ($R_{\rm b} < 1$), the photons decouple at a redshift z_* slightly before the decoupling of the baryons at redshift $z_{\rm d}$. The redshifts z_* and $z_{\rm d}$ can found analytically using fitting formulas [28, 138], which for $z_{\rm d}$ is

$$\begin{aligned} z_{\rm d} &= \frac{1291\omega_{\rm m}^{0.251}}{1+0.659\omega_{\rm m}^{0.828}}(1+b_1\omega_{\rm b}^{b_2}), \\ b_1 &= 0.313\omega_{\rm m}^{-0.419}(1+0.607\omega_{\rm m}^{0.674}), \\ b_2 &= 0.238\omega_{\rm m}^{0.223}. \end{aligned} \tag{7.6}$$

The redshifts of decoupling and the sound horizon at these times can be inferred by the use of CMB data. The values inferred by Planck [29] are

$$z_* = 1090.37 \pm 0.65, \qquad r_{\rm s}(z_*) = 144.75 \pm 0.66\,{\rm Mpc}, \tag{7.7a}$$

$$z_{\rm d} = 1059.29 \pm 0.65, \qquad r_{\rm s}(z_{\rm d}) = 147.53 \pm 0.64\,{\rm Mpc}. \tag{7.7b}$$

7.3.2 Baryon Acoustic Oscillations

Using the BAO as a standard ruler, by measuring its size at a variety of redshifts we can reconstruct an expansion history. The BAO manifests itself as a bump in the correlation function and so will appear as an oscillation in the power spectrum. The BAO may potentially be measured in both radial and perpendicular directions, which appear as an angular separation φ and redshift separation Δz that are related to the expansion parameters by

$$\varphi(z) = \frac{r_{\rm d}}{(1+z)D_A(z)}, \tag{7.8a}$$

$$\Delta z(z) = r_{\rm d} H(z)/c, \tag{7.8b}$$

where $r_{\rm d} = r_{\rm s}(z_{\rm d})$. The first generation of BAO detections did not have sufficient data to accurately measure the parallel and perpendicular BAO scales separately and instead used the spherically averaged measure

$$[\varphi(z)^2 \Delta z(z)]^{1/3} = \frac{r_{\rm d}}{[c(1+z)^2 D_A(z)^2/H(z)]^{1/3}}. \tag{7.9}$$

With new experiments such as CHIME that will be able to rapidly map large areas of the sky over a wide range of redshifts, D_A and H may be measured separately, increasing the constraining power of our telescope.

7.4 Fisher Matrix Formalism

Creating a detailed model of the full likelihood function for a forecast of an experiment is often a difficult task. For the purposes of forecasting, it is often more useful to introduce some assumptions to simplify this task [1, 5]. In this vein, we assume that we have a fiducial model that is sufficiently accurate such that it is reasonable to find the maximum likelihood by expanding the likelihood function L about the fiducial model, which yields

$$\ln L(\boldsymbol{\theta}) \approx \ln L(\tilde{\boldsymbol{\theta}}) + \left.\frac{\partial \ln L(\boldsymbol{\theta})}{\partial \theta_i}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} (\theta_i - \tilde{\theta}_i) + \frac{1}{2}\left.\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} (\theta_i - \tilde{\theta}_i)(\theta_j - \tilde{\theta}_j), \quad (7.10)$$

where the likelihood is a function of our model parameters $\boldsymbol{\theta}$ that have values $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ in the fiducial mode and assume that $\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}$ is small. If $\boldsymbol{\theta}$ is near the maximum likelihood values, after taking the expectation value of the above expression, the first derivative term should vanish (or at least be small). The Fisher matrix $\mathbf{F}$ then can be approximated as

$$F_{ij} = -\left\langle \frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right\rangle \approx -\left.\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}. \quad (7.11)$$

Assuming that the parameters $\boldsymbol{\theta}$ are normally distributed, then Eq. (7.11) implies that the Fisher matrix is approximately equal to the inverse of the covariance matrix $\mathbf{C}_\theta$ for the parameters $\boldsymbol{\theta}$. The likelihood function is then approximately

$$L \approx \frac{1}{\sqrt{(2\pi)^n |\mathbf{F}^{-1}|}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^{\mathrm{T}} \mathbf{F} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})\right), \quad (7.12)$$

where $|\mathbf{F}^{-1}|$ is the determinant of the inverse of the $n \times n$ fisher matrix $\mathbf{F}$. Even if the parameters $\boldsymbol{\theta}$ not not have Gaussian uncertainties, the Fisher matrix is a useful measure since as long as $\boldsymbol{\theta}$ are unbiased estimators of the true values with covariance matrix $\mathbf{C}_\theta$, the Cramér-Rao inequality implies that $\mathbf{C}_\theta \geq \mathbf{F}^{-1}$. We proceed assuming Gaussian statistics and that the

chosen fiducial model has parameter values close to the maximum likelihood values.

We now summarize frequently used operations on the Fisher matrix. If we have the Fisher matrix $\mathbf{F}_\theta$ for parameters $\boldsymbol{\theta}$, the Fisher matrix $\mathbf{F}_\phi$ for parameters $\boldsymbol{\phi}$ can be calculated by use of the Jacobian $J_{ij} = \partial\theta_i/\partial\phi_j$ evaluated in the fiducial model. Since $\theta_i - \tilde{\theta}_i = J_{ij}(\phi_j - \tilde{\phi}_j)$, from Eq. (7.12) the Fisher matrix transforms as

$$\mathbf{F}_\phi = \mathbf{J}^T \mathbf{F}_\theta \mathbf{J}. \tag{7.13}$$

To maximize the likelihood with respect to a parameter, we simply remove the row and column of the Fisher matrix corresponding to that variable. Marginalizing over a variable amounts to removing the row and column corresponding to that variable from the covariance matrix. If we order our Fisher matrix as

$$\mathbf{F} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{M} \end{pmatrix}, \tag{7.14}$$

where $\mathbf{M}$ is the submatrix that has both rows and columns corresponding to variables that we wish to marginalize over, then the marginalized Fisher matrix $\mathbf{F}_\mathrm{M}$ is given by

$$\mathbf{F}_\mathrm{M} = \mathbf{A} - \mathbf{B}\mathbf{M}\mathbf{B}^T. \tag{7.15}$$

Since adding a Gaussian prior amounts to summing the inverse of covariance matrices, a prior can be added by simply adding Fisher matrices, assuming that they use the same fiducial model and have their rows and columns ordered in the same manner.

7.5 Measuring the 21-cm Power Spectrum

In this section we outline our method for forecasting the uncertainties in measuring the 21-cm power spectrum for a cylinder transit telescope. This method, based on the analysis in Refs. [139, 140], provides a straightforward and computationally cheap procedure for estimating power spectrum uncertainties. More complex and computationally expensive forecasting methods, which also include the effects of foreground subtraction, can be found in Refs. [141, 142]. For our forecast, we use the cosmological parameters $\Omega_\mathrm{m} = 0.266, \Omega_\mathrm{b} = 0.0449, \Omega_\mathrm{k} = 0, n_\mathrm{s} = 0.963, h = 0.71, \sigma_8 = 0.8$, consistent with WMAP7.

The visibilities measured by a telescope may be processed in several ways to produce maps and power spectra. In the flat-sky approximation, each

visibility measures a small number of Fourier modes on the sky. Extending this notion to a wide field of view telescope, a visibility measures a finite set of spherical harmonics on the sky. Alternatively, the process of beamforming combines the measured visibilities to form localized beams on the sky, which we will employ for our forecasting. In this process, the localized beams are formed from the spatial Fourier transform of the feed responses or visibilities. These beams can be characterized by the point spread function (or *dirty* or *synthetic beam*)

$$\mathrm{PSF}(\mathbf{p}) = |\mathrm{A}(\mathbf{p})|^2 \int \mathrm{d}^2\Delta\mathbf{r}\mathrm{S}(\Delta\mathbf{r})\mathrm{e}^{2\pi\mathrm{i}\mathbf{p}\cdot\Delta\mathbf{r}}, \tag{7.16}$$

where $\Delta\mathbf{r}$ is the baseline vector, $\mathbf{p}$ is the wave vector of the radiation, $A(\mathbf{p})$ is the primary beam response, and $S(\Delta\mathbf{r})$ is a sampling function equal to unity for each baseline measured and zero otherwise. We have assumed here that all primary beams are identical. An image of the sky, known as the *dirty image*, can be found by convolution of the point spread function and the true sky intensity.

Each cylinder in our telescope will have N_f equally spaced (dual polarization) feeds along its focal line and will consist of N_cyl cylinders. To start with, consider the response of a single cylinder that lies in the $\hat{e}_z = 0$ plane with its focal line along the $\hat{e}_y$ axis, so that the feed locations are given by $\mathbf{r}_m = (0, md_\mathrm{f}, 0)$ for $m = 0, \ldots, N_\mathrm{f} - 1$. By sampling $\mathbf{p}$ at discrete locations, the Fourier transform in Eq. (7.16) along the cylinder can be expressed as a discrete Fourier transform, where p_y is sampled at $(p_y)_n = n/N_\mathrm{f}d_\mathrm{f}$ for $n = -N_\mathrm{f}/2 + 1, \ldots, N_\mathrm{f}/2$. If θ is the angle between zenith and $\mathbf{p}$ projected into the $\hat{e}_y - \hat{e}_z$ plane then we sample $\sin\theta$ at even intervals with

$$\sin\theta_n = \frac{(n - \frac{1}{2})\lambda}{N_\mathrm{f}d_\mathrm{f}} \qquad n = -\frac{N_\mathrm{f}}{2} + 1, \ldots, \frac{N_\mathrm{f}}{2}. \tag{7.17}$$

The resolution of the telescope is set by the longest baseline in the array $L = N_\mathrm{f}d_\mathrm{f}$, assumed to be large enough for small angle approximations to be valid. Our feed and cylinder spacing will be such that L is the longest baseline in both North-South and East-West directions. The resolution of the synthetic beams is $\Delta\theta \approx \lambda\ /L$, which by Eq. (7.17) is also the angular sampling rate. Fourier transforming between the cylinders segments each synthetic beam into multiple beams in the azimuth direction to improve the resolution in that direction.

The foreground emissions, predominately from synchrotron radiation from our Galaxy as well as from extragalactic sources, will dominate over

the cosmological 21-cm signal. However, these foregrounds have very smooth spectra, unlike the 21-cm signal, and thus there are a variety of techniques that rely on this difference in spectra to separated the foregrounds from the desired signal [141, 142, 148, 149, 150]. In the following we assume that foregrounds may be completely cleaned from the 21-cm power spectrum at scales near the BAO scale. We direct the reader to Refs. [141, 142] for an in-depth analysis of the effect of foreground subtraction on the recovered 21-cm signal.

In a sidereal day, we can create a map of a large fraction of the sky. With adequate foreground subtraction, by relating angular and frequency scales to comoving position, the power spectrum of the underlying density field $\delta(\mathbf{x})$ may be measured. If the response of our telescope to the overdensity field is given by $W(\mathbf{x})\delta(\mathbf{x})$, where $W(\mathbf{x})$ is a window function appropriate for our experiment, then the 'raw' measured power spectrum $\Delta^2(\mathbf{k})$ can be written as [143, 144]

$$\Delta^2(\mathbf{k}) = |W(\mathbf{k})|^2 P(\mathbf{k}) + P_{\rm shot} + P_N, \tag{7.18}$$

where $P(\mathbf{k})$ is the 'true' underlying power spectrum appearing as the sample variance, $P_{\rm shot}$ is the shot noise, and P_N is the additional noise introduced by the instrument.[52] The power spectrum under investigation is the 21-cm brightness temperature power spectrum $P_{\rm 21cm}(\mathbf{k}, z) = \bar{T}_{\rm b}^2(z) b^2 P_{\rm m}(\mathbf{k}, z)$, where $P_{\rm m}$ is the matter power spectrum and b is the bias. The shot noise can be expressed as $P_{\rm shot} = 1/\bar{n}$ with $\bar{n}$ the expected average number density of emitters detectable by the experiment.

We can use $\hat{P}(\mathbf{k}) = \Delta^2(\mathbf{k}) - P_{\rm shot} - P_N$ as the estimator of $P(\mathbf{k})$. Assuming Gaussian statistics, the covariance of $\hat{P}(\mathbf{k})$ is simply $\langle \hat{P}(\mathbf{k})\hat{P}(\tilde{\mathbf{k}})\rangle = (P(\mathbf{k})|W(\mathbf{k})|^2 + P_{\rm shot} + P_N)^2 \delta_{\mathbf{k},\tilde{\mathbf{k}}}$. If in surveying a real space volume V_s we measure the average power in $\mathbf{k}$ within a k-space volume $V_{\mathbf{k}}$, then we may observe $N_{\mathbf{k}} = V_{\mathbf{k}} V_{\rm s}/2(2\pi)^3$ independent modes within the volume, and thus may reduce the variance of our power spectrum measurement by a factor of $1/N_{\mathbf{k}}$.[53] The covariance matrix for the 21-cm power spectrum measurements for modes $\mathbf{k}_i$ is then

$$C_{\rm P}(\mathbf{k}_i, \mathbf{k}_j) = 2\frac{(2\pi)^3}{V_{\mathbf{k}_i} V_{\rm s}} \left[\bar{T}_{\rm b}^2(z) b^2 (P(\mathbf{k}_i)|W(\mathbf{k}_i)|^2 + P_{\rm shot}) + P_N\right]^2 \delta_{ij}. \tag{7.19}$$

[52] In this chapter, we refer to the *dimensionful* power spectrum as simply the power spectrum, unless otherwise stated, so that the power spectrum for field ϕ is $P_\phi(\mathbf{k}) = V|\phi_{\mathbf{k}}|^2$ with V the volume.

[53] The extra factor of $\frac{1}{2}$ accounts for the fact that since the field is real valued, the modes $\mathbf{k}$ and $-\mathbf{k}$ are not independent of one another.

We will now specify the parameter values necessary to compute the covariance in Eq. (7.19) for our forecast. Since during the redshifts under consideration we expect $T_{\rm S} \gg T_\gamma$, the mean brightness temperature simplifies to [145]

$$\delta\bar{T}_{\rm b} \approx 0.1 \frac{\Omega_{\rm HI}}{10^{-3}} \frac{(1+z)^2}{H/H_0}\,{\rm mK}. \tag{7.20}$$

For the HI density parameter we set $\Omega_{\rm HI} b = 6.2 \times 10^{-4}$ [146] and take the bias as $b = 1$. The matter power spectrum can be expressed as $P_{\rm m}(k,\mu) = R(\mu)P_{\rm m}(k)$, where μ is the cosine of the angle between $\mathbf{k}$ and the line of sight, $R(\mu) = (1+\beta\mu^2)^2$ is the linear redshift-space distortion factor, $\beta = f/b$, and $f = (a/G)dG/da$ is the linear growth rate. We use the transfer function of Ref. [28] when computing the matter power spectrum.

For our forecast, we will divide our coverage in frequency into bins assuming constant cosmological values (i.e. H and D_A) in each. For a redshift bin of size Δz, the comoving real-space volume measured within the bin is given by

$$V_s = \frac{D_A^2(1+z)^2}{H}\Delta z \Omega_{\rm s}, \tag{7.21}$$

where $\Omega_{\rm s}$ is the solid angle covered by the survey.

The coverage in declination is determined by both the array configuration as well as the primary beams of the feeds. We can determine the locations of outermost synthetic beams by used of Eq. (7.17), which for large $N_{\rm f}$ is [54]

$$\sin\theta_{n_{\rm max}} = \frac{\lambda}{2d_{\rm f}}. \tag{7.22}$$

Synthetic beams close to the horizon may be dampened by the primary beam pattern $A(\mathbf{p})$. To account for the decreasing sensitivity close to the horizon, we approximate the survey solid angle by

$$\Omega_{\rm s} \approx 2\pi \int_{\theta_{\rm min}}^{\theta_{\rm max}} |A(\theta)|^2 \cos(\theta) d\theta, \tag{7.23}$$

where $\theta_{\rm min} = \min(\theta_{\rm lat} - \theta_{\rm n_{max}}/2, -\pi/2)$, $\theta_{\rm max} = \max(\theta_{\rm lat} + \theta_{\rm n_{max}}/2, \pi/2)$, and $\theta_{\rm lat}$ is the latitude of the telescope. We take the primary beam pattern in the meridian as the used in Ref. [141], which is proportional to the flux

[54] Note that Eq. (7.17) is more properly written with the addition of an arbitrary integer since this expression originates from inside the exponential in Eq. (7.16), which ensures that $|\sin\theta_{n_{\rm max}}| \leq 1$. We set $\theta_{n_{\rm max}} = \pi/2$ in cases where $\lambda/2d_{\rm f} > 1$.

passing through the ground-plane, so that

$$|A(\theta)|^2 = \begin{cases} \cos\theta & -\pi/2 \leq \theta \leq \pi/2 \\ 0 & \text{else} \end{cases}, \tag{7.24}$$

The parameters for our telescope used for the forecast are listed in Table 7.1. From these values we see that for wavelengths $\lambda < 2d_{\rm f}$ ($\nu \gtrsim 484\,\text{MHz}$ in our case) the synthetic beams will be aliased with directions pointed closer to the horizon. However, since these aliases appear closer to the horizon where the sensitivity of the primary beam is reduced, the impact of the aliasing is diminished, although not completely removed.

Parameter	Value
$\theta_{\rm lat}$	$49.5°$
$N_{\rm cyl}$	5
$N_{\rm f}$	256
$d_{\rm f}$	$31\,\text{cm}$
$T_{\rm sys}$	$50\,\text{K}$
$t_{\rm tot}$	2 yrs
ν	$400-800\,\text{MHz}$

Table 7.1: Telescope parameters for CHIME used for BAO forecasting.

We now specify the window function in Eq. (7.19), dealing with modes perpendicular and parallel to the line of sight separately. Assuming an adequately small angular resolution $\Delta\theta$, our resolution in comoving distances perpendicular to the line of sight is $\Delta x \approx (\lambda/L)D_A(1+z)$. We may then sample at the Nyquist rate $2\pi(1/(2\Delta x)) = \pi(L/\lambda)/D_A(1+z)$, which acts as our low-scale cutoff mode $k_{\perp\rm max}$ in the perpendicular direction. We assume a sharp cutoff at $k_{\perp\rm max}$ and take $W(\mathbf{k})$ to be a top hat function in this direction. As the frequency resolution of our telescope of $\sim 1\,\text{MHz}$ corresponds to scales smaller than those relevant for measuring the BAO, we forego adding such a cutoff in the $k_{\parallel}$ direction.

Assuming uncorrelated noise between feeds, the variance on a real-space pixel is given by the expression in Eq. (5.17), where we assume an identical system temperature for all feeds. We can conservatively estimate the integration time for a real-space pixel by assuming that a pixel is seen by each cylinder for a fraction $\Delta\theta/2\pi$ of each sidereal day so that the total integration time for a real-space pixel is

$$t_{\rm int} \approx \frac{\Delta\theta}{2\pi} N_{\rm cyl} t_{\rm tot}, \tag{7.25}$$

where $N_{\rm cyl}$ is the number of cylinders and $t_{\rm tot}$ is the total observation time of the experiment.[55] In addition, CHIME will be equipped with dual-polarization feeds, which in effect doubles the integration time.

