# Gravitational Waves and the Early Universe

Latham A. Boyle

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Physics

November, 2006

© Copyright 2006 by Latham A. Boyle.

All rights reserved.

# Abstract

Can we detect primordial gravitational waves (*i.e.* tensor perturbations)? If so, what will they teach us about the early universe? These two questions are central to this two part thesis.

First, in chapters 2 and 3, we compute the gravitational wave spectrum produced by inflation. We argue that if inflation is correct, then the scalar spectral index $n_s$ should satisfy $n_s \lesssim 0.98$; and if $n_s$ satisfies $0.95 \lesssim n_s \lesssim 0.98$, then the tensor-to-scalar ratio $r$ should satisfy $r \gtrsim 0.01$. This means that, if inflation is correct, then primordial gravitational waves are likely to be detectable. We compute in detail the "tensor transfer function" $T_t(k, \tau)$ which relates the tensor power spectrum at two different times $\tau_1$ and $\tau_2$, and the "tensor extrapolation function" $E_t(k, k_*)$ which relates the primordial tensor power spectrum at two different wavenumbers $k$ and $k_*$. By analyzing these two expressions, we show that inflationary gravitational waves should yield crucial clues about inflation itself, and about the "primordial dark age" between the end of inflation and the start of big bang nucleosynthesis (BBN).

Second, in chapters 4 and 5, we compute the gravitational wave spectrum produced by the cyclic model. We examine a surprising duality relating expanding and contracting cosmological models that generate the same spectrum of gauge-invariant Newtonian potential fluctuations. This means that, if the cyclic model is correct, then it cannot be distinguished from inflation by observing primordial scalar perturbations alone. Fortunately, gravitational waves may be used to cleanly discriminate between the inflationary and cyclic scenarios: we show that BBN constrains the gravitational wave spectrum generated by the cyclic model to

be so suppressed that it cannot be detected by any known experiment. Thus, the detection of a primordial gravitational wave signal would rule out the cyclic model.

# Acknowledgements

I was not a very science-oriented child — never particularly passionate about the planets, or collecting/torturing insects, or blowing things up; and I certainly never dreamed of having any remotely scientific occupation as an adult. My high school physics teacher, a remarkable man named Robert Shurtz, was primarily responsible for changing my perspective. Then, in college, one of my professors — Gerry Gabrielse, an atomic physicist — invited me to join his research group. His encouragement, and the excitement of working in his lab, were crucial factors in my decision to pursue a PhD in physics.

Since coming to graduate school, I have learned from many people. First and foremost, I would like to thank my advisor, Paul Steinhardt. Many aspects of his approach to research in general, and cosmology in particular, have made a deep impression on me: the relative weight that he assigns to different arguments; his judgement about which problems are most important, or most likely to yield fruit; his sense of which pieces of the cosmological puzzle are firmly in place, and which pieces are looser and may yet be rearranged; his simultaneous attention to both the broader features and the finer details of that puzzle. Much of the research in this thesis was suggested by him, directly or indirectly, and all of it was done in collaboration with him.

I would also like to acknowledge my other collaborators: Robert Caldwell, Neil Turok, and particularly Marc Kamionkowski who introduced me to theoretical research and cosmology.

I am quite in awe of the swirl of research — experimental and theoretical, past and present — encompassed by the Gravity Group at Princeton. It has been a privelege to be

a part of it. It has also been a *pleasure* to be a part of it, thanks in large measure to the friendly tone set by the senior members.

I have learned about physics and math and various other neat things from many conversations with Chris Beasley, Joel Erickson, Steve Gratton, Chris Hirata, Kevin Huffenberger, Justin Khoury, Max Lieblich, Katie Mack, Alexey Makarov, Liam McAllister, Annika Peter, Sri Raghu, David Shih, Keitaro Takahashi, and Amol Upadhye. I am particularly happy to acknowledge Mike Kesden, with whom I learned cosmology and played squash during my first year of graduate school; Kumar Raman, who helped me to properly appreciate the movie *Purple Rain*; and Daniel Baumann, Nikhil Padmanabhan, Andrew Tolley, and Daniel Wesley — fellow cosmologists and friends — for countless discussions about physics and life in general, and for enduring my whistling habit nonviolently. My graduate school experience has been enriched by friendship and conversation with Aparna and Richard Bole, Oleg Evnin, Jenny Hoffman and Daniel Larson, Inna Kozinsky, Chris Lee, Gabe Mendlow, Sandi Milbourne, Andrew Mitchell, Aparna Ravi, Ben Rahn, Jeremy Smith, Lauren Williams, and James Wilson. Laurel Lerner and Angela Glenn have been unfailingly helpful and fun to talk to. My friend David Schur and my brother Matt Boyle have had a profound effect on the way I view the world, and the things that I think are important.