The above approximations yield a position independent estimate of the instrument noise on a real-space pixel. The final step is to Fourier transform to k-space. Although the real-space pixels are arranged on concentric spheres, we are only interested in smaller scales near the BAO scale and will therefore use the flat-sky approximation here. As we approximate the real-space noise as being position independent, the Fourier transform to k-space is trivial and simply introduces a factor of $1/N_{\rm pix}$, where $N_{\rm pix}$ is the number of real-space pixels measured. The instrument noise power spectrum then becomes

$$P_N = V_{\rm pix} \frac{T_{\rm sys}^2}{t_{\rm int} \Delta\nu}, \tag{7.26}$$

where $\Delta\nu$ is frequencies resolution of our telescope and $V_{\rm pix}$ is a average volume of a real-space pixel.[56]

We take the shot noise as constant over the times of interest and, as in Ref. [139], set its value to $\bar{n} = 0.01\, h^3 {\rm Mpc}^{-3}$ as inferred from the catalogue of 4315 extragalactic HI sources with $z < 0.042$ measured by the HIPASS survey [147].

The different noise contributions to the measurement of the 21-cm power spectrum are plotted in Fig. 7.1, where each curve corresponds to a term in the square brackets in Eq. (7.19). At low redshifts, the sample variance dominates at all relevant scales, while the instrumental noise dominates at smaller scales for higher redshifts. As seen in Fig. 7.2, the contribution to the total survey volume is higher at larger redshifts, allowing more modes to be measured at lower frequencies.

[55] The integration time per real-space pixel in actuality varieties with declination. The integration time in Eq. (7.25) is an estimate of the lower bound and thus leads to a conservative estimate of the noise.

[56] A volume factor was added to Eq. (7.26) to conform to our definition of the power spectrum.

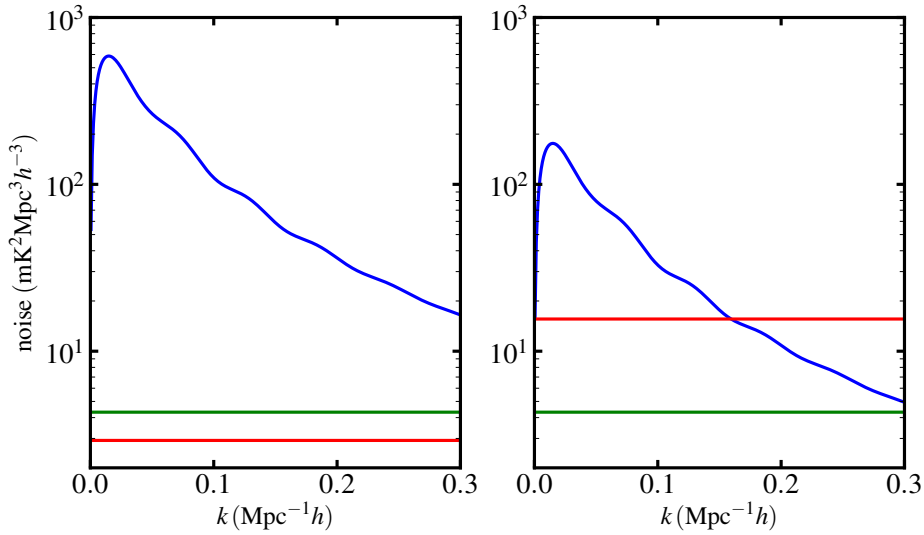


Figure 7.1: Contributions to the 21-cm power spectrum noise per mode as described in Eq. (7.19) at $z = 0.8$ (left) and $z = 2.5$ (right). The curves correspond to the different terms in the square brackets in Eq. (7.19), where the blue curve is the sample variance term at $\mu = 0$ and the green and red curves are the shot and instrument noise terms, respectively.

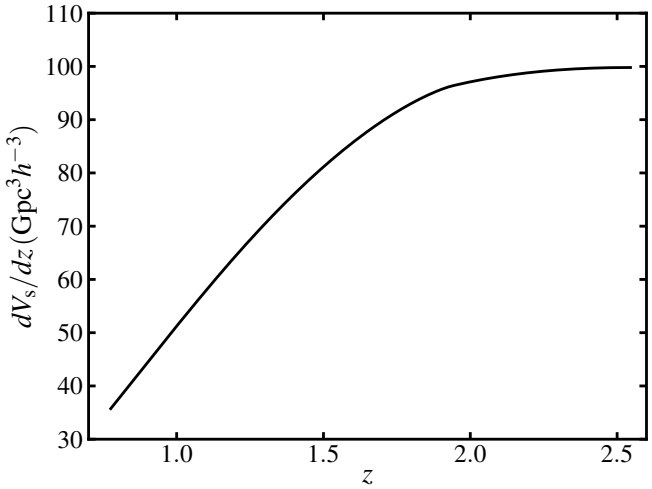


Figure 7.2: Survey volume per redshift over the CHIME band.

We are now in a position to calculate our forecasted power spectrum uncertainties. For our analysis, we divide our band into 16 redshift bins of equal size $\Delta z \approx 0.11$. The uncertainties in the power spectrum σ_P for spherically averaged k-bins with width $\Delta k = 0.02\,\mathrm{Mpc}^{-1}h$ can be seen in Fig. 7.3. The top panel shows the power spectrum uncertainties for z-bins centred at $z = 0.83, 1.61, 2.50$. At all redshifts, the trend of decreasing uncertainty with k at larger scales is due to the increased k-space volume available for spherical shells at higher values of k. At smaller scales, the uncertainties decrease with redshift as larger volumes can be surveyed for z-bins of equal size at higher redshifts. The resolution of the array decreases with redshift and thus the power spectrum uncertainties worsen at smaller scales for higher redshifts as compared to lower redshifts. The bottom panel of Fig. 7.3 shows the power spectrum uncertainties at $z = 1.61$ for the function P/P_{sm}, where P_{sm} is the 'smooth only' power spectrum which has the baryonic oscillations removed (see Section 7.6 for more details).

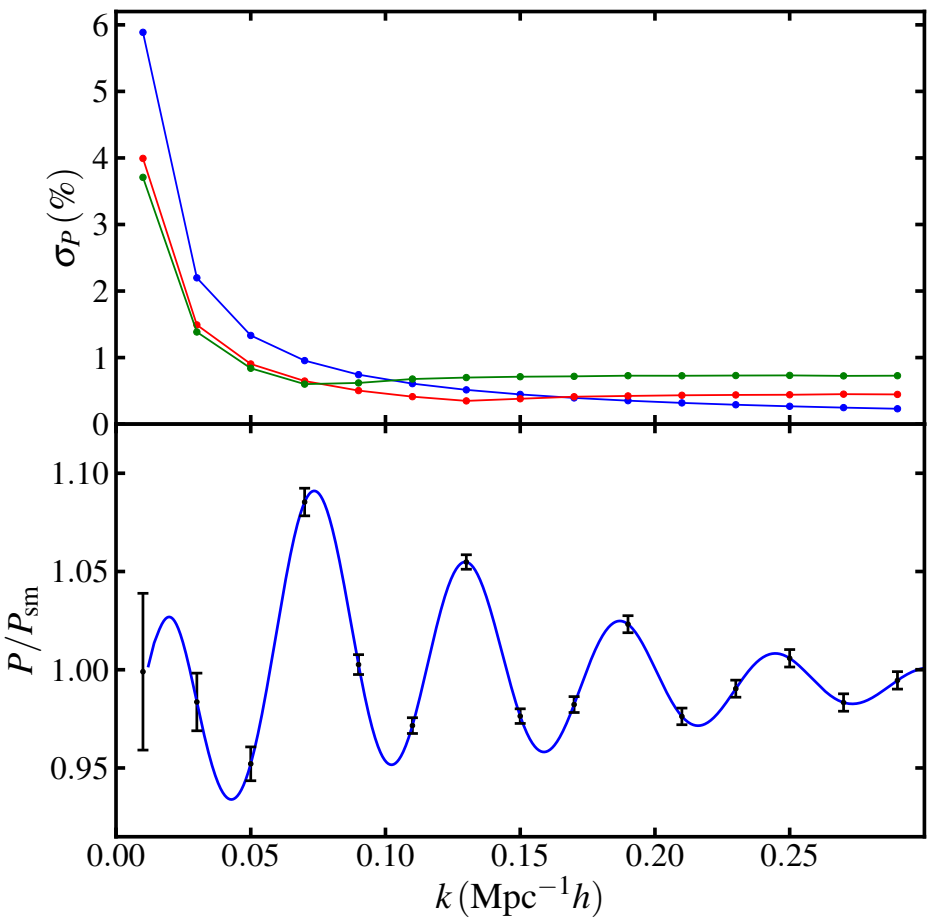


Figure 7.3: Forecasted power spectrum uncertainties for spherically averaged k-bins with width $\Delta k = 0.02\,\mathrm{Mpc}^{-1}h$ for redshift bins of size $\Delta z \approx 0.11$. Top: Power spectrum uncertainties at $z = 0.83$ (blue), 1.61 (red), and 2.50 (green). Bottom: Total power spectrum over 'smooth only' power spectrum with uncertainties at $z = 1.61$.

7.6 The 'Wiggles Only' Method

Once we have the uncertainties in the matter power spectrum, we would like to propagate these to uncertainties in our measurements of the BAO scale. In other words, we would like to find the Jacobian $\mathbf{J}_s$ to transform the Fisher matrix $\mathbf{F}_\mathrm{P} \approx \mathbf{C}_\mathrm{P}^{-1}$ for the measurements of the power spectrum

into the Fisher matrix $\mathbf{F}_s$ for the sound horizon. To accomplish this task, we use the 'wiggles only' method [140], which assumes we can remove the power from scales much larger then the BAO scale and models the remaining 'wiggly' BAO power spectrum as $P_b = P - P_{\rm sm}$, where P is the full power spectrum and $P_{\rm sm}$ is the power spectrum smoothed on a scale larger than the BAO scale.

7.6.1 Modelling the BAO Power Spectrum

Our task is to build a model of the BAO-only power spectrum that will contain the acoustic oscillations as well as effects that degrade this signal, which we now outline.

Acoustic Oscillations

The BAO manifests itself as preferred separation in the two-point correlation function at a particular angular and redshift separation, which can be translated into the comoving lengths $s_\perp$ and $s_\parallel$, respectively. As such, the un-deteriorated BAO signal is approximated as a ellipsoidal Dirac delta function with semi-axes $s_\parallel$ and $s_\perp$ parallel and perpendicular to the line of sight, respectively, the Fourier transform of which is $P_b \propto \mathrm{sinc(x)}$ with $x = \sqrt{(k_\perp s_\perp)^2 + (k_\parallel s_\parallel)^2}$.

Silk Damping

Although photons are tightly coupled to baryons at early times, near the time of recombination the mean free path of the photons grows to a non negligible size, allowing them to steam out of overdense regions and into underdense ones. This effect, known as Silk damping, smooths out perturbations on small scales. We can make a rough estimate of the Silk scale $\lambda_{\rm silk}$ by use of the mean free path $\lambda_{\rm mfp} = (\sigma_T n_e)^{-1}$, where σ_T is the Thomson cross-section. As the relevant timescale is the Hubble time H^{-1}, we expect the number of collisions to be of the order $H^{-1}/\lambda_{\rm mfp} = \sigma_T n_e H^{-1}$. For a Gaussian random walk, the rms distance travelled is given by the product of the mean free path and the root of the number of steps, so we can estimate the Silk scale by $\lambda_{\rm silk} \sim (\sigma_T n_e H)^{-1/2}$. Evaluating our estimate of $\lambda_{\rm silk}$ just prior to recombination and using $\sigma_T n_e \approx 2.307 \times 10^{-5}\,\mathrm{Mpc}^{-1}\omega_{\rm b}\mathrm{a}^{-3}$ results in $\lambda_{\rm silk} \sim 57\,\mathrm{Mpc}$ ($k_{\rm silk} \approx 0.11\,\mathrm{Mpc}^{-1}$). A more detailed examination can be preformed using the Boltzmann equations, from which the following fitting

formula was derived [28]

$$k_{\rm silk} = 1.6\omega_{\rm b}^{0.52}\omega_{\rm m}^{0.73}[1 + (10.4\omega_{\rm m})^{-0.95}]\,{\rm Mpc}^{-1}, \tag{7.27}$$

where it was found that the baryonic power spectrum is dampened as $P_b \propto D_{\rm silk} = \exp(-(k/k_{\rm silk})^{1.4})$.

BAO Power Spectrum Amplitude

We can now write our model for the linear BAO power spectrum amplitude as

$$P_{b,{\rm lin}}(\mathbf{k}) = \sqrt{8}\pi A_0 P_{0.2}{\rm sinc}({\rm x}(\mathbf{k}))D_{\rm silk}({\rm k}), \tag{7.28}$$

where is A_0 is a normalization constant and $P_{0.2}$ is the linear power spectrum evaluated at $k = 0.2\,h{\rm Mpc}^{-1}$. To find the normalization constant A_0, we first apply a low-pass filter with a sharp cutoff just below the BAO scale to our fiducial linear power spectrum to estimate $P_{\rm sm}$ and then subtract it from the total power spectrum to yield P_b. This estimate is subsequently fit to Eq. (7.28), which for the parameters used yields a best fit value of $A_0 = 0.42$.

Nonlinear Damping

Nonlinear behaviour distorts the BAO signal by displacing matter at the $\sim 10\,{\rm Mpc}$ scale, which smears out the BAO peak in the correlation function, thereby damping the oscillations in the power spectrum. Ref. [151] found that the nonlinear displacement distribution is well approximated by an elliptical normal distribution with rms values $\Sigma_\parallel$ and $\Sigma_\perp$ parallel and perpendicular to the line of sight, respectively. As the distorted correlation function is found by convolution with this elliptical Gaussian, the resulting effect on the power spectrum P_b is a multiplication by the Fourier transform of the distorting elliptical normal distribution $D_{\rm nl}$

$$D_{\rm nl} = \exp\left(-\frac{(k_\perp\Sigma_\perp)^2}{2} - \frac{(k_\parallel\Sigma_\parallel)^2}{2}\right). \tag{7.29}$$

The rms displacements parallel and perpendicular to the line of sight were found to be [151]

$$\Sigma_\parallel = (1+f)\Sigma_\perp, \tag{7.30}$$

$$\Sigma_\perp = 8.35h^{-1}{\rm Mpc}\frac{G(z)}{G(0)}\frac{\sigma_8}{0.8}. \tag{7.31}$$

7.6.2 Distance Uncertainties

With an analytical expression for the power spectrum as a function of $s_\parallel$ and $s_\perp$, we can find the Jacobian $\mathbf{J}_s$ to transform the Fisher matrix $\mathbf{F}_\mathrm{P}$ for the power spectrum into the Fisher matrix $\mathbf{F}_s$ for the parameters $\boldsymbol{\theta}_s = (\ln s_\perp^{-1}, \ln s_\parallel)$.

Including the effects of Silk and nonlinear damping, our model for P_b becomes

$$P_b(\mathbf{k}) = \sqrt{8}\pi A_0 P_{0.2} \mathrm{sinc}(\mathrm{x}(\mathbf{k})) D_\mathrm{silk}(\mathrm{k}) D_\mathrm{nl}(\mathbf{k}). \tag{7.32}$$

The Jacobian $\mathbf{J}_s$ is then found to be

$$\begin{aligned}(J_s)_{ij} &= \frac{\partial P(\mathbf{k}_i)}{\partial(\theta_s)_j} \\ &= \frac{\partial P_b(\mathbf{k}_i)}{\partial \ln x_i} \frac{\partial \ln x_i}{\partial(\theta_s)_j} \\ &= \sqrt{8}\pi A_0 P_{0.2} \left[\cos(x_i) - \mathrm{sinc}(\mathrm{x_i})\right] D_\mathrm{silk}(k_i) D_\mathrm{nl}(\mathbf{k}_i) \frac{\partial \ln x_i}{\partial(\theta_s)_j},\end{aligned} \tag{7.33}$$

where $x_i = x(\mathbf{k}_i)$ and

$$\frac{\partial \ln x}{\partial \ln s_\perp^{-1}} = \mu^2 - 1, \tag{7.34a}$$

$$\frac{\partial \ln x}{\partial \ln s_\parallel} = \mu^2. \tag{7.34b}$$

As $s_\parallel$ and $s_\perp$ are equivalent to an angular and redshift separation, respectively, from Eqs. (7.8) we see that

$$s_\perp \propto r_\mathrm{d}/D_A, \qquad s_\parallel \propto r_\mathrm{d} H, \tag{7.35}$$

so the Fisher matrix $\mathbf{F}_\mathrm{dist}$ for the variables $\boldsymbol{\theta}_\mathrm{dist} = (\ln D_A, \ln H, \Omega_\Lambda, \Omega_\mathrm{k}, \omega_\mathrm{m}, \omega_\mathrm{b})$ can be found from $\mathbf{F}_\mathrm{dist} = \mathbf{J}_\mathrm{dist}^T \mathbf{F}_s \mathbf{J}_\mathrm{dist}$, where the Jacobian $\mathbf{J}_\mathrm{dist}$ is given by

$$\begin{aligned}(J_\mathrm{dist})_{0j} &= \frac{\partial \ln s_\perp^{-1}}{\partial(\theta_\mathrm{dist})_j} = \frac{\partial \ln D_A}{\partial(\theta_\mathrm{dist})_j} - \frac{\partial \ln r_\mathrm{d}}{\partial(\theta_\mathrm{dist})_j}, \\ (J_\mathrm{dist})_{1j} &= \frac{\partial \ln s_\parallel}{\partial(\theta_\mathrm{dist})_j} = \frac{\partial \ln H}{\partial(\theta_\mathrm{dist})_j} + \frac{\partial \ln r_\mathrm{d}}{\partial(\theta_\mathrm{dist})_j},\end{aligned} \tag{7.36}$$

where the derivatives of r_d can be found using Eq. (7.5).

The uncertainties on D_A and H for our forecast as a function of redshift can be seen in Fig. 7.4. The fractional uncertainties on H show a decreasing trend towards higher redshifts as greater volumes are contained within each

spherical shell with equal redshift spacing. The same trend would be seen in D_A as well if not for the cutoff imposed on $k_\perp$, which decreases at increasing redshift. The correlation between D_A and H remains roughly at $\rho \approx 0.4$, which is close to the value one would expect from a spherically symmetric model.

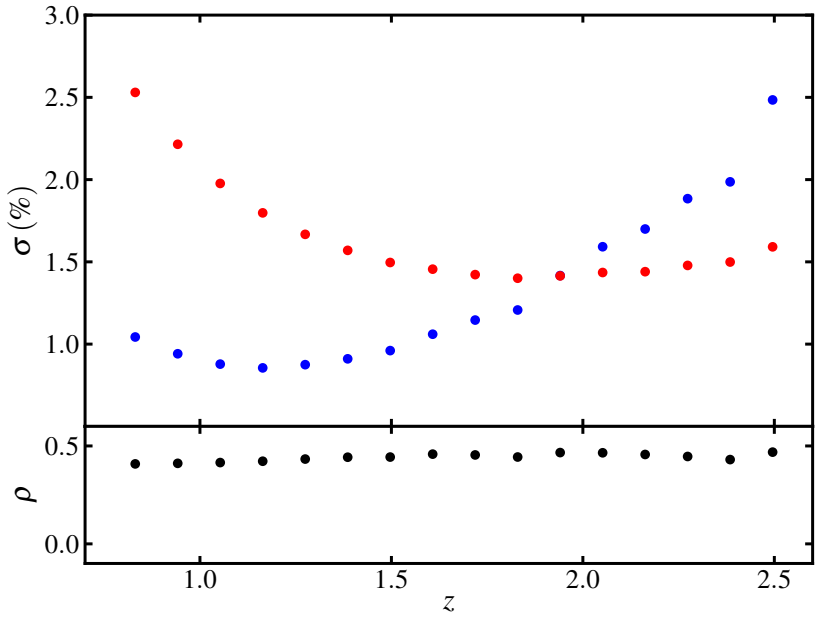


Figure 7.4: Top: Forecast uncertainties for D_A (blue) and H (red) as a function of redshift. Bottom: Correlation coefficient between D_A and H.

Most current BAO measurements only produce a high signal to noise measurement when averaged over angle, and therefore do not measure D_A and H separately. The 'dilation scale' D_V is often constrained instead, whose cube root consists of two factors of the angular diameter distance and one factor of the proper radial distance

$$D_V(z) = \left[(1+z)^2 D_A(z)^2 \frac{cz}{H(z)}\right]^{1/3} . \tag{7.37}$$

The forecasted uncertainties on D_V for CHIME can be seen in Fig. 7.5, along with uncertainties from current experiments. The CHIME band compliments these experiments by covering a higher redshift range, a band which is particularly sensitive to the dark energy equation of state (see Fig. 7.6).

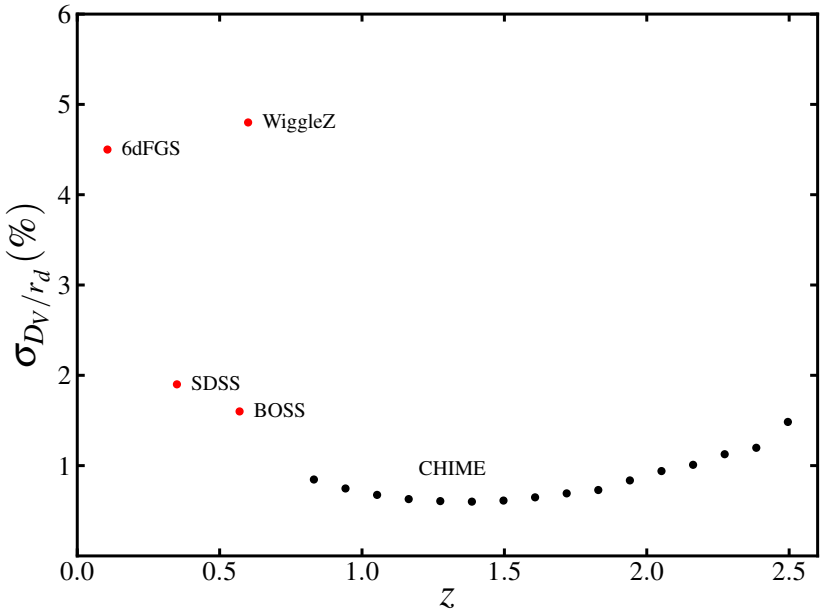


Figure 7.5: Measurement uncertainties on D_V from CHIME forecasts as well as from current detections from 6dFGS [13], SDSS [152], WiggleZ [153], and BOSS [15].

7.7 Dark Energy Constraints

With the uncertainties on D_A and H, we can finally forecast the uncertainties on the dark energy equation of state parameters w_0 and w_a. We form the Fisher matrix $\mathbf{F}_{\rm DE} = \mathbf{J}_{\rm DE}^T \mathbf{F}_{\rm dist} \mathbf{J}_{\rm DE}$ for the variables $\boldsymbol{\theta}_{\rm DE} = (w_0, w_a, \Omega_\Lambda, \Omega_{\rm k}, \omega_{\rm m}, \omega_{\rm b})$ with the Jacobian $\mathbf{J}_{\rm DE}$ given by

$$\mathbf{J}_{\rm DE} = \frac{\partial \boldsymbol{\theta}_{\rm dist}}{\partial \boldsymbol{\theta}_{\rm DE}}, \tag{7.38}$$

where Eq. (7.2) can be used to compute derivatives of $\ln H$ and $\ln D_A$. As seen in Fig. 7.6, the redshift range covered by CHIME encompasses regions where the derivatives of $\ln H$ and $\ln D_A$ with respect to w_0 and w_a are large, allowing for tight constraints on w_0 and w_a.

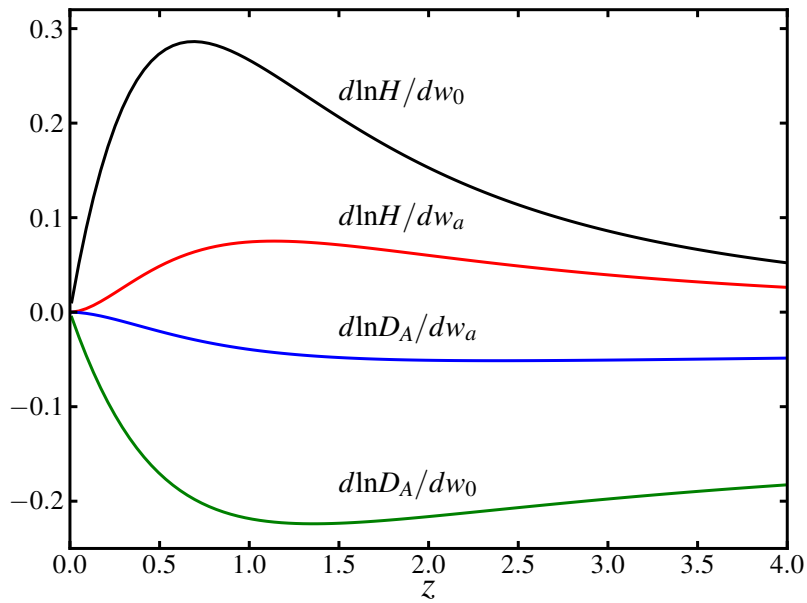


Figure 7.6: Derivatives of $\ln H$ and $\ln D_A$ with respect to w_0 and w_a, which appear in $\mathbf{J}_{\rm DE}$, as a function of redshift.

The final steps in forecasting our constraint contours in the $w_0 - w_a$ plane is to add any relevant priors to $\mathbf{F}_{\rm DE}$ and then marginalize over all other variables. The 95% CL contours in the $w_0 - w_a$ plane can be seen in Fig. 7.7, which shows the forecasts for CHIME combined with Planck, Stage II, and BOSS data. Stage II priors, which represent the anticipated constraints that will be available after currently running experiments are complete [137], are comprise of cluster, supernovae, and weak lensing surveys and were estimated using the DETFast [57] software. The BOSS survey aims to measure BAO over the redshift range $0.15 < z < 0.7$ and anticipates a precision of 1.0% and 1.8% on the measured values of D_A and H, respectively, at $z = 0.35$ and of 1.0% and 1.7% at $z = 0.6$, with a correlation coefficient of 0.4 between D_A and H at both redshifts [155]. As seen in Fig. 7.7, adding the anticipated constraints from CHIME to Planck+Stage II drastically improves the dark energy equation of state constraints, which may be further improved upon by adding the lower redshift BAO measurements from BOSS.

[57] http://www.physics.ucdavis.edu/DETFast/

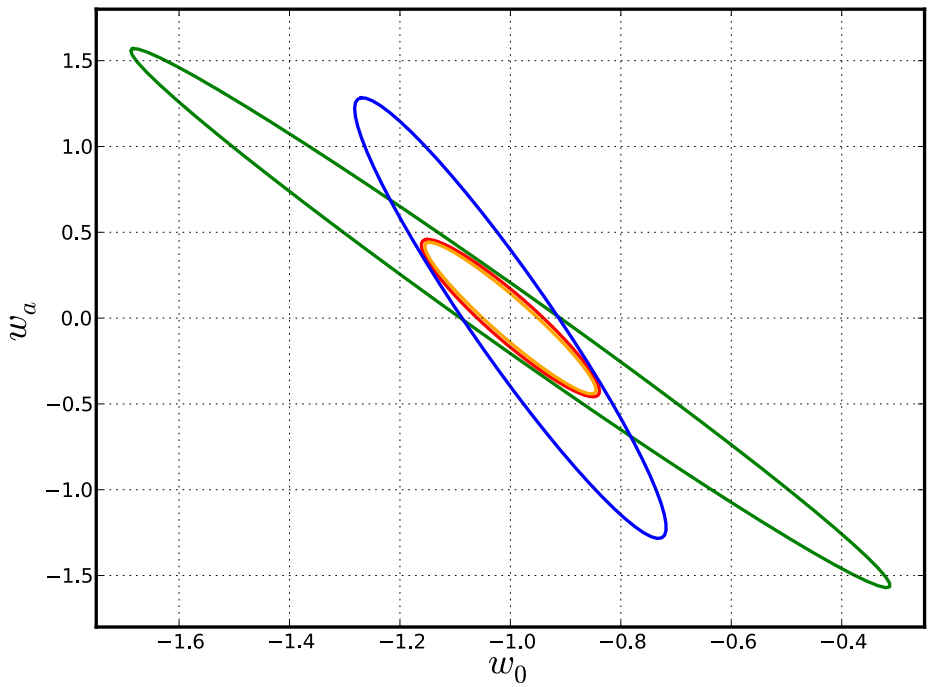


Figure 7.7: Forecasted constraints in the $w_0 - w_a$ plane for CHIME+Planck (green, FOM = 42.2), CHIME+Planck+Stage II (red, FOM = 217.0), CHIME+Planck+Stage II+BOSS (orange, FOM = 270.9), and Planck+Stage II (blue, FOM = 53.3). Each contour represents the 95% CL contours of each forecast.

The figure of merit (FOM) for the $w_0 - w_a$ contour is defined as being proportional to the inverse of the area of the contour, with a larger FOM indicating a more constraining measurement. We use the normalization of the FOM given in Ref. [137], which can be easily computed from the marginalized Fisher matrix $\mathbf{F}^{\mathrm{M}}_{\mathrm{DE}}$ by

$$\mathrm{FOM} = \sqrt{\det(\mathbf{F}^{\mathrm{M}}_{\mathrm{DE}})}. \tag{7.39}$$

For CHIME+Planck+Stage II the FOM is approximately 217.0 and increases to 270.9 when BOSS is added as a prior. The constraints from CHIME are competitive with previous forecasts for Planck+Stage II +Stage III,[58] which estimate a FOM of 279.5 [145]. Fig. 7.8 shows the relative improvement of the figure of merit over the fiducial value $\mathrm{FOM}_0 = 85.8$

[58] As defined in Ref. [137], Stage III refers to near-future, medium cost experiments.

from the forecast for Planck+Stage II+BOSS. Each curve represent a possible survey with a different lowest redshift $z_{\rm min}$ probed as a function of the survey's maximum redshift $z_{\rm max}$. The chosen CHIME band has been indicated. Fig. 7.8 shows that there would be little constraining power gained by extending the CHIME band to cover lower redshifts, as these redshifts have already been or will soon be probed by existing surveys and that there is limited benefit extending the band to higher redshifts where the changes in D_A and H with w_0, w_a are smaller than at lower redshifts, as indicated in Fig. 7.6.

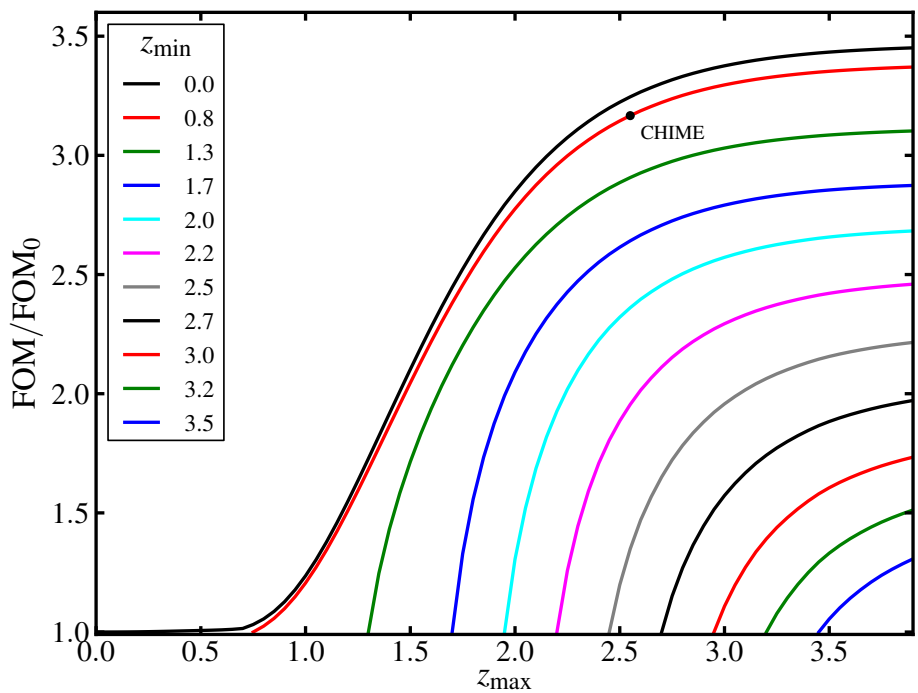


Figure 7.8: Relative improvement of figure of merit FOM with CHIME over fiducial value FOM$_0$ as a function of redshift coverage. Each curve represent a survey with a different lowest redshift measured $z_{\rm min}$ as a function of maximum redshift $z_{\rm max}$. The fiducial figure of merit FOM$_0$ = 85.8 is taken as the forecast for Planck+Stage II+BOSS. The black point denotes the CHIME band.

We can now ask at which redshift our experiment best constrains $w_{\rm DE}(z)$.

Longer term projects such as the SKA are included in Stage IV.

To this end, instead of expanding $w_{\rm DE}$ in a Taylor series about $a = 1$ as done in Eq. (7.1), we expand around the point $a = a_p$, so that

$$w_{\rm DE}(a) = w_p + w_a(a_p - a), \tag{7.40}$$

where $w_p = w_0 + w_a(1 - a_p)$. a_p is defined to be the scale factor when the uncertainty in $w_{\rm DE}$ is minimized, referred to as the *pivot point*. In this case, the pivot point coincides with the point when the uncertainty in w_p is minimized. By use of the Jacobian

$$\mathbf{J}_{\rm pivot} = \begin{pmatrix} 1 & a_p - 1 \\ 0 & 1 \end{pmatrix}, \tag{7.41}$$

to change from the parameters (w_0, w_a) to (w_p, w_a), the uncertainty in w_p is found to be

$$\sigma^2_{w_p} = \sigma^2_{w_0} + (1 - a_p)^2\sigma^2_{w_a} + 2(1 - a_p)\rho_{w_0,w_a}\sigma_{w_0}\sigma_{w_a}, \tag{7.42}$$

where ρ_{w_0,w_a} is the correlation coefficient between w_0 and w_a, so that the minimum in $\sigma^2_{w_p}$ occurs at

$$a_p = 1 + \frac{\rho_{w_0,w_a}\sigma_{w_0}}{\sigma_{w_a}}, \tag{7.43}$$

and $\sigma^2_{w_p}$ then simplifies to

$$\sigma^2_{w_p} = \sigma^2_{w_0}(1 - \rho^2_{w_0,w_a}). \tag{7.44}$$

It is easily verified that w_p and w_a are uncorrelated. In addition, since $\det(\mathbf{J}^{\rm T}_{\rm pivot}\mathbf{F}^{\rm M}_{\rm DE}\mathbf{J}_{\rm pivot}) = \det(\mathbf{J}^{\rm T}_{\rm pivot})\det(\mathbf{F}^{\rm M}_{\rm DE})\det(\mathbf{J}_{\rm pivot}) = \det(\mathbf{F}^{\rm M}_{\rm DE})$, the FOM for the $w_p - w_a$ contour will be the same as the $w_0 - w_p$ contour. The constraints on $w_{\rm DE}$ for CHIME+Planck+Stage II as a function of redshift can be seen in Fig. 7.9, where the pivot point occurs at redshift $z_p = 0.49$, near where changes in $w_{\rm DE}$ produce the largest changes in H and D_A.

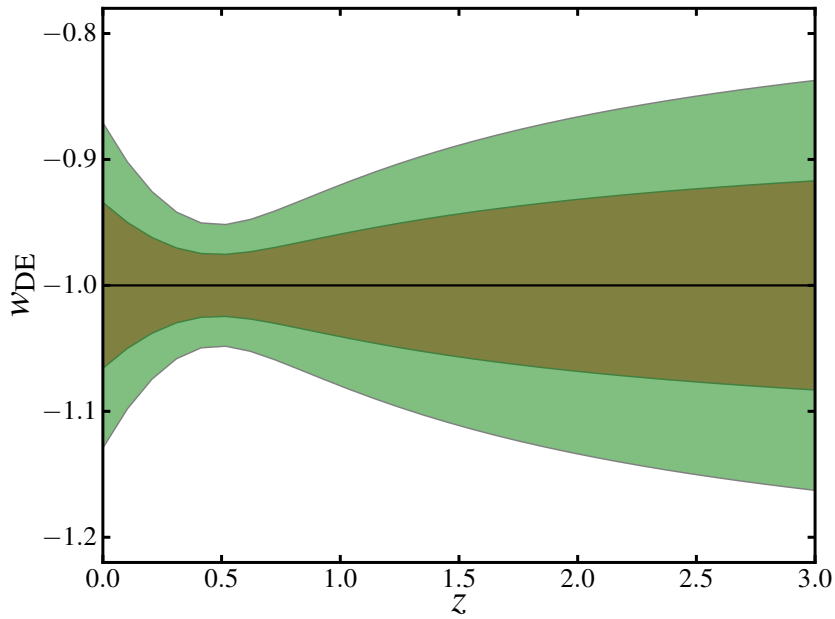


Figure 7.9: Constraints on $w_{\rm DE}$ from the forecast for CHIME+Planck+Stage II. Filled regions denote the $1-\sigma$ and 95% CL bounds. The constraints are most stringent at the pivot point occurring at redshift $z_p = 0.49$.

7.8 Conclusions

The forecasting methods and models presented in this chapter provide a rapid and easily implementable method for generating dark energy forecasts for 21-cm intensity mapping experiments such as CHIME. These forecasts allow one to optimize the design of 21-cm intensity mapping experiments for the detection of the BAO at different redshifts as well as how to best constrain the dark energy equation of state in combination with existing and future experiments. These methods may be extended to allow for more detailed forecasts which include effects due to foreground subtraction [141, 142]. We have shown that CHIME will add competitive and complimentary bounds on the dark energy equation of state compared to Stage II and III experiments.

Chapter 8

Redundant Baseline Calibration

8.1 Introduction

Radio interferometric telescopes with a large number of elements will form an important category of radio telescopes in the near future, with many experiments currently being built or planned for in the next decade [156, 157, 158, 159, 160, 161]. These experiments will typically have to remove bright foregrounds. For example, 21-cm intensity mapping experiments will be required to remove foregrounds that are four orders of magnitude larger than the desired 21-cm signal. These requirements will introduce many operational challenges for these projects, such as the precise calibration of system gains, characterization of the beam shapes, and removal of radio frequency interference, amongst others.

There exist many calibration techniques already employed by radio interferometers, for example the use of noise injection from a known noise source or the use of calibrator sources in the sky. While these methods have proven successful in the past, their application to 21-cm mapping experiments will introduce many new uncertainties, especially when using calibrator sources, as many of these experiments will operate at frequencies where few precise maps of the sky exist. Due to the need for great precision in the calibration of these telescopes, we expect that a variety of new and existing calibration techniques will be used. In this vein, in this chapter we examine new techniques for the calibration of the direction-independent gains of interferometric arrays that contain a large number of redundant baselines.

Many array designs for new telescopes contain many redundant or nearly redundant baselines. Redundant baseline calibration exploits this redundancy to extract both the antenna gains and calibrated sky visibilities with minimal a priori information about the sky. As such, redundant baseline calibration may be a valuable calibration technique for such experiments.

Redundant baseline calibration has been tested on the Westerbork Syn-

thesis Radio Telescope [162] as well as LOFAR [163] and several different algorithms have been developed [164, 165, 166, 167, 168]. The common factor among algorithms is the assumption that all primary beams in the array are identical, so that (up to noise levels and excluding the effect of the different gains of the system) visibilities formed from perfectly redundant baselines will be identical, excluding other unintended signals such as crosstalk. In this model, the differences between measured visibilities with the same baseline will be due to variation of the gains used to form these visibilities. Redundant baseline algorithms essentially compare the measured visibilities that share a baseline to one another to achieve a nearly sky independent measurement of the gains.

An essential assumption of the basic implementation of the redundant baseline algorithm is that all primary beams are identical. In actuality, the beams will differ from one another to some degree. If these differences are significant, calibrations done with the redundant baseline algorithm will be poor. In this chapter, we examine the effects of introducing variations between primary beams in the array on redundant baseline calibrations. We develop a model of redundant baseline calibration for the gain amplitudes that accounts for the variation between beams which can result in improved calibrations.

This chapter is organized as follows: The model of the gains is introduced in Section 8.3. We will use different algorithms for the gain amplitude and phase calibrations, which are examined in Sections 8.4 and 8.5, respectively. The different redundant baselines calibration algorithms discussed in this chapter will be tested on simulated data, the generation of which is described in Section 8.4.4.

8.2 Calibration Requirements for CHIME

Although our redundant baseline calibration algorithm can be applied to any multi-element interferometric telescope that contains enough redundant baselines, we will be particularly interested in comparing the performance of the calibration algorithms that will be examined to the calibration requirements for CHIME. Here we will give a brief overview of these requirements before discussing calibration algorithms.

In Ref. [142], simulated calibration errors were propagated through the CHIME pipeline, the results of which show that if the gain amplitude is calibrated to an accuracy of a few percent and the phase calibrated to an accuracy of a few degrees, then the 21-cm power spectrum can be con-

structed to an accuracy of $\sim 10\%$ or better. At higher levels of calibration error, the extraction of the 21-cm power spectrum is significantly degraded, with $\sim 10\%$ gain calibration errors resulting in systematic errors dominating over statistical uncertainties. As such, we desire a gain amplitude calibration accuracy of at least a few percent and a phase gain calibration accuracy of a few degrees. Furthermore, by achieving a phase calibration accurate to a few degrees, we ensure that such phase errors are subdominant to those induced by a warping of the reflector on the scale of $\sim 1\,\mathrm{cm}$, the expected scale to which CHIME's reflector could be considered accurate to.

8.3 Gain Model

We model the response S_i of antenna i within an array of receivers as $S_i = g_i F_i + n_i$, where g_i is the (complex) gain of feed i and n_i is an additive noise term. Taking the time-averaged correlation between feeds produces the measured visibilities

$$V_{ij}^{\mathrm{meas}} = \langle S_i^* S_j \rangle = g_i^* g_j V_{ij} + n_{ij}, \tag{8.1}$$

where n_{ij} is the noise on the measured visibility and V_{ij} are the 'true' visibilities. Each feed is sensitive to a particular polarization. As we will be considering only visibilities formed by correlating feeds that measure the same polarization, we do not add any references in our notation to the particular polarization measured. For arrays that measure two polarization states, the calibration algorithms to be described may be done on each polarization separately. The true visibility can be represented as

$$V_{ij} = \frac{1}{\sqrt{\Omega_i \Omega_j}} \int d^2\hat{\mathbf{n}} A_i^*(\hat{\mathbf{n}}) A_j(\hat{\mathbf{n}}) e^{2\pi i \hat{\mathbf{n}}\cdot\mathbf{u}_{ij}} T(\hat{\mathbf{n}}), \tag{8.2}$$

where $A_i(\hat{\mathbf{n}})$ is the primary beam shape of feed i in the direction $\hat{\mathbf{n}}$ and $T(\hat{\mathbf{n}})$ is the sky intensity for the polarization of interest. $\mathbf{u}_{ij}$ is the feed separation between feeds i and j and $\Omega_i = \int d^2\hat{\mathbf{n}} |A_i(\hat{\mathbf{n}})|^2$ is the beam solid angle.

If all primary beams are identical, V_{ij} is dependent only on the separation $\mathbf{u}_{ij}$ and thus V_{ij} will be identical for pairs of feeds separated by the same distance. In this case, we use the notation of Ref. [164] to write V_{ij} as V_{i-j} to emphasize this point, although the subscript $_{i-j}$ should not be taken literally.

In the standard redundant baseline algorithm where all primary beams are identical, the measured visibilities V_{ij}^{meas} are used to give a simultaneous estimate of both the gains g_i and true visibilities V_{i-j}. For an array with N

feeds, we have $N(N-1)/2$ correlations (excluding autocorrelations) and N gains. If all baselines are unique, then we will have $N(N-1)/2$ true visibilities and solving for both gains and true visibilities is an underdetermined problem. On the other hand, if enough of the baselines in the array are the same, the problem will be overdetermined. In particular, a regular array is overdetermined for sizes larger than a few elements.

Removing the assumption that all beams are identical, we can write the beams in terms of a set of basis functions $\{A^\mu(\hat{\mathbf{n}})\}$, where μ labels each basis function in the set. By expanding beam i with coefficients a_i^μ as

$$A_i(\hat{\mathbf{n}}) = \sum_\mu a_i^\mu A^\mu(\hat{\mathbf{n}}), \tag{8.3}$$

the true visibilities V_{ij} are then given by

$$V_{ij} = \frac{1}{\sqrt{\Omega_i \Omega_j}} \sum_{\mu\nu} a_i^{\mu *} a_j^\nu V_{i-j}^{\mu\nu}, \tag{8.4}$$

where

$$V_{i-j}^{\mu\nu} = \int d^2\hat{\mathbf{n}} A^{\mu *}(\hat{\mathbf{n}}) A^\nu(\hat{\mathbf{n}}) e^{2\pi i \hat{\mathbf{n}}\cdot\mathbf{u}_{ij}} T(\hat{\mathbf{n}}). \tag{8.5}$$

In this case, although the visibilities V_{ij} are no longer identical for pairs of feeds with the same baseline, the quantities $V_{i-j}^{\mu\nu}$ are the same for these pairs.

If the beam can be well approximated by a small number of basis functions with known coefficients a_i^μ, we may employ a calibration algorithm very similar to the standard redundant baseline algorithm, except we must now solve for the parameters $V_{i-j}^{\mu\nu}$ instead of V_{i-j} in addition to the gains.

8.4 Amplitude Calibration

In this section we will examine redundant baseline algorithms useful for gain amplitude calibrations. We first review the so-called 'logarithm method' that assumes that all beams are identical and subsequently describe a novel extension to this method that can account for variation between beams.

8.4.1 The Logarithm Method

The logarithm method provides a straightforward way to solve for the model parameters, as well as for determining any degeneracies that are present between parameters.

We can linearize the calibration problem by taking the logarithm of Eq. (8.1), which will allow us to use well known linear least-squares solutions to estimate the model parameters. After taking the logarithm, we can separate the real and imaginary parts as

$$\ln|V_{ij}^{\mathrm{meas}}| = \ln|g_i| + \ln|g_j| + \ln|V_{ij}| + \mathrm{Re}(\eta_{\mathrm{ij}}), \tag{8.6a}$$

$$\arg(\mathrm{V}_{\mathrm{ij}}^{\mathrm{meas}}) = -\arg(\mathrm{g_i}) + \arg(\mathrm{g_j}) + \arg(\mathrm{V_{ij}}) + \mathrm{Im}(\eta_{\mathrm{ij}}) + 2\pi\mathrm{c}, \tag{8.6b}$$

where $\eta_{ij} = \ln(1 + n_{ij}/g_i^* g_j V_{ij})$ and c is an arbitrary integer. By taking the logarithm, we have turned the product in Eq. (8.1) into a sum and have separated the amplitudes and phases into two separate equations, although the amplitude and phase equations are weakly coupled through the noise term η_{ij}. We can then solve for the gain amplitudes and phases separately.

As noted in Ref. [164], while the logarithm method can accurately recovers the gain amplitudes, large errors may appear in its estimate of the phases. This problem arises from the freedom of c in Eq. (8.6b) to assume the value of any integer, which if a least-squares estimate is to be formed creates an ambiguity in the phases of the gains. As a result, while the logarithm method may be useful in calibrating the gain amplitudes, in its current formulation it is ill-suited for producing estimates of the phases and we will only examine its performance for amplitude calibrations for this method.

8.4.2 Identical Beams

In the case where all primary beams are identical, we have $V_{ij} = V_{i-j}$ so Eq. (8.6a) becomes

$$\ln|V_{ij}^{\mathrm{meas}}| = \ln|g_i| + \ln|g_j| + \ln|V_{i-j}| + \mathrm{Re}(\eta_{\mathrm{ij}}). \tag{8.7}$$

We can write Eq. (8.7) as the matrix equation

$$\mathbf{d} = \mathbf{M}\mathbf{x} + \boldsymbol{\eta}, \tag{8.8}$$

where the vector $\mathbf{d}$ holds the logarithms of the measured visibilities, $\mathbf{x}$ contains the logarithms of both the gains and true visibilities, and the noise terms are put into $\boldsymbol{\eta}$. The information regarding the array configuration is incorporated into the matrix $\mathbf{M}$. For example, for a regular one-dimensional array with feeds numbered sequentially down the array, the amplitude equa-

tion could be written as

$$\begin{pmatrix} \ln|V_{12}^{\text{meas}}| \\ \ln|V_{23}^{\text{meas}}| \\ \ln|V_{13}^{\text{meas}}| \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & & 1 & 0 & \\ 0 & 1 & 1 & \cdots & 1 & 0 & \cdots \\ 1 & 0 & 1 & & 0 & 1 & \\ & \vdots & & & \vdots & & \ddots \end{pmatrix} \begin{pmatrix} \ln|g_1| \\ \ln|g_2| \\ \ln|g_3| \\ \vdots \\ \ln|V_1| \\ \ln|V_2| \\ \vdots \end{pmatrix} + \begin{pmatrix} \text{Re}(\eta_{12}) \\ \text{Re}(\eta_{23}) \\ \text{Re}(\eta_{13}) \\ \vdots \end{pmatrix}, \tag{8.9}$$

where the subscript b on V_b here labels the baseline lengths in units of the smallest baseline b_{min}.

Once the matrix $\mathbf{M}$ is formed for our array configuration, we can estimate the gains and true visibility amplitudes contained in the vector $\mathbf{x}$ with the least-squares estimator

$$\hat{\mathbf{x}} = (\mathbf{M}^T\mathbf{N}^{-1}\mathbf{M})^{-1}\mathbf{M}^T\mathbf{N}^{-1}\mathbf{d}, \tag{8.10}$$

where $\mathbf{N} = \langle \boldsymbol{\eta}\boldsymbol{\eta}^T \rangle$ is the (logarithmic) noise covariance matrix and the covariance of the estimator $\hat{\mathbf{x}}$ is given by

$$\mathbf{C}_{\hat{\mathbf{x}}} = (\mathbf{M}^T\mathbf{N}^{-1}\mathbf{M})^{-1}. \tag{8.11}$$

Although redundant baseline calibration is nearly independent of the sky, degeneracies in the model require a few additional constraints to be added in order to fully solve for the system. In particular, the calibration is not sensitive to the overall absolute gain of the array as the transformations $g_i \to Cg_i$ and $V_{ij} \to V_{ij}/C^2$, where C is a real constant, leave the measured visibilities unchanged. To allow us to solve for $\hat{\mathbf{x}}$ via Eq. (8.10), one may supplement the series of linear equations by a 'gauge-fixing' equations, such as

$$\sum_i \ln|g_i| = 0, \tag{8.12}$$

or by adding a condition to fix the value of a particular gain or calibrated visibility. In some situations only the relative gains are desired, so adjusting the overall absolute gain of the calibration is unimportant. If an absolute calibration is required, then the post-calibrated gains may be adjusted via the above transformations to fit additional information.

8.4.3 Nonidentical Beams

We now adapt the logarithm method to accommodate beams that vary from one another, where the departure of the beams from each other will be treated as a perturbation. As a shorthand, in later sections we will refer to this method as the *extended* algorithm and the algorithm that models each beam (as identical as described in the previous section) as the *basic* algorithm.

We characterize the beams by two real-valued functions $A^0(\hat{\mathbf{n}})$ and $A^1(\hat{\mathbf{n}})$ and will set $a_i^0 = 1$, as well as relabel the coefficients $\delta_i \equiv a_i^1$, which are taken to be real. It will be assumed that the beam profiles (and thus $\{\delta_i\}$ and $\{\Omega_i\}$) are known to some degree of accuracy. With this, Eq. (8.6a) becomes

$$\ln|V_{ij}^{\text{meas}}\sqrt{\Omega_i\Omega_j}| = \ln|g_i| + \ln|g_j| + \ln|V_{i-j}^{00}| + \ln|1+\epsilon_{ij}| + \text{Re}(\eta_{\text{ij}}), \quad (8.13)$$

where

$$\epsilon_{ij} = (\delta_i + \delta_j)\frac{V_{i-j}^{01}}{V_{i-j}^{00}} + \delta_i\delta_j\frac{V_{i-j}^{11}}{V_{i-j}^{00}}, \quad (8.14)$$

and note that we have $V_{i-j}^{01} = V_{i-j}^{10}$. We will assume that the deviation of each primary beam from $A_i(\hat{\mathbf{n}}) = A^0(\hat{\mathbf{n}})$ is small and that $|\epsilon_{ij}| \ll 1$. Approximating the second last term in Eq. (8.13) by a Taylor series to first order in ϵ_{ij} yields the linear equation

$$\begin{aligned}\ln|V_{ij}^{\text{meas}}\sqrt{\Omega_i\Omega_j}| \approx \ln|g_i| + \ln|g_j| + \ln|V_{i-j}^{00}| \\ + (\delta_i+\delta_j)\Gamma_{i-j}^{(0)} + \delta_i\delta_j\Gamma_{i-j}^{(1)} + \text{Re}(\eta_{\text{ij}}),\end{aligned} \quad (8.15)$$

where

$$\Gamma_{i-j}^{(0)} = \text{Re}\left(\frac{\text{V}_{\text{i-j}}^{01}}{\text{V}_{\text{i-j}}^{00}}\right), \quad (8.16a)$$

$$\Gamma_{i-j}^{(1)} = \text{Re}\left(\frac{\text{V}_{\text{i-j}}^{11}}{\text{V}_{\text{i-j}}^{00}}\right) \quad (8.16b)$$

are unknown sky parameters. Note that we can write Ω_i as

$$\Omega_i = \Omega^{00} + 2\delta_i\Omega^{01} + \delta_i^2\Omega^{11}, \quad (8.17)$$

where

$$\Omega^{\mu\nu} = \int d^2\hat{\mathbf{n}}A^\mu(\hat{\mathbf{n}})A^\nu(\hat{\mathbf{n}}). \quad (8.18)$$

With a preexisting estimate of the δs, we can write Eq. (8.15) as a matrix equation analogous to Eq. (8.8), which we write as

$$\tilde{\mathbf{d}} = \tilde{\mathbf{M}}\tilde{\mathbf{x}} + \boldsymbol{\eta}. \tag{8.19}$$

As in the basic algorithm, the vector $\tilde{\mathbf{x}}$ holds all model parameters, but now also includes the parameters $\Gamma^{(0)}_{i-j}$ and $\Gamma^{(1)}_{i-j}$ and may be packed as $\tilde{\mathbf{x}} = (\{\ln|g_i|\}, \{\ln|V^{00}_{i-j}|\}, \{\Gamma^{(0)}_{i-j}\}, \{\Gamma^{(1)}_{i-j}\})$. The vector $\tilde{d}_{(ij)} = \ln|V^{\text{meas}}_{ij}\sqrt{\Omega_i\Omega_j}|$ still holds the logarithms of the measured visibilities, but now has factors of $\sqrt{\Omega_i}$ added. $\tilde{\mathbf{M}}$ is constructed in a similar manner to $\mathbf{M}$, but now includes elements dependent on the δs.

As we have reduced our model to a set of linear equations, we can employ the same least-squares solution in Eq. (8.10) that was used to solve for the model parameters in the basic algorithm. However, as will be discussed in the following section, there exists an additional degeneracy between the model parameters that must be fixed in some way before the least-squares solution may be applied.

Lastly, we note that when using the basic algorithm in a situation with nonidentical primary beams, the vector $\mathbf{d}$ used to solve Eq. (8.10) can be replaced by the vector $\tilde{\mathbf{d}}$. This accounts for the variability of the beam solid angle between feeds while not changing the basic algorithm as described in Section 8.4.2, which will be taken into account when showing calibration results in Section 8.4.4.

Fixing the Degeneracies

A number of additional degeneracies are introduced by having expanded the number of model parameters, which may be found by calculating the null space of $\tilde{\mathbf{M}}$. However, not all of these degeneracies affect the gains and as such will not be of concern here. For example, for a regular linear array, there are three additional distinct degeneracies that leave the gains unaffected. On the other hand, the degeneracies that do affect the gains are

$$\ln|V^{00}_{i-j}| \to \ln|V^{00}_{i-j}| + 1 \quad \text{and} \quad \ln|\mathrm{g_i}| \to \ln|\mathrm{g_i}| - \frac{1}{2}, \tag{8.20a}$$

$$\Gamma^{(0)}_{i-j} \to \Gamma^{(0)}_{i-j} + 1 \quad \text{and} \quad \ln|\mathrm{g_i}| \to \ln|\mathrm{g_i}| - \delta_\mathrm{i}. \tag{8.20b}$$

The degeneracy in Eq. (8.20a) is the perturbed version of the overall amplitude degeneracy described in Section 8.4.2, which may be fixed in the same manner as done in the identical beams model. On the other hand, as the

degeneracy of Eq. (8.20b) is not present in the identical beams model, we will require an additional 'gauge-fixing' condition.

We can fix all degeneracies (both those that do and do not affect the gains) by using the matrix $\tilde{\mathbf{M}}_{\mathrm{f}}$ in Eq. (8.10) instead of $\tilde{\mathbf{M}}$, which is constructed by appending the null space $\tilde{\mathbf{M}}_{\mathrm{null}} = \mathrm{Null}(\tilde{\mathbf{M}})$ to $\tilde{\mathbf{M}}$

$$\tilde{\mathbf{M}}_{\mathrm{f}} = \begin{pmatrix} \tilde{\mathbf{M}} \\ \tilde{\mathbf{M}}_{\mathrm{null}} \end{pmatrix}, \tag{8.21}$$

The vector $\tilde{\mathbf{d}}$ must also be appended by a vector of length equal to the total number of degeneracies. For this part of the calculation, we choose to set these additional values to zero. For the degeneracies that do not affect the gains, setting those values in $\tilde{\mathbf{d}}$ to zero is inconsequential. In both the identical and nonidentical beams algorithms we will enforce the overall gain amplitude condition given in Eq. (8.12) and will only be interested in the relative gain calibration. The only remain degeneracy is that of Eq. (8.20b), which we will use to adjust our initial estimate of $\tilde{\mathbf{x}}$, found by use of Eq. (8.21), to fit to prior information (the details of which will be discussed in the following section).

8.4.4 Simulation

To test of our calibration algorithms, we simulate the response of a 12 feed regular linear array aligned in the North-South direction with beams pointed at zenith. The primary beam of each feed is modelled as being nearly Gaussian, with a narrow width in the East-West direction and wide field of view in the NS direction, to mimic the response of feeds placed along the focal line of a cylindrical reflector oriented with its axis along the NS direction. The zeroth order beam A^0 is taken as a two dimensional Gaussian with widths σ_u and σ_v in the EW and NS directions, respectively. Our beam basis functions are chosen from the Hermite functions, with zeroth order function of which is a Gaussian. We choose the perturbing first order beam A^1 to be the second Hermite function.[59] Specifically, we have $A^0(x^2) \propto \exp(-x^2/2)$ and $A^1(x^2) = (2x^2 - 1)A^0(x^2)/\sqrt{2}$, where $x^2 = (\hat{\mathbf{n}} \cdot \hat{\mathbf{u}})^2/\sigma_u^2 + (\hat{\mathbf{n}} \cdot \hat{\mathbf{v}})^2/\sigma_v^2$, with $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ pointing in the East and North directions, respectively. With

[59] We have chosen A^1 as the second Hermite function as there has been a significant variability in the beam widths of the CHIME pathfinder observed. Choosing the first Hermite function instead would correspond to the feeds having different pointings, which has been observed as well in the CHIME pathfinder. As the conclusions are the same in either case, we only examine the single case where the second Herminte function is used.

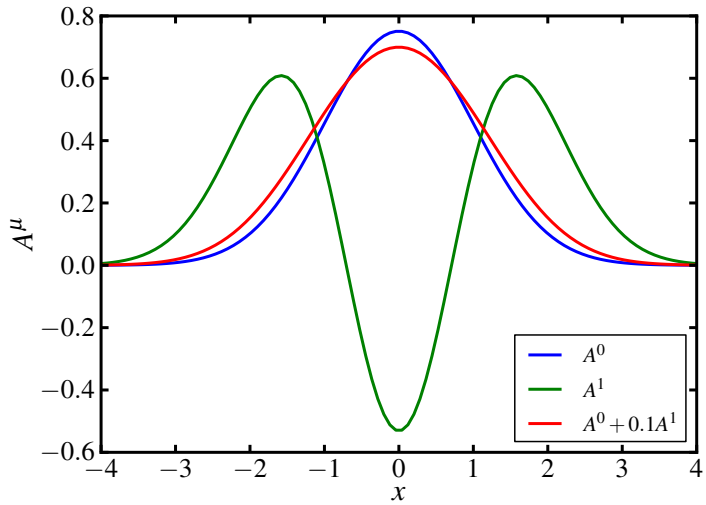


Figure 8.1: Beam basis functions chosen as the zeroth and second order Hermite functions. A linear combination of the two beam basis function is displayed, illustrating that for small values of δ_i, the chosen basis functions perturb the width of a Gaussian beam.

this basis, having a nonzero value of δ_i perturbs the beam width, as seen in Fig. 8.1.

The visibilities used for the test calibrations in this section where generated using the 408 MHz Haslam map [169].[60] These visibilities were generated assuming a feed separation of 30 cm, where the latitude of the telescope was chosen to be at 45°. The time step used to generate our test set of visibilities was taken with a transiting right ascension of 0°.

We set the EW beam width as $\sigma_u = \lambda/W$, where λ is the wavelength corresponding to the frequency of 408 MHz and W is a length scale that in the present context can be thought of as approximating the width of our cylindrical reflector, chosen to be $W = 20\,\mathrm{m}$. The NS beam width is set as $\sigma_v = 10\sigma_u$ so that each feed has a large field of view in that direction.

The noise on the visibility V_{ij}^{meas} is constructed with a variance of $(\sigma_n^2)_{ij} = |g_i||g_j|T_{\mathrm{sys}}^2/t_{\mathrm{int}}\Delta\nu$ and is uncorrelated between different visibilities. The noise covariance matrix $\mathbf{N}$ for the logarithmic noise variables η_{ij} can be

[60] The map used is available at http://lambda.gsfc.nasa.gov/product/foreground/haslam_408.cfm

approximated by

$$N_{(ij),(kl)} = \frac{\sigma_n^2}{|V_{ij}^{\text{meas}}|^2}\delta_{(ij),(kl)} \tag{8.22}$$

where (ij) specifies the correlation index number between feeds i and j and $\delta_{(ij),(kl)}$ is the Kronecker delta function. Although in our model σ_n depends on the gain amplitudes, as we do not assume detailed a priori information about the gain amplitudes, we do not vary our estimate of σ_n^2 for each visibility when forming $\mathbf{N}$ for use in Eq. (8.10). Instead, an expected average gain amplitude level between all feeds may be incorporated into σ_n^2 so that it remains constant for all visibilities, although including variations of σ_n^2 between different visibilities may easily be incorporated if such information is available. For the generation of the noise we use a system temperature of $T_{\text{sys}} = 200\,\text{K}$, a bandwidth of $\Delta\nu = 1\,\text{MHz}$, and integration time of $t_{\text{int}} = 10\,\text{s}$.[61]

We select the true values of the gain amplitudes randomly from a uniform distribution between -0.5 and 1.5 and assign a random phase, and pick the δs from a uniform random distribution from $-\delta_{\text{max}}$ to δ_{max}, the particular chosen values of which can be seen in Fig. 8.2. For δ_{max} we use a fiducial value of $\delta_{\text{max,fid}} = 0.1$.

8.4.5 Amplitude Calibration Results

As described in Section 8.4.3, we require at least one additional piece of information to properly fix the extra relevant degeneracy in Eq. (8.20b) arising in the nonidentical beam model. To do this, we will provide a prior on $\Gamma_{b_{\text{min}}}^{(0)}$ for the smallest baseline b_{min}. This choice of priors may be useful when accurate prior sky information is available for only large scales probed by only the smallest baseline of an array, while small scale information that contributes significantly to the signal measured by larger baselines is poorly constrained by prior information.

To begin with, we assume that both the δ parameters and the prior on $\Gamma_{b_{\text{min}}}^{(0)}$ are known to arbitrary accuracy, both of which will later be relaxed. The calibration results for both basic and extended redundant baseline algorithms can be seen in Fig. 8.3, which shows the relative bias and standard deviation for each feed in the array over a set of visibilities taken with 100 different realizations of the noise.[62] The set of visibilities used for this cal-

[61] The value of $T_{\text{sys}} = 200\,\text{K}$ was chosen to stress test the algorithm and as such is a bit higher than the nominal value of $T_{\text{sys}} = 50\,\text{K}$ for CHIME.

[62] The same set of 100 noise realizations are used for all example calibrations throughout the chapter.

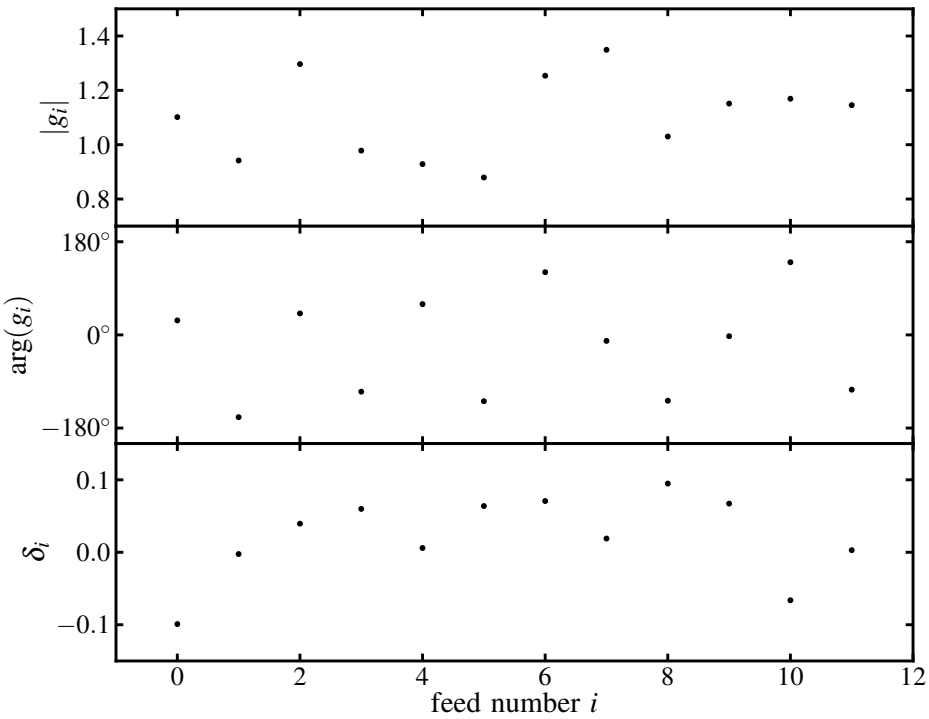


Figure 8.2: Fiducial simulated values of the gains g_i and the beam perturbation parameters δ_i with $\delta_{\mathrm{max,fid}} = 0.1$.

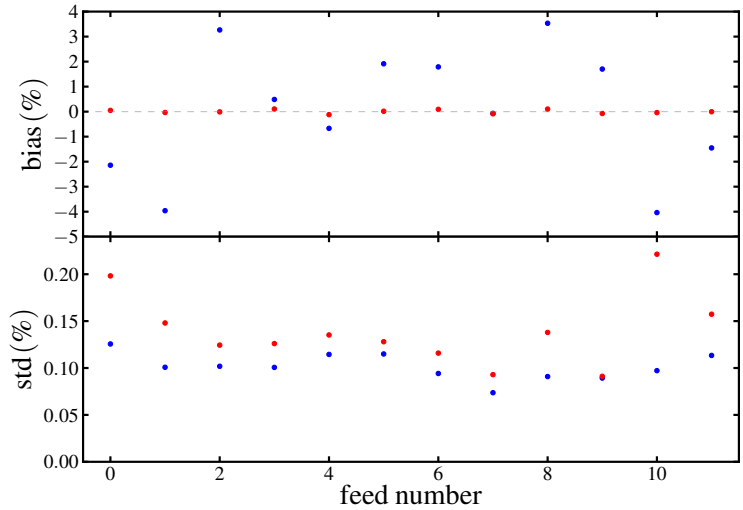


Figure 8.3: Relative calibrated gain amplitude $|g_i|$ bias and standard deviation over a set of 100 noise realizations for the basic (blue, Section 8.4.2) and extended (red, Section 8.4.3) redundant baseline calibration methods. Simulated data was generated with the fiducial values for the gains and δs as seen in Fig. 8.2. The extended calibration algorithm was performed with the δs known to arbitrary accuracy and by fixing $\Gamma^{(0)}_{b_{\min}}$ for the smallest baseline to a prior that is known without error.

ibration was generated using the fiducial values of the δs, as depicted in Fig. 8.2. It is clear from Fig. 8.3 that in this situation, the calibrated gain amplitudes produced by the basic redundant baseline algorithm are highly biased, while the extended algorithm results in a small bias. In general, the resulting biases in the basic algorithm will be correlated with the beam perturbation parameters δ_i, which for this particular calibration has a correlation coefficient of -0.86. As the extended algorithm uses the same amount of information to fit for more parameters, we expect that its gain calibrations will have a larger variance, as can be seen in the lower panel of Fig. 8.3.

The improvement of the extended algorithm over the basic algorithm depends on the size of the beam perturbations. Fig. 8.4 shows the calibration results of the basic and extended algorithms as a function of the maximum beam perturbation parameter $\delta_{\max}$, where the beam perturbation variables are given by $\delta_i = (\delta_{\text{fid}})_i \times \delta_{\max}/\delta_{\max,\text{fid}}$. The absolute value of the bias averaged over all feeds (we denoted the average over all feeds by $\langle \ldots \rangle_{\text{f}}$) is plotted with errors bars that represent the square root of $\langle \sigma_g^2 \rangle_{\text{f}}$, which is the variance

σ_g^2 of the calibrated gain amplitudes over the set of noise realizations averaged over all feeds (note that the error bars do not represent the variance of the estimate of the bias. Error bars in all further plots in this chapter should be interpreted in this way). Since in our example the beam perturbations act to perturb the width of the beams, the value of each δ_i can be translated into a perturbation of the full width at half maximum (FWHM) of the beam. The relative difference between the largest and average beam width, $\mathrm{FWHM}_{\mathrm{max}}$ and FWHM_0, respectively, is displayed on the top horizontal axis of Fig. 8.4.[63] If the perturbations are negligible, the extended algorithm will produce similar averages for the calibrated gain amplitudes, but with a higher variance. With large perturbations, the truncated terms in the Taylor series in Eq. (8.15) become more important and the bias in the extended algorithm grows, although the calibration may still show significant improvement over the basic algorithm. Intermediate values of the δs show a significant improvement in the calibrations of the extended over the basic algorithm, with little bias appearing in the extended calibrations.

Although using the logarithm method as described by Eq. (8.15) can yield accurate estimates of the gain amplitudes, as with the basic logarithmic redundant baseline algorithm, if the noise n_{ij} is normally distributed, the logarithmic noise term η_{ij} will not be, leading to a statistical bias in the gain amplitude calibration. However, with noise levels attainable with modern low noise amplifiers and integration times on the order of a second or greater, the distributions of η_{ij} will be very close to normal, resulting in only a small statistical bias.

We now relax the assumption that the prior on $\Gamma^{(0)}_{b_{\mathrm{min}}}$ is known exactly and run the extended calibration algorithm with an error added to this prior, the results of which can be seen in Fig. 8.5. Although obscuring our knowledge of the prior degrades the calibration, only after adding a large error to the prior does the basic algorithm outperform the extended algorithm.

As a final test of the extended algorithm, the assumption that the beam perturbation parameters are known exactly is relaxed by adding a random error to our knowledge of each δ_i, which are used to construct the matrix $\tilde{\mathbf{M}}$. To generate this error, we draw values for the fractional error of each beam width from a normal distribution centred on zero with variance $\sigma^2_{\ln\delta}$. The resulting calibrations can be seen in Fig. 8.6 for the extended algorithm with no errors on the prior for $\Gamma^{(0)}_{b_{\mathrm{min}}}$ and as well as for 50% and 100% errors on this prior. The top horizontal axis in Fig. 8.6 displays the RMS error

[63] Based on preliminary analysis, the beam widths of the CHIME pathfinder can vary by $\sim 15\%$, which corresponds to a value of δ slightly less than $\delta_{\mathrm{max,fid}} = 0.1$.

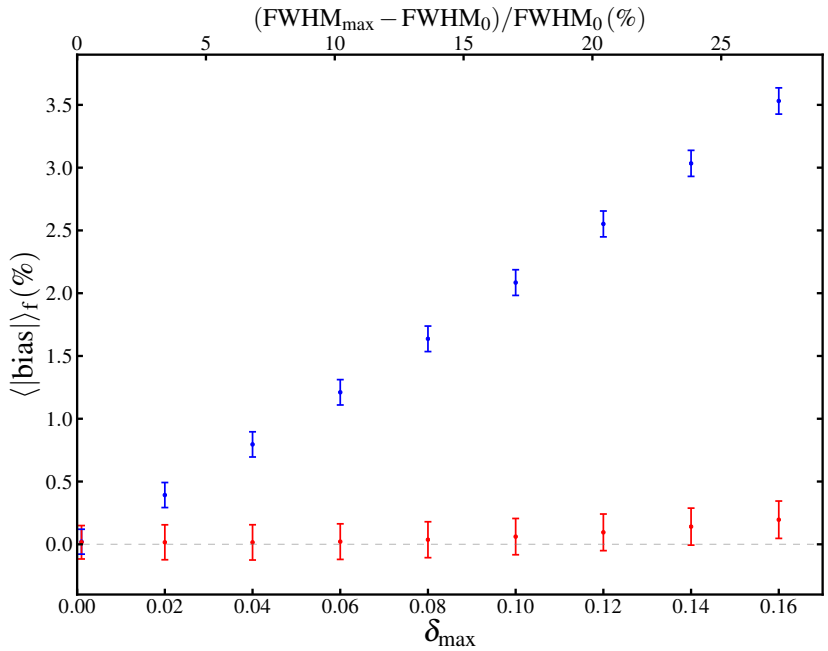


Figure 8.4: Gain amplitude calibration for the basic (blue) and extended (red) algorithms as a function of the maximum beam perturbation parameter $\delta_{\rm max}$. At each value of $\delta_{\rm max}$ the beam perturbation parameters are scaled as $\delta_i = (\delta_{\rm fid})_i \times \delta_{\rm max}/\delta_{\rm max,fid}$. Each point is the absolute value of the relative bias averaged over all of the feeds. The error bars represent the square root of the variance in the calibrated gain amplitude σ_g^2 averaged over all feeds $\sqrt{\langle \sigma_g^2 \rangle_{\rm f}}$.

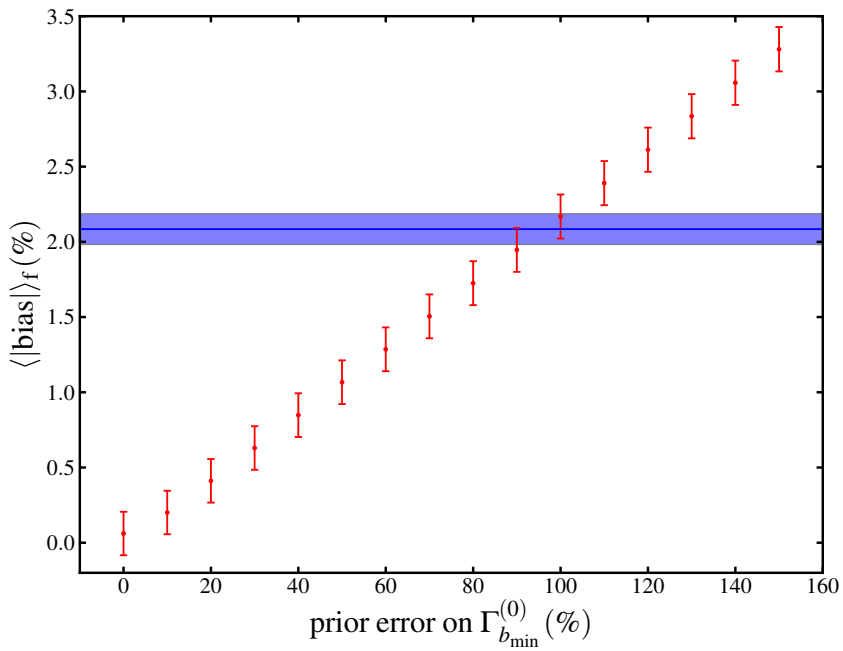


Figure 8.5: Amplitude calibration as a function of error on the prior of $\Gamma^{(0)}_{b_{\text{min}}}$ used to fix the extra degeneracy in the extended algorithm. The extended algorithm is shown in red and the basic algorithm, which is independent of the prior on $\Gamma^{(0)}_{b_{\text{min}}}$, is shown as the blue band. Fiducial values of the δs are used and are assumed to be known perfectly. Error bars are to be interpreted as in Fig. 8.4.

on the beam FWHM over all the feeds in the array. With the beam widths known to a few percent and a reasonably accurate prior on $\Gamma^{(0)}_{b_{\min}}$, we can see that the mean bias on the gain amplitude calibration performed with the extended algorithm remains near the percent level or better.

8.5 Phase Calibration

In the previous section, we examined the performance of the logarithm method (and its extension to accommodate for varying primary beams) for the calibration of the gain amplitudes of our array. As mentioned in Section 8.4.1, the logarithm method has difficultly in estimating the phases of the gains and thus we look for another method better suited for the phase calibration using redundant baselines. In this section we examine such a calibration algorithm, although no attempt is made to accommodate for variations between primary beams, which we leave for future work.

8.5.1 The Eigenvector Method

The eigenvector method is an iterative calibration technique close in vein to the self-calibration algorithm. We begin by seeking an estimate of the gain matrix $\mathbf{G} = |g\rangle\langle g|$. To form our estimate, we divide V_{ij}^{meas} with an estimate of the true visibilities (we will soon see that this estimate can be made arbitrary), yielding the matrix $\hat{G}_{ij} = V_{ij}^{\text{meas}}/V_{ij}$. We would now like to find the gain vector $|g\rangle$ that minimizes the chi-squared

$$\chi^2 = \sum_{ijkl} (\hat{G}_{ij} - G_{ij})^* C^{-1}_{(ij),(kl)} (\hat{G}_{kl} - G_{kl}), \tag{8.23}$$

where $C_{(ij),(kl)}$ is the covariance matrix for $\hat{\mathbf{G}}$. We assume that the covariance matrix is uncorrelated between different baselines and weights all baselines with the same variance σ^2. Under these assumptions, Eq. (8.23) simplifies to

$$\chi^2 = \frac{1}{\sigma^2} \sum_{ij} |\hat{G}_{ij} - G_{ij}|^2. \tag{8.24}$$

Before we attempt to minimize the above chi-squared, we note that since $\mathbf{G}|g\rangle = \langle g|g\rangle|g\rangle$, then $|g\rangle$ is an eigenvector of $\mathbf{G}$ with eigenvalue $\langle g|g\rangle$. The matrix $\hat{\mathbf{G}}$ is Hermitian by construction and therefore has an eigendecomposition $\mathbf{G} = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|$, where λ_i is the ith eigenvalue of $\hat{\mathbf{G}}$ corresponding to eigenvector $|\lambda_i\rangle$. As $\mathbf{G}$ is the outer product of a vector with

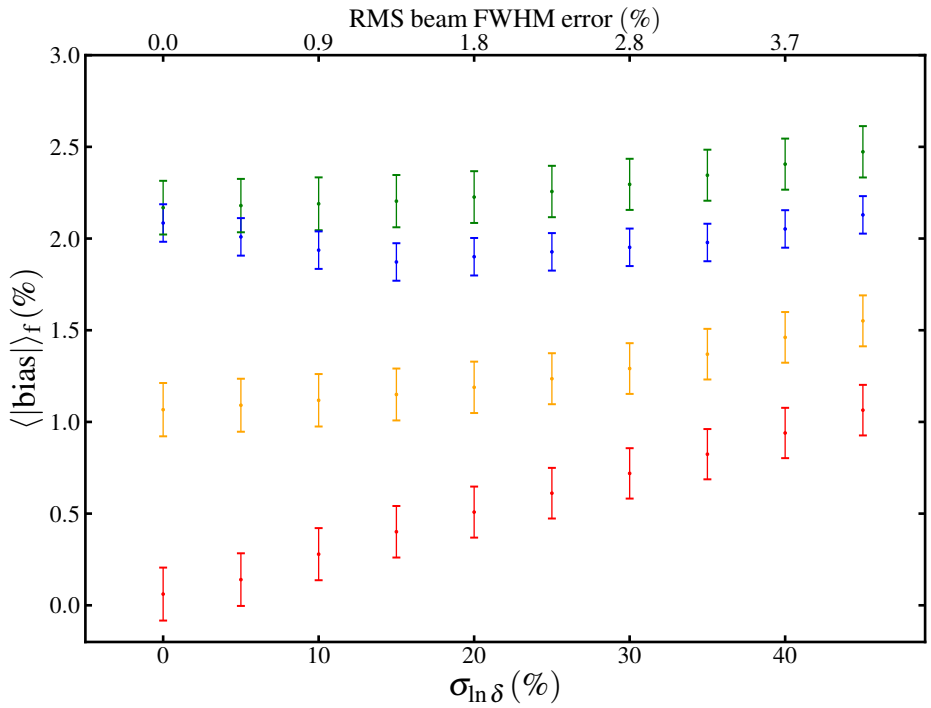


Figure 8.6: Amplitude calibration as a function of the level of uncertainty on the beam perturbation parameters δ_i. The accuracy to which the beams are known is parameterized by $\sigma_{\ln\delta}$. The extended algorithm is performed both with a perfect prior on $\Gamma^{(0)}_{b_{\min}}$ (red) as well as with 50% (orange) and 100% (green) error on this prior. The basic algorithm is shown in blue. The top horizontal axis shows the corresponding RMS error on the FWHM of the beams.

itself, we can use an eigenvector $|\lambda_i\rangle$ of $\hat{\mathbf{G}}$ (with some overall normalization) as an estimator for the gain vector $|g\rangle$. The question is now which eigenvector of $\hat{\mathbf{G}}$ is the best estimate of the gain vector?

Since $\hat{\mathbf{G}}$ is Hermitian and positive semi-definite, its eigen-decomposition coincides with its singular value decomposition (SVD). This recognition is very useful, as the SVD can be used to find a lower rank approximation for a matrix. As $\mathbf{G}$ is a rank-1 matrix, we would like to find the rank-1 matrix closest to $\hat{\mathbf{G}}$. We can see that the chi-square in Eq. (8.24) is proportional to the square of the Frobenius norm of the matrix $\hat{\mathbf{G}} - \mathbf{G}$.[64] The rank-$m$ matrix that minimizes the Frobenius norm between it and a rank-n matrix with $n > m$ can be found by decomposing the rank-n matrix with SVD and then replacing its lowest $n - m$ singular values by zeros. Reducing $\hat{\mathbf{G}}$ to a rank-1 matrix via the SVD is equivalent to minimizing the chi-square in Eq. (8.24). Since for $\hat{\mathbf{G}}$ its eigen-decomposition and SVD are the same, its singular values are its eigenvalues, so the eigenvector that will be our best estimate of the gains will be the eigenvector with the largest corresponding eigenvalue.

Once we have an estimate for the gains, we can refine our estimate of the true visibilities by minimizing Eq. (8.23) with respect to V_{i-j}, which, from differentiating with respect to the visibility V_b with baselines b, gives

$$V_b^* = \frac{\sum_i g_i g_{i+b}^* V_{i,i+b}^{\text{meas}}}{\sum_i |g_i|^2 |g_{i+b}|^2}. \tag{8.25}$$

The process of solving for the gains and subsequently the true visibilities as described above can be done iteratively until the desired level of convergence is reached.

A shortcoming of the eigenvector method as formulated above is that we had to assume that the covariance matrix $C_{(ij),(kl)}$ in Eq. (8.23) was proportional to the identity matrix and thereby have weighted all correlations equally, including autocorrelations. In this case, we would need to solve for the system temperature of each feed, in addition to the other parameters in the model. This method is then not ideal for a redundant baseline calibration of the gain amplitudes, since we would have to specify the diagonal elements of the matrix $\hat{\mathbf{G}}$. However, if we are only concerned with the phases of the gains, we can make the substitution $\hat{G}_{ij} \to \hat{G}_{ij}/|\hat{G}_{ij}|$ before preforming the SVD, so that the matrix $\hat{\mathbf{G}}$ contains only complex elements with unity norms. The diagonal elements of $\hat{\mathbf{G}}$, which estimate the autocorrelation, are then all scaled to a value of one in all cases. Therefore, we

[64] The Frobenius norm of a matrix is the square root of the sum of the squares of each element in the matrix.

can calibrate the phases of the gains using the eigenvector method without having to solve for the system temperature of each feed.

8.5.2 Phase Degeneracies

The phase solution of the redundant baseline scheme is insensitive to the degeneracies

$$V_{i-j} \to e^{i\boldsymbol{\alpha}\cdot(\mathbf{r}_i-\mathbf{r}_j)}V_{i-j} \quad \text{and} \quad \mathrm{g_i} \to \mathrm{e}^{\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{r}_\mathrm{i}}\mathrm{g_i}, \tag{8.26}$$

where $\mathbf{r}_i$ is the position of feed i and $\boldsymbol{\alpha}$ is an arbitrary vector. This transformation corresponds to a rotation of the sky in either direction or equivalently tilting the entire array by a certain angle. Note that unlike with the logarithm method, an initial arbitrary 'gauge-fixing' condition, such as that added in Eq. (8.21), does not need to be employed with our iterative method for determining the phases.

Additionally, adding a constant phase to the gains leaves the measured visibilites unchanged. However, since we are only interested in the product $g_i^* g_j$, which is invariant under this transformation, this degeneracy is inconsequential and may be fixed arbitrarily.

8.5.3 Phase Calibration Results

To illustrate the performance of the eigenvector redundant baseline phase calibration, we employ the same array configuration and simulated signals as were used for the gain amplitude calibration, which were described in Section 8.4.4, and perform the phase calibrations over the same set of 100 noise realizations.

Since we are considering a one-dimensional array, the only phase degeneracy of consequence is a tilt of the array in a plane containing the baseline vectors. As with the amplitudes, we fix this degeneracy by fitting to a prior of the phase of the visibility with the smallest baseline $V_{b_{\min}}$.

To start the iterative process, we assume that the true sky visibilities are purely real. Although in practice it may be sensible to begin with a more realistic set of phases, for an array of this size, since each iteration can be computed quickly and the solution converges with a small number of iteration, the starting point is of little consequence. This can be seen in Fig. 8.7, which shows the estimated calibrated visibilities after each iteration of the algorithm. For this calibration, all primary beams were taken to be identical and given by the two-dimensional Gaussian described in Section 8.4.4 (in other words $\delta_i = 0$ for all beams).

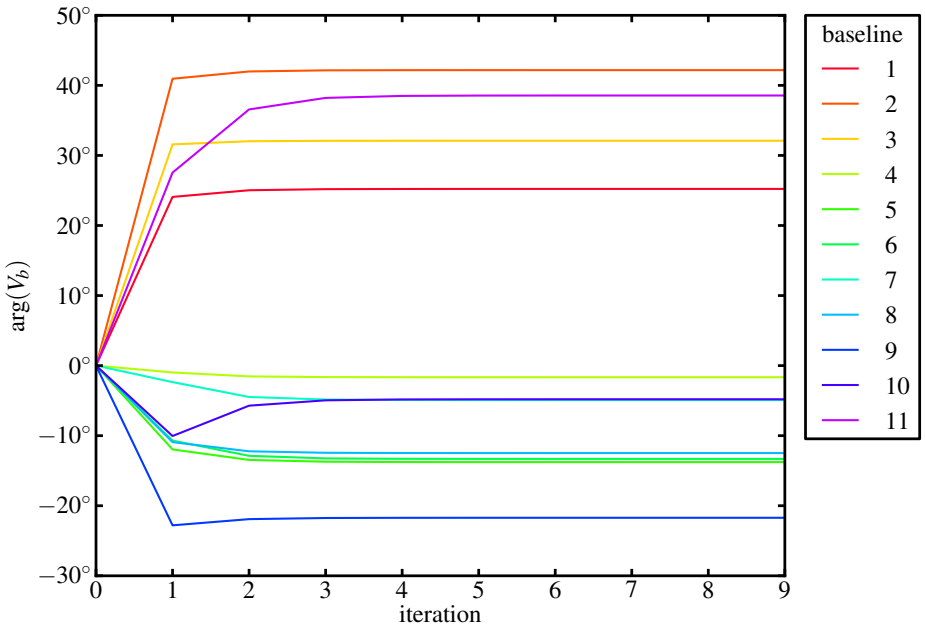


Figure 8.7: Phases of calibrated visibilities for the example calibration after each iteration of the phase calibration algorithm. Each curve represents the solution for a visibility with a particular baseline, labelled in multiples of the smallest baseline length $b_{\min}$. The simulated measured visibilities were produced with all primary beams identical.

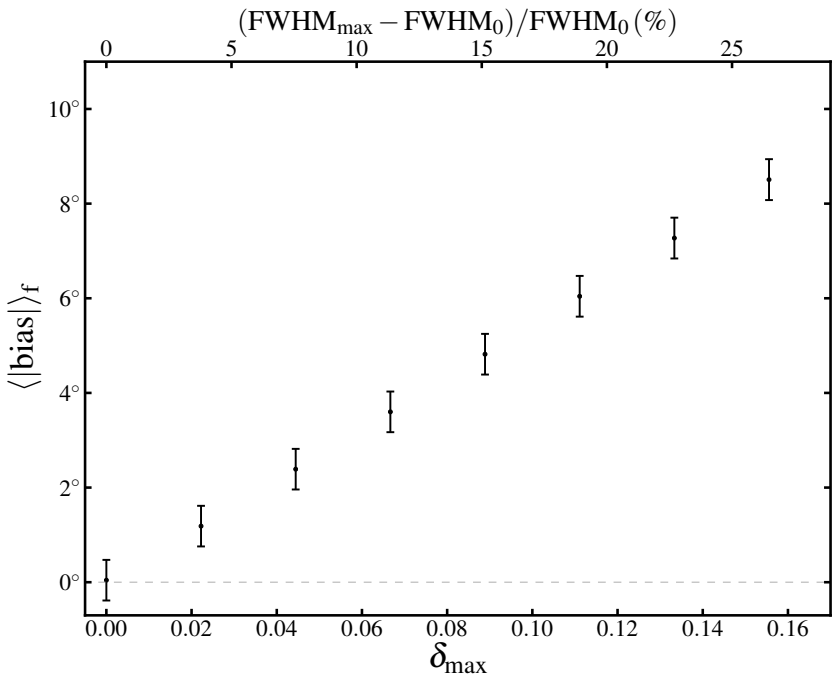


Figure 8.8: Phase calibration as a function of maximum beam perturbation parameter $\delta_{\max}$. Each point is the absolute value of the bias in the phase averaged over all of the feeds. The error bars represent the square root of the variance in the calibrated gain phases averaged over all feeds.

Although the phase calibration does not account for variations among the beams, it is crucial to evaluate its performance when the beams do vary. The bias and variance of the phase calibration as a function of the amplitude of the beam perturbations can be seen in Fig. 8.8, where the phase degeneracy is fixed using a perfect prior on the phase of $V_{b_{\min}}$. As with the gain amplitudes, the data points plotted are the absolute value of the bias averaged over all feeds and the error bar represent the square root of the variance averaged over all feeds. With all beams identical, little bias is seen in the phase calibration, but when the differences between beams are increased the bias grows. From Fig. 8.8, we can see that to achieve a phase calibration accurate to a few degrees, the widths of the beams may not vary by more than a few percent (in the absence of any other complicating factors).

The effect of introducing an error on the prior of the phase of $V_{b_{\min}}$ (used

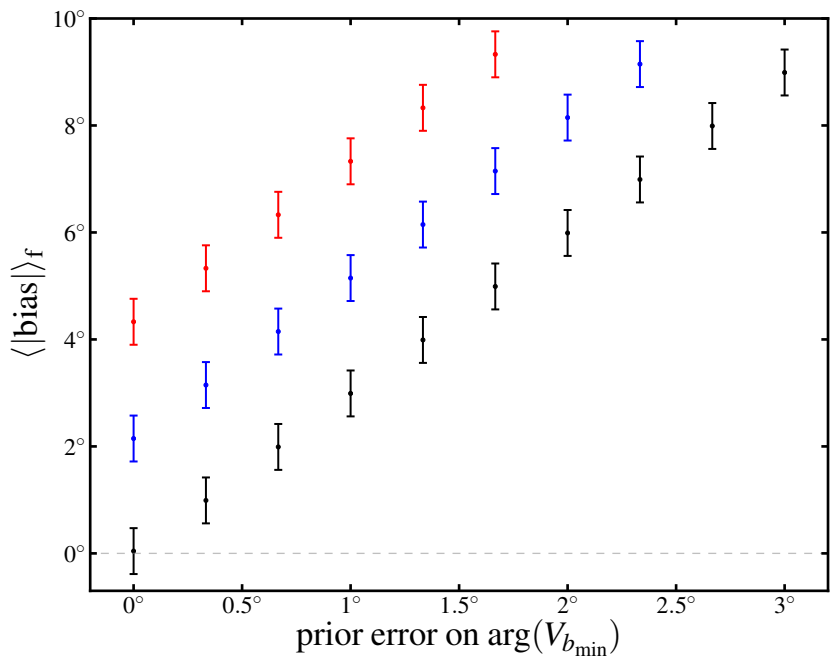


Figure 8.9: Phase calibrations as a function of the error on the phase prior of the visibility $V_{b_{\min}}$ for the smallest baseline $b_{\min}$. The black points are calibrations for which the simulated visibilities have all primary beams identical, while the blue and red points are for calibrations that have $\delta_{\max} = 0.04$ and 0.08, respectively, corresponding to a variability of the beam widths of $(\mathrm{FWHM}_{\max} - \mathrm{FWHM}_0)/\mathrm{FWHM}_0 = 6.8\%$ and 13.6%.

to fix the phase degeneracy) can be seen in Fig. 8.9, which shows the effect of increasing this error for the case of identical beams as well as when the widths differ by as much as 6.8% and 13.6% from the unperturbed beam. With all beams identical, there is room for a prior error of around a degree to achieve a phase calibration biased by only a few degrees or less, while there is little leeway for prior errors to achieve this accuracy when the beam widths differ from one another by more than a few percent.

8.6 Conclusions

With many new interferometric radio telescopes being built currently or planned for the near future, containing a large number of feeds and many

redundant baselines, redundant baseline calibration may prove to be a valuable tool for calibration. The strongest appeal of redundant baseline calibration is that the calibration is done nearly independently of our knowledge of the sky, a feature which may be important when mapping regions with little prior knowledge of the sky at the frequencies under examination.

A crucial assumption in the basic redundant baseline calibration algorithm is that each primary beam in the array is identical. However, the actual beams will be at least slightly different from one another. We have shown that by decomposing the beams into a basis of functions that can represent the beams accurately with only a small number of functions, we can adapt the amplitude redundant baseline calibration algorithm to accommodate beams that vary from feed to feed. As this introduces more parameters into our model, a larger array with more visibilities is required for this extension compared to the basic algorithm.

The extended algorithm for the amplitude calibration can yield better results than the basic algorithm when a small perturbation is added to the beams that is well represented by a single additional function. However, the extended algorithm requires as input a reasonably accurate model of the beams. In addition, the extended algorithm requires an additional piece of prior information (about either the sky or the gains) to fix an extra degeneracy that appears when the beam model is expanded to account for two different beam basis functions.

The logarithm method provides a simple approach for solving for the calibration model parameters. As with the basic algorithm, a drawback of this method is that with the split into amplitude and phase equations, the cyclic nature of the phase equations leads to phase errors in a naive implementation of this calibration algorithm. However, as the amplitude and phase equations may be solved independently of one another, this method may still be used successfully for amplitude calibration.

The eigenvector redundant baseline calibration algorithm uses an iterative scheme to solve for both the gains and true sky visibilities that treats both the gain amplitudes and phases simultaneously. However, due to a difficulty in assigning different weights to measured visibilities, for our calibration tasks this algorithm is not well suited for estimating the gain amplitudes as this would require solving for the system temperature of each feed in addition to the other parameters in the model. On the other hand, by modifying this algorithm to use only complex unit vectors, accurate phase calibrations may be produced when all beams are nearly identical. By simulating visibilities from an array with varying beam widths, we have estimated the degree of variance between the widths that is allowed if a phase calibration

accuracy of a few degrees is to be achieved.

Chapter 9

Conclusions

In this thesis, we have developed models for the behaviour and detection of inflation, dark matter, and dark energy. This work only represents a fraction of that needed to fully understand the nature of these mysterious substances/events, which will likely be a major focus of both theoretical and experimental work in cosmology for many years to come.

Precision measurements of the CMB will likely continue to be an essential tool for constraining models of inflation well into the foreseeable future. A large part of this work will be refining polarization measurements that are sensitive to primordial gravitational waves.

The examination of macroscopic models of inflation may provide a useful framework in which to view inflation. Of particular interest is the realization that there exist inflationary elastic models that have an equation of state far from $w = -1$. This work highlights the different superhorizon behaviour that may occur in some models of inflation and demonstrates that the 'separate universe approach' is a useful tool in understanding this behaviour.

The potential benefit of developing new 21-cm radiation experiments to cosmology has been highlighted throughout this thesis. It may prove to be one of the most important tools for exploring the physics of the high-redshift Universe. Although predicting the pre-reionization 21-cm signal is complex, since it is influenced by both astrophysical sources as well as cosmological phenomena, measuring both its mean value and power spectrum may provide a rich set of data from which we may learn about early structure formation, the dark ages, and the properties of dark matter.

In the later stages of the Universe's evolution, 21-cm intensity mapping can potentially measure the BAO scale during a wide range of redshifts in which the Universe was transitioning into a dark energy dominated state. Such experiments will complement BAO detections made at lower redshifts, which when combined with data from a variety of other experimental methods, are expected to provide new strong constraints on the dark energy equation of state.

Although the potential benefit of these 21-cm experiments to cosmology may be immense, a number of challenges must be overcome before rele-

vant cosmological data may be extracted. In particular, efficiently removing extremely bright foregrounds that may be four orders of magnitude larger than the 21-cm signal will likely be a daunting task. However, with new radio telescopes such as CHIME, used with novel data processing algorithms, these difficulties may be overcome.

Bibliography

[1] S. Dodelson, "Modern cosmology," Academic press, 2003.

[2] V. Mukhanov, "Physical Foundations of Cosmology," Cambridge University Press, 2005.

[3] E. W. Kolb, & M. S. Turner, "The Early Universe," Addison-Wesley Publishing Company, 1988.

[4] G. Bertone, "Particle Dark Matter: Observations, Models and Searches," Cambridge University Press, 2010.

[5] L. Amendola, & S. Tsujikawa, "Dark Energy: Theory and Observations," Cambridge University Press, 2010.

[6] Y. Wang, "Dark Energy," Wiley-VCH, 2010.

[7] K. A. Olive, G. Steigman, & T. P. Walker, "Primordial nucleosynthesis: theory and observations," *Phys Rep*, vol. 333, p. 389, 2000.

[8] H. Kodama, & M. Sasaki, "Cosmological Perturbation Theory," *Prog. Theor. Phys. Suppl.*, vol. 78, p. 1, 1984

[9] V. F. Mukhanov, H. A. Feldman, & R. H. Brandenberger, "Theory of Cosmological Perturbations," *Phys. Rep.*, vol. 215, p. 203, 1992.

[10] M. Halpern, et al., "A Digital Radio Telescope for CHIME: Three dimensional mapping of the largest volume of the Universe to date," Canada Foundation for Innovation grant proposal, 2012.

[11] D. J. Eisenstein, et al., "Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies," *ApJ*, vol. 633, p. 560, 2005.

[12] S. Cole, et al., "The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications," *MNRAS*, vol. 362, p. 505, 2005.

[13] F. Beutler, et al., "The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant," *MNRAS*, vol. 416, p. 3017, 2011.

[14] C. Blake, et al., "The WiggleZ Dark Energy Survey: mapping the distanceredshift relation with baryon acoustic oscillations," *MNRAS*, vol. 418, p. 1707, 2011.

[15] L. Anderson, et al., "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample," *MNRAS*, vol. 427, p. 3435, 2012.

[16] N. G. Busca, et al., "Baryon acoustic oscillations in the Lyα forest of BOSS quasars," *A. & A.*, vol. 552, p. 96, 2013.

[17] A. Slosar, et al., "Measurement of baryon acoustic oscillations in the Lyman-α forest fluctuations in BOSS data release 9," *JCAP*, vol. 4, p. 026, 2013.

[18] A. Font-Ribera, et al., "Quasar-Lyman α Forest Cross-Correlation from BOSS DR11: Baryon Acoustic Oscillations," arXiv:1311.1767, 2013.

[19] S. R. Furlanetto, S. Peng Oh, & F. H. Briggs., "Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe," *Phys. Rep.*, vol. 433, p. 181, 2006.

[20] J. E. Gunn, & B. A. Peterson, "On the Density of Neutral Hydrogen in Intergalactic Space," *ApJ*, vol. 142, p. 1633, 1965.

[21] R. H. Becker, et al., "Evidence for Reionization at z$\sim$6: Detection of a Gunn-Peterson Trough in a z=6.28 Quasar," *Astron. J.*, vol. 122, p. 2850, 2001.

[22] "Evolution of the Ionizing Background and the Epoch of Reionization from the Spectra of z$\sim$6 Quasars," *Astron. J.*, vol. 123, p. 1247, 2002.

[23] A. G. Riess, et al., "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *Astron. J.*, vol. 116, p. 1009, 1998.

[24] S. Perlmutter et al., "Measurements of Ω and Λ from 42 high-redshift supernovae," *ApJ*, vol. 517, p. 565, 1999.

[25] G. Hinshaw, et. al., "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results," *ApJSuppl.*, vol. 208, 19, 2013.

[26] C. -P. Ma, & E. Bertschinger, "Cosmological perturbation theory in the synchronous and conformal Newtonian gauges," *ApJ*, vol. 455, p. 7, 1995.

[27] J. M. Bardeen, et al., "The statistics of peaks of Gaussian random fields," *ApJ*, vol. 304, p. 15, 1986.

[28] D. J. Eisenstein, & W. Hu, "Baryonic features in the matter transfer function," *ApJ*, vol. 496, p. 605, 1998.

[29] Planck collaboration, "Planck 2013 results. XVI. Cosmological parameters," arXiv:1303.5076, 2013.

[30] H. Mo, F. Van den Bosch, & S. White, "Galaxy formation and evolution," Cambridge University Press, 2010.

[31] N. N. Weinberg, & M. Kamionkowski, "Constraining dark energy from the abundance of weak gravitational lenses," *MNRAS*, vol. 341, p. 251, 2003.

[32] W. H. Press, & P. Schechter, "Formation of galaxies and clusters of galaxies by self-similar gravitational condensation," *ApJ*, vol. 187, p. 425, 1974.

[33] R. K. Sheth, H. J. Mo, & G. Tormen, "Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes," *MNRAS*, vol. 323, p. 1, 2001.

[34] J. R. Bond, S. Cole, G. Efstathiou, & N. Kaiser, "Excursion set mass functions for hierarchical Gaussian fluctuations," *ApJ*, vol. 379, p. 440, 1991.

[35] A. R. Zentner, "The Excursion Set Theory of Halo Mass Functions, Halo Clustering, and Halo Growth," *Int. J. Mod. Phys. D*, vol. 16, p. 763, 2007.

[36] G. ,L. Bryan, & M. L. Norman, "Statistical properties of x-ray clusters: Analytic and numerical comparisons," *ApJ*, vol. 495, p. 80, 1998.

[37] A. H. Guth, "Inflationary universe: A possible solution to the horizon and flatness problems," *Phys. Rev. D*, vol. 23, 347, 1981.

[38] A. D. Linde, "A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems," *Phys. Lett. B*, vol. 108, p. 389, 1982.

[39] A. Albrecht, & P. J. Steinhardt, "Cosmology for grand unified theories with radiatively induced symmetry breaking," *Phys. Rev. Lett.*, vol. 48, p. 1220, 1982.

[40] S. A. Fulling, "Nonuniqueness of canonical field quantization in Riemannian space-time," *Phys. Rev. D*, vol. 7, 2850, 1973.

[41] S. W. Hawking, "Black hole explosions," *Nature* vol. 248, p. 30, 1974.

[42] P. C. W. Davies "Scalar production in Schwarzschild and Rindler metrics," *Journal of Physics A*, vol. 8, p. 609, 1975.

[43] W. G. Unruh, "Notes on black-hole evaporation," *Phys. Rev. D*, vol. 14, 870, 1976.

[44] N. D. Birrell, & P. C. W. Davies, "Quantum fields in curved space," Cambridge university press, 1984.

[45] V. Mukhanov, & S. Winitzki, "Introduction to quantum effects in gravity," Cambridge University Press, 2007.

[46] T. S. Bunch, & P. C. W. Davies, "Quantum field theory in de Sitter space: renormalization by point-splitting," *Proc. R. Soc. A*, vol. 360, p. 117, 1978.

[47] L. D. Landau, & E. M. Lifshitz, "Theory of Elasticity," Pergamon Press Ltd., New York, 3rd ed, 1986.

[48] M. Bucher, & D. N. Spergel, "Is the Dark Matter a Solid?" *Phys. Rev. D*, vol. 60, 043505, 1999.

[49] R. A. Battye, & A. Moss, "Cosmological Perturbations in Elastic Dark Energy Models," *Phys. Rev. D*,vol 76, 023005, 2007.

[50] R. A. Battye, & A. J. Pearson, "Massive gravity, the elasticity of spacetime and perturbations in the dark sector," *Phys. Rev. D*, vol 88, 084004, 2013.

[51] A. Gruzinov, "Elastic Inflation," *Phys. Rev. D*, vol. 70, 063518, 2004.

[52] S. Endlich, A. Nicolis, & J. Wang, "Solid Inflation," arXiv:1210.0569, 2012.

[53] N. Bartolo, S. Matarrese, M. Peloso, & A. Ricciardone, "Anisotropy in solid inflation," arXiv:1306.4160, 2013.

[54] B. Carter, & H. Quintana, "Foundations of general relativistic high-pressure elasticity theory," *Proc. R. Soc. A*, vol. 331, p. 57, 1972.

[55] R. A. Battye, B. Carter, E. Chachoua, & A. Moss, "Rigidity and stability of cold dark solid universe model," *Phys. Rev. D*, vol. 72, 023503, 2005.

[56] R. A. Battye, E. Chachoua, & A. Moss, "Elastic properties of anisotropic domain wall lattices," *Phys. Rev. D*, vol. 73, 123528, 2006.

[57] R. A. Battye, M. Bucher, & D. Spergel, "Domain wall dominated universes," *arXiv*:astro-ph/9908047, 1999.

[58] A. Friedland, H. Murayama, & M. Perelstein, "Domain walls as dark energy," *Phys. Rev. D*, vol. 67, 043519, 2003.

[59] B. Carter, "Speed of Sound in a High-Pressure General-Relativistic Solid," *Phys. Rev. D*, vol. 7, 1590, 1973.

[60] V. Fock, "The Theory of Space, Time and Gravitation," Pergamon Press Ltd., Oxford, 2nd rev. ed., 1964.

[61] D. Wands, K. A. Malik, D. H. Lyth, & A. R. Liddle, "New approach to the evolution of cosmological perturbations on large scales," *Phys. Rev. D*, vol. 62, 043527, 2000.

[62] D. H. Lyth, & D. Wands, "Conserved cosmological perturbations," *Phys. Rev. D*, vol 68, 103515, 2003.

[63] G. F. R. Ellis, & H. van Elst, "Cosmological Models," *Theoretical and Observational Cosmology: Proceedings of the NATO Advanced Study Institute on Theoretical and Observational Cosmology.* NATO Science Series C., vol. 541, p. 1, 1999.

[64] A. Pontzen, & A. Challinor, "Linearization of homogeneous, nearly-isotropic cosmological models," *Class. Quant. Grav.*, vol. 28, 185007, 2011.

[65] K. A. Malik, & D. Wands, "Adiabatic and entropy perturbations with interacting fluids and fields," *JCAP*, vol. 2, 7, 2005.

[66] D. Baumann, L. Senatore, & M. Zaldarriaga, "Scale-Invariance and the Strong Coupling Problem," *JCAP*, vol. 5, 4, 2011.

[67] S. Weinberg, *Phys. Rev. D*, vol. 67, 123504, 2003.

[68] A. Loeb & S. R. Furlanetto "The First Galaxies in the Universe," Princeton University Press, 2013.

[69] J. R. Pritchard & A. Loeb, "21 cm cosmology in the 21st century," *Rep. Prog. Phys.*, vol. 75, 086901, 2012.

[70] G. B. Field, "Excitation of the hydrogen 21-cm line," *Proc. I.R.E.*, vol. 46, p. 240, 1958.

[71] S. A. Wouthuysen, "On the excitation mechanism of the 21-cm (radio-frequency) interstellar hydrogen emission line," *Astron. J.*, vol. 57, p. 31, 1952.

[72] J. R. Pritchard, & S. R. Furlanetto, "Descending from on high: Lyman series cascades and spin-kinetic temperature coupling in the 21 cm line," *MNRAS*, vol. 367, p. 1057, 2006.

[73] C. M. Hirata, "Wouthuysen-Field coupling strength and application to high-redshift 21-cm radiation," *MNRAS*, vol. 367, p. 259, 2006.

[74] G. B. Field, "The Time Relaxation of a Resonance-Line Profile," *ApJ*, vol. 129, p, 551, 1959.

[75] S. Seager, D. D. Sasselov, & D. Scott, "A new calculation of the recombination epoch," *ApJ*, vol 523, L1, 1999.

[76] A. Klypin, A.V. Kravtsov, O. Valenzuela, & F. Prada, "Where Are the Missing Galactic Satellites?" *ApJ*, vol. 522, p. 82, 1999.

[77] B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel, & P. Tozzi, "Dark Matter Substructure within Galactic Halos," *ApJ*, vol. 524, L19, 1999.

[78] E. Papastergis, A. M. Martin, R. Giovanelli, & M. P. Haynes, "The Velocity Width Function of Galaxies from the 40*ApJ*, vol. 739, p. 38, 2011.

[79] P. J. E. Peebles, "The Void Phenomenon," *ApJ*, 557, 495, 2001.

[80] W. J. G. de Blok, S. S. McGaugh, A. Bosma, & V. C. Rubin, "Mass Density Profiles of Low Surface Brightness Galaxies," *ApJ*, vol. 552, L23, 2001.

[81] F. Donato, G. Gentile, P. Salucci, C. Frigerio Martins, M. I. Wilkinson, G. Gilmore, E. K. Grebel, A. Koch, & R. Wyse, "A constant dark matter halo surface density in galaxies," *MNRAS*, vol. 397, p. 1169, 2009.

[82] A. B. Newman, T. Treu, R. S. Ellis, D. J. Sand, J. Richard, P. J. Marshall, P. Capak, & S. Miyazaki, "The Distribution of Dark Matter Over Three Decades in Radius in the Lensing Cluster Abell 611," *ApJ*, vol. 706, p. 1078, 2009.

[83] M. Boylan–Kolchin, J. S. Bullock, & M. Kaplinghat, "Too big to fail? The puzzling darkness of massive Milky Way subhaloes," *MNRAS*, vol. 415, L40, 2011.

[84] M. Boylan–Kolchin, J. S. Bullock, & M. Kaplinghat, "The Milky Ways bright satellites as an apparent failure of ΛCDM," *MNRAS*, vol. 422, p. 1203, 2012.

[85] S. Garrison-Kimmel, M. Rocha, M. Boylan-Kolchin, J. Bullock, & J. Lally, "Can feedback solve the too-big-to-fail problem?," *MNRAS*, vol. 433, p. 3539, 2013.

[86] F. Governato, B. Willman, L. Mayer, A. Brooks, G. Stinson, O. Valenzuela, J. Wadsley, & T. Quinn, "Forming disc galaxies in ΛCDM simulations," *MNRAS*vol. 374, p. 1479, 2007.

[87] A. Pontzen, & F. Governato, "How supernova feedback turns dark matter cusps into cores," *MNRAS*, vol. 421, p. 3464, 2012.

[88] E. Sobacchi, & A. Mesinger, "How does radiative feedback from an ultraviolet background impact reionization?" *MNRAS*, vol. 432, p. 3340, 2013.

[89] R. Teyssier, A. Pontzen, Y. Dubois, & J. I. Read, "Cusp-core transformations in dwarf galaxies: observational predictions," *MNRAS*, vol. 429, p. 3068, 2013.

[90] D. N. Spergel, & P. J. Steinhardt, "Observational Evidence for Self-Interacting Cold Dark Matter," *Phys. Rev. Lett.*, vol. 84, p. 3760, 2000.

[91] A. Burkert, "The Structure and Evolution of Weakly Self-interacting Cold Dark Matter Halos," *ApJ*, vol. 534, L143, 2000.

[92] R. Davé, D. N. Spergel, P. J. Steinhardt, & B. D. Wandelt, "Halo Properties in Cosmological Simulations of Self-interacting Cold Dark Matter," *ApJ*, vol. 547, p. 574, 2001.

[93] F.-Y. Cyr-Racine, & K. Sigurdson, "Cosmology of atomic dark matter," *Phys. Rev. D*, vol. 87, 103515, 2013.

[94] D. E. Kaplan, G. Z. Krnjaic, K. R. Rehermann, & C. M. Wells, "Atomic dark matter," *JCAP*, vol. 5, p. 1475, 2010.

[95] A. Boyarsky, J. Lesgourgues, O. Ruchayskiy, & M. M. Viel, "Realistic Sterile Neutrino Dark Matter with keV Mass does not Contradict Cosmological Bounds," *Phys. Rev. Lett.*, vol. 102, p. 201304, 2009.

[96] K. Abazajian, G. M. Fuller, & M. Patel, "Sterile neutrino hot, warm, and cold dark matter," *Phys. Rev. D*, vol. 64, 023501, 2001.

[97] S. Dodelson, & L. M. Widrow, "Sterile neutrinos as dark matter," *Phys. Rev. Lett.*, vol. 72, p. 17, 1994.

[98] J. R. Bond, A. S. Szalay, & M. S Turner, "Formation of Galaxies in a Gravitino-Dominated Universe," *Phys. Rev. Lett.*, vol. 48, p. 1636, 1982.

[99] H. Pagels, & J. R. Primack, "Supersymmetry, Cosmology, and New Physics at Teraelectronvolt Energies," *Phys. Rev. Lett.*, vol. 48, p. 223, 1982.

[100] A. Mesinger, R. Perna, & Z. Haiman, "Constraints on the Small-Scale Power Spectrum of Density Fluctuations from High-Redshift Gamma-Ray Bursts," *ApJ*, vol. 623, p. 1, 2005.

[101] F. Pacucci, A. Mesinger, & Z. Haiman, "Focusing on warm dark matter with lensed high-redshift galaxies," *MNRAS*, vol. 435, L53, 2013.

[102] R. Barkana, Z. Haiman, & J. P. Ostriker, "Constraints on Warm Dark Matter from Cosmological Reionization," *ApJ*, vol. 558, p. 482, 2001.

[103] R. S. de Souza, A. Mesinger, A. Ferrara, Z. Haiman, R. Perna, & N. Yoshida, "Constraints on warm dark matter models from high-redshift long gamma-ray bursts," *MNRAS*, vol. 432, p. 3218, 2013.

[104] X. Kang, A. V. Macciò, & A. A. Dutton, "The Effect of Warm Dark Matter on Galaxy Properties: Constraints from the Stellar Mass Function and the Tully-Fisher Relation," *ApJ*, vol. 767, p. 22, 2013.

[105] V. K. Narayanan, D. N. Spergel, R. Davé, & C. P. Ma, "Constraints on the Mass of Warm Dark Matter Particles and the Shape of the Linear Power Spectrum from the Ly? Forest," *ApJ*, vol. 543, L103, 2000.

[106] U. Seljak, A. Makarov, P. McDonald, & H. Trac, "Can Sterile Neutrinos Be the Dark Matter?," *Phys. Rev. Lett.*, vol. 97, p. 191303, 2006.

[107] M. Viel, G. D. Becker, J. S. Bolton, M. G. Haehnelt, M. Rauch, & W. L. Sargent, "How Cold Is Cold Dark Matter? Small-Scales Constraints from the Flux Power Spectrum of the High-Redshift Lyman-α Forest," *Phys. Rev. Lett.*, vol. 100, p. 041304, 2008.

[108] M. Viel, J. Lesgourgues, M. G. Haehnelt, S. Matarrese, & A. Riotto, "Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-α forest," *Phys. Rev. D*, vol. 71, 063534, 2005.

[109] M. Viel, G. D. Becker, J. S. Bolton, & M. G. Haehnelt, "Warm Dark Matter as a solution to the small scale crisis: new constraints from high redshift Lyman-alpha forest data," *Phys. Rev. D*, vol. 88, 043502, 2013.

[110] M. R. Lovell, V. Eke, C. S. Frenk, L. Gao, A. Jenkins, T. Theuns, J. Wang, S. D. M. White, A. Boyarsky, & O. Ruchayskiy, "The haloes of bright satellite galaxies in a warm dark matter universe," *MNRAS*, vol. 420, p. 2318, 2012.

[111] A. V. Macciò, S. Paduroiu, D. Anderhalden, A. Schneider, & B. Moore, "Cores in warm dark matter haloes: a Catch 22 problem," *MNRAS*, vol. 424, p. 1105, 2012.

[112] F. Villaescusa-Navarro, & N. Dalal, "Cores and cusps in warm dark matter halos," *JCAP*, vol. 3, p. 1475, 2011.

[113] H. J. de Vega, M. C. Falvella, & N. G. Sanchez, "Towards the Chalonge 17th Paris Cosmology Colloquium 2013: highlights and conclusions of the Chalonge 16th Paris Cosmology Colloquium 2012," arXiv:1307.1847, 2013.

[114] P. Madau, A. Meiksin, & M. J. Rees, "21 Centimeter Tomography of the Intergalactic Medium at High Redshift," *ApJ*, vol. 475, p. 429, 1997.

[115] A. Mesinger, A. Ewall-Wice, & J. Hewitt, "Reionization and beyond: detecting the peaks of the cosmological 21-cm signal," *MNRAS*, vol. 439, p. 3262, 2013a

[116] M. F. Morales, & J. S. B. Wyithe, "Reionization and Cosmology with 21-cm Fluctuations," *ARA&A*, vol. 48, p. 127, 2010.

[117] M. Zaldarriaga, S. Furlanetto, & L. Hernquist, "21 Centimeter Fluctuations from Cosmic Gas at High Redshifts," *ApJ*, vol. 608, p. 622, 2004.

[118] D. S. Gorbunov, & V. A. Rubakov, "Introduction to the theory of the early Universe: Hot Big Bang Theory," World Scientific, 2011.

[119] D. S. Gorbunov, & V. A. Rubakov, "Introduction to the theory of the early Universe: Cosmological perturbations and inflationary theory," World Scientific, 2011.

[120] P. Bode, J. P. Ostriker, & N. Turok, "Halo Formation in Warm Dark Matter Models," *ApJ*, vol. 556, p. 93, 2001.

[121] A. Jenkins, C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, & N. Yoshida, "The mass function of dark matter haloes," *MNRAS*, vol. 321, p. 372, 2001.

[122] S. Bharadwaj, & S. S. Ali, "The cosmic microwave background radiation fluctuations from HI perturbations prior to reionization," *MNRAS*, vol. 352, p. 142, 2004.

[123] A. Lewis, & A. Challinor, "21 cm angular-power spectrum from the dark ages," *Phys. Rev. D*, vol. 76, 083005, 2007.

[124] A. Loeb, & M. Zaldarriaga, "Measuring the Small-Scale Power Spectrum of Cosmic Density Fluctuations through 21 cm Tomography Prior to the Epoch of Structure Formation," *Phys. Rev. Lett.*, vol. 92, p. 211301, 2004.

[125] S. Naoz, & R. Barkana, "Growth of linear perturbations before the era of the first galaxies," *MNRAS*, vol. 362, p. 1047, 2005.

[126] M. Mapelli, A. Ferrara, & E. Pierpaoli, "Impact of dark matter decays and annihilations on reionization," *MNRAS*, vol. 369, p. 1719, 2006.

[127] M. Valdés, C. Evoli, A. Mesinger, A. Ferrara, & N. Yoshida, "The nature of dark matter from the global high-redshift HI 21?cm signal," *MNRAS*, vol. 429, p. 1705, 2013.

[128] A. Mesinger, A. Ferrara, & D. S. Spiegel, "Signatures of X-rays in the early Universe," *MNRAS*, vol. 431, p. 621, 2013b.

[129] A. Mesinger, S. Furlanetto, & R. Cen, "21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal," *MNRAS*, vol. 411, p. 955, 2011.

[130] H.-J. Grimm, M. Gilfanov, & R. Sunyaev, "High-mass X-ray binaries as a star formation rate indicator in distant galaxies," *MNRAS*, vol. 339, p. 793, 2003.

[131] R. Barkana, & A. Loeb, "Detecting the earliest galaxies through two new sources of 21 centimeter fluctuations," *ApJ*, vol. 626, p. 1, 2005.

[132] R. Barkana, & A. Loeb, "Unusually Large Fluctuations in the Statistics of Galaxy Formation at High Redshift," *ApJ*, vol. 609, p. 474, 2004.

[133] Z. Haiman, T. Abel, & M. J. Rees, "The Radiative Feedback of the First Cosmological Objects," *ApJ*, vol. 534, p. 11, 2000.

[134] A. Mesinger, G. L. Bryan, & Z. Haiman, "Relic HII regions and radiative feedback at high redshifts," *MNRAS*, vol. 399, p. 1650, 2009.

[135] R. E. Smith, & K. Markovic, "Testing the warm dark matter paradigm with large-scale structures," *Phys. Rev. D*, vol. 84, 063507, 2011.

[136] J. C .Pober, et al., "Opening the 21cm EoR Window: Measurements of Foreground Isolation with PAPER," *ApJ*, vol. 768, L36, 2013.

[137] A. Albrecht, et al., "Report of the dark energy task force," arXiv:0609591, 2006.

[138] W. Hu, & N. Sugiyama, "Small-scale cosmological perturbations: an analytic approach," *ApJ*, vol. 471, p. 542, 1996.

[139] H. Seo, et al., "A ground-based 21 cm baryon acoustic oscillation survey," *ApJ*, vol. 721, p. 164, 2010.

[140] H. Seo, & D. J. Eisenstein, "Improved forecasts for the baryon acoustic oscillations and cosmological distance scale," *ApJ*, vol. 665, p. 14, 2007.

[141] J. R. Shaw, K. Sigurdson, U. -L. Pen, A. Stebbins, & M. Sitwell, "All-Sky Interferometry with Spherical Harmonic Transit Telescopes," *ApJ*, vol. 781, p. 57, 2014.

[142] J. R. Shaw, K. Sigurdson, M. Sitwell, A. Stebbins, & U. -L. Pen, "Coaxing Cosmic 21cm Fluctuations from the Polarized Sky using m-mode Analysis," arXiv:1401.2095, 2014.

[143] H. A. Feldman, N. Kaiser, & J. A. Peacock, "Power spectrum analysis of three-dimensional redshift surveys," *ApJ*, vol. 426, p. 23, 1994.

[144] M. Tegmark, "Measuring Cosmological Parameters with Galaxy Surveys," *Phys. Rev. D*, vol. 79, 3806, 1997.

[145] T. -C. Chang, et al., "Baryon acoustic oscillation intensity mapping of dark energy," *Phys. Rev. Lett.*, vol. 100, p. 091303, 2008.

[146] E. R. Switzer, et al., "Determination of $z \sim 0.8$ neutral hydrogen fluctuations using the 21 cm intensity mapping auto-correlation," *MNRAS*, vol. 434, L46, 2013.

[147] M. A. Zwaan, et al., "The HIPASS catalogue: ΩHI and environmental effects on the HI mass function of galaxies," *MNRAS*, vol. 359, L30, 2005.

[148] A. Liu, & M. Tegmark, "A method for 21 cm power spectrum estimation in the presence of foregrounds," *Phys. Rev. D*, vol. 83, 103006, 2011.

[149] A. R. Parsons, & D. C. Backer, "Calibration of low-frequency, wide-field radio interferometers using delay/delay-rate filtering," *Astron. J.*, vol. 138, 219, 2009.

[150] A. R. Parsons, et al., "A per-baseline, delay-spectrum technique for accessing the 21 cm cosmic reionization signature," *ApJ*, vol. 756, p. 165, 2012.

[151] D. J. Eisenstein, H. Seo, & M. White, "On the robustness of the acoustic scale in the low-redshift clustering of matter," *ApJ*, vol. 664, p. 660, 2007.

[152] N. Padmanabhan, et al., "A 2 per cent distance to z= 0.35 by reconstructing baryon acoustic oscillationsI. Methods and application to the Sloan Digital Sky Survey," *MNRAS*, vol. 427, p. 2132, 2012.

[153] C. Blake, et al., "The WiggleZ Dark Energy Survey: testing the cosmological model with baryon acoustic oscillations at z= 0.6," *MNRAS*, vol. 415, p. 2892, 2011.

[154] M. Tegmark, A. N. Taylor, & A. F. Heavens, "Karhunen-Loève eigenvalue problems in cosmology: How should we tackle large data sets?," *ApJ*, vol. 480, p.22, 1997.

[155] K. S. Dawson, et al., "The Baryon Oscillation Spectroscopic Survey of SDSS-III," *Astron. J.*, vol. 145, p. 10, 2013.

[156] M. F. Morales, & J. Hewitt, "Toward epoch of reionization measurements with wide-field radio observations," *ApJ*, vol. 615, p. 7, 2004.

[157] Y. Mao, et al., "How accurately can 21 cm tomography constrain cosmology?" *Phys. Rev. D*, vol. 78, 023529, 2008.

[158] M. Tegmark, & M. Zaldarriaga, "The Fast Fourier Transform Telescope," *Phys. Rev. D*, vol. 79, 083530, 2009.

[159] M. Tegmark, & M. Zaldarriaga, "Omniscopes: Large Area Telescope Arrays with only N log N Computational Cost," *Phys. Rev. D*, vol. 82, 103501, 2010.

[160] A. Parsons, et al., "A sensitivity and array-configuration study for measuring the power spectrum of 21 cm emission from reionization," *ApJ*, vol. 753, p.81, 2012.

[161] A. Liu, et al., "Global 21cm signal experiments: A designer's guide," *Phys. Rev. D*, vol. 87, 043002, 2013.

[162] J. E. Noordam, & A. G. De Bruyn, "High dynamic range mapping of strong radio sources, with application to 3C84," *nature*, vol. 299, p. 597, 1982.

[163] P. Noorishad, et al., "Redundancy Calibration of Phased Array Stations," *A. & A.*, vol. 545, p. 108, 2012.

[164] A. Liu, et al., "Precision calibration of radio interferometers using redundant baselines," *MNRAS*, vol. 408, p. 1029, 2010.

[165] V. R. Marthi, & J. Chengalur, "Non-linear redundancy calibration," *MNRAS*, vol. 437, p. 524, 2014.

[166] S. J. Wijnholds, & P. Noorishad, "Statistically optimal self-calibration of regular imaging arrays," *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, p. 1304, 2012.

[167] Y. Yang, "A comparison of two self-calibration techniques for radio interferometric data," *A. & A.*, vol. 189, p. 361, 1988.

[168] M. H. Wieringa, "An investigation of the telescope based calibration methods redundancy and self-cal," *Experimental Astronomy*, vol. 2, p. 203, 1992.

[169] C. G. T. Haslam, et al., "A 408 MHz all-sky continuum survey. II-The atlas of contour maps," *A. & A.S.*, vol. 47, p. 1, 1982.

Appendix A

Supplemental Details for Elastic Solid Model of Inflation

A.1 Equations of Motion for Scalar and Tensor Perturbations

In Section 4.4, we derived the equations of motion for the mode functions χ and X for the scalar and tensor linear perturbations, respectively, from the action, which can then be used to find the equations of motion for R and h_{ij}^T. Alternatively, we can derive these equations of motion directly from the Einstein equations and the properties of the elastic solid given in Eq. (4.20). Although we can derive the equations of motion for u and U_p in this manner, it is only through the action in which we can properly identify u and U_p as the canonical variables for the scalar and tensor linear perturbations, respectively.

As before, to derive the equations of motion for the scalar perturbations, it is convenient to work in the comoving gauge and will also move to Fourier space. We can combine the Fourier space versions of Eqs. (4.31a) and (4.31c) to eliminate δ and then combine the derivative of this equation with Eqs. (4.9a), (4.9c), and (4.9d) in the comoving gauge to eliminate δ, E' and E'', then use Eq. (4.31b) and its derivative to eliminate ϕ and ϕ', which gives

$$\begin{aligned}\psi_{\mathbf{k}}'' + \left[\left(2 + 3\left[w - \frac{dP}{d\rho}\right]\right)\mathcal{H} - \left(\ln\left[\frac{dP}{d\rho}\right]\right)'\right]\psi_{\mathbf{k}}' + \frac{dP}{d\rho}k^2\psi_{\mathbf{k}} \\ + \frac{\mathcal{H}}{3(1+w)\frac{dP}{d\rho}}\left[3\mathcal{H}\frac{dP}{d\rho}\left(3w + w^2 - 2\frac{dP}{d\rho}\right) - 2w\left(\frac{dP}{d\rho}\right)'\right]\Pi_{\mathbf{k}} \\ + \frac{2}{3}\frac{w}{1+w}\mathcal{H}\Pi_{\mathbf{k}}' = 0. \quad \text{(A.1)}\end{aligned}$$

This result does not assume any properties of the substance occupying our spacetime, other than being able to write its stress-energy tensor in the form in Eq. (4.8).

We now specify the elastic properties of our substance by employing Eq. (4.22), which can be written in terms of $\psi_{\mathbf{k}}$ by use of the Fourier version of Eq. (4.35). Substituting this into Eq. (A.1) yields

$$R''_{\mathbf{k}} + 2\frac{z'}{z}R'_{\mathbf{k}} + \left[c_{\mathrm{s}}^2 k^2 + m_{\mathrm{eff,S}}^2 + \frac{z''}{z}\right] R_{\mathbf{k}} = 0, \tag{A.2}$$

where z and $m_{\mathrm{eff,S}}^2$ were defined in Eqs. (4.37) and (4.40), respectively, and have used the fact that $R = \psi$ in the comoving gauge. By substituting u for R using Eq. (4.38), we can recover the equation of motion for u in Eq. (4.41).

Obtaining the equation of motion for the tensor perturbations is very straightforward, since there is only one Einstein equation for the tensor perturbations, given in Eq. (4.12). With the tensor part of the anisotropic stress for the elastic solid in Eq. (4.24), the equation of motion for the tensor perturbations is

$$(h_{\mathbf{k}}^{\mathrm{T}})^{i\prime\prime}_{j} + 2\mathcal{H}(h_{\mathbf{k}}^{\mathrm{T}})^{i\prime}_{j} + (k^2 + 4c_{\mathrm{v}}^2\beta)(h_{\mathbf{k}}^{\mathrm{T}})^{i}_{j} = 0, \tag{A.3}$$

where we have switched to Fourier space.

A.2 Multicomponent System with Energy-Momentum Transfer

In this appendix, we review the equations governing the energy-momentum transfer between multiple 'fluid-like' substances.[65] This discussion will follow the work of Ref. [65]. We will begin by stating the general equations for energy-momentum transfer between any number of 'fluid-like' substances and then specialize to the case of an elastic solid decaying into a perfect fluid. As in Section 4.8, we will use the coordinate time instead of conformal time in this section.

The energy-momentum transfer 4-vector $Q^{\nu}_{(\alpha)}$ that appears in Eq. (4.114) must vanish when summed over all substances α,

$$\sum_{\alpha} Q^{\nu}_{(\alpha)} = 0, \tag{A.4}$$

[65] By 'fluid-like', we are not referring to a perfect fluid, but instead any substance whose stress-energy tensor can be parameterized by Eq. (4.8), which includes an elastic solid.

for the total stress-energy tensor to be covariantly conserved. The conservation equation for the background energy density ρ_α of substance α is

$$\dot{\rho}_\alpha = -3H(\rho_\alpha + P_\alpha) + Q_\alpha, \tag{A.5}$$

where $Q_\alpha = Q^0_{(\alpha)}$. Importantly, from Eq. (A.4), the conservation equation for the total energy density given in Eq. (4.4) still holds.

If all substances have vanishing intrinsic nonadiabatic pressure ($\delta P_\alpha = (dP_\alpha/d\rho_\alpha)\delta\rho_\alpha$) then the equations of motion for ζ_α and R_α, defined in Eqs. (4.115) and (4.117), are given by

$$\dot{\zeta}_\alpha = -\frac{H}{\dot{\rho}_\alpha}(\delta Q_{\mathrm{intr},\alpha} + \delta Q_{\mathrm{rel},\alpha}) - \frac{\dot{H}}{H}(R - \zeta) + \frac{k^2}{3a^2 H}\left(1 - \frac{Q_\alpha}{\dot{\rho}_\alpha}\right) R_\alpha \tag{A.6}$$

and

$$\dot{R}_\alpha = \frac{\dot{H}}{H}(R_\alpha - R) - \frac{\dot{\rho}_\alpha}{\rho_\alpha + P_\alpha}\frac{dP_\alpha}{d\rho_\alpha}(R_\alpha - \zeta_\alpha) - \frac{H}{\rho_\alpha + P_\alpha}\left(\frac{2}{3}P_\alpha \Pi_\alpha + f_{\mathrm{rel},\alpha}\right), \tag{A.7}$$

where we have written the equations in Fourier space, but neglected to write the subscript k to minimize the number of subscripts written on each variable. In the above equations, $\delta Q_{\mathrm{intr},\alpha}$ and $\delta Q_{\mathrm{rel},\alpha}$ are the intrinsic and relative nonadiabatic energy transfer perturbations, respectively, and $f_{\mathrm{rel},\alpha}$ is the relative momentum transfer, which are defined as

$$\delta Q_{\mathrm{intr},\alpha} \equiv \delta Q_\alpha - \frac{\dot{Q}_\alpha}{\dot{\rho}_\alpha}\delta\rho_\alpha, \tag{A.8a}$$

$$\delta Q_{\mathrm{rel},\alpha} \equiv -\frac{Q_\alpha}{6H\rho}\sum_\beta \dot{\rho}_\beta \mathcal{S}_{\alpha\beta}, \tag{A.8b}$$

$$f_{\mathrm{rel},\alpha} \equiv aQ_\alpha \sum_\beta \frac{\rho_\beta + P_\beta}{\rho + P} v_{\alpha\beta}, \tag{A.8c}$$

where δQ_α is the perturbation to Q_α.

With Q^ν_e for the decay of the elastic solid specified in Eq. (4.121), we have $Q_e = -Q_f = -\Gamma\rho_e(1 + w_e)$. Since Q_e is a function of ρ_e, $\delta Q_{\mathrm{intr},e} = 0$. Using this fact and $\delta Q_f = -\delta Q_e$, we find that $\delta Q_{\mathrm{intr},f} = \dot{Q}_e \mathcal{S}_{ef}/3H$.

A.3 Scalar Amplitude

In this appendix we give detailed expressions for the scalar amplitude evaluated at the pivot scale k_p, as given in Eq. (4.126). We first examine the case where the sound speeds and equation of state are varying slowly in time and $\epsilon_0 = c_{s0} = 0$. From Eqs. (4.103) and (4.125), Δ is given by

$$\Delta = 0.11 \times 10^{5.23(n_s-0.96)}\Gamma^2(\gamma_S) \\ (3 - 18\epsilon_1 - 8c_{s1}^2\epsilon_1 - 2\tau_s + 2\tau_\epsilon) \\ \times g_{s0}^{\frac{1-n_s}{3}}(T_0/k_p)^{1-n_s}, \quad \text{(A.9)}$$

where we have assumed that the reheating process is very rapid. Choosing the pivot scale as $k_p = 0.002\,\mathrm{Mpc}^{-1}$, and using the CMB temperature $T_0 = 2.725\,K$ and $g_{s0} = 43/11$, Δ becomes

$$\Delta = 1.53 \times 10^{-28.5(n_s-0.96)}\Gamma^2(\gamma_S)(3 - 18\epsilon_1 - 8c_{s1}^2\epsilon_1 - 2\tau_s + 2\tau_\epsilon). \quad \text{(A.10)}$$

In the case of constant sound speeds and equation of state, Δ is found to be

$$\Delta = 5.82 \times 10^{-5-22.9(n_s-1)}\Gamma^2(\nu)(21 + n_s(n_s - 10))^2|1 + 3w|^{7-n_s}. \quad \text{(A.11)}$$