My family. I have written and rewritten this paragraph and nothing seems good enough. I am indebted to Matt, Alex, Nelya, Joseph, and Anna for their love and support. I am especially grateful to my parents, Ivy and James Boyle, to whom I dedicate this thesis.

Finally, I would like to thank my wife, Elina, who has been my best friend since before I knew anything about physics.

*

As a graduate student, I have received much-appreciated funding from a National Science Foundation Graduate Research Fellowship, from my advisor's Department-of-Energy grant, and from Princeton University.

# Contents

**Abstract** **iii**

**Acknowledgements** **v**

**Contents** **vii**

**List of Figures** **ix**

**List of Tables** **xii**

**1 Introduction** **1**

**2 Observable predictions from inflation** **7**

2.1 Introduction . . . . . 7

2.2 Inflationary predictions for scalar and tensor fluctuations reconsidered . . . 8

2.3 Additional remarks . . . . . 17

**3 What can inflationary gravitational waves teach us?** **22**

3.1 Introduction . . . . . 22

3.2 The tensor extrapolation function $E_t(k, k_*)$ . . . . . 30

3.3 The tensor transfer function $T_t(k, \tau)$ . . . . . 41

3.3.1 Tensor perturbations: fundamentals and conventions . . . . . 44

3.3.2 Organizing the calculation . . . . . 50

3.3.3 The redshift-suppression factor, $C_1$ . . . . . 52

3.3.4 The horizon-crossing factor, $C_2$ . . . 55
3.3.5 The anisotropic-stress damping factor, $C_3$ . . . 57
3.3.6 Equation-of-state corrections, $\delta w_r$ . . . 60
3.3.7 Transfer function summary . . . 65
3.4 Discussion . . . 66

**4 An expansion/contraction duality in ordinary FRW cosmology 74**
4.1 Introduction . . . 74
4.2 Background model . . . 78
4.3 Scalar perturbations . . . 79
4.4 Dominant and subdominant modes . . . 84
4.5 Tensor perturbations . . . 88
4.6 Other dualities . . . 89
4.7 Geometric interpretation . . . 92
4.8 Generalization to arbitrary spacetime dimension . . . 92
4.9 Generalization to time-varying $\epsilon$ . . . 96
4.9.1 $2+1$ dimensions . . . 97
4.9.2 $3+1$ dimensions . . . 100
4.10 Discussion . . . 103

**5 The gravitational wave spectrum from the cyclic model 105**
5.1 Introduction . . . 105
5.2 Background evolution . . . 107
5.3 Primordial strain spectrum, $\Delta h(k, \tau_r)$. . . . 108
5.4 Present-day strain spectrum, $\Delta h(k, \tau_0)$. . . . 110
5.5 Observational constraints and detectability. . . . 113

**References 116**

# List of Figures

2.1 Predictions for tensor/scalar ratio $r$ versus spectral tilt $n_s$ for minimal tuning ($Z_\eta = 0$) and for different degrees of extra fine-tuning ($Z_\eta \geq 1$). The small white circles correspond to monomial potentials (from right to left: quadratic, cubic, quartic). The thick curve enclosing all models with zero or one extra degree of fine-tuning has $n_s < 0.98$ and $r > 10^{-2}$; hatched portion has $Z_\eta = 0$ or 1 but is only accessible for polynomials of degree greater than four ($Z_{order} \geq 1$). Nine or more extra degrees of fine-tuning are required to obtain $n_s$ close to 1 or small $r$ (gray). . . . . 12

2.2 Inflationary predictions of $\Omega_{gw}(f)$ vs. $f$ with present (solid bars) and future (dashed bars) observational limits. The solid blue and green bands represent the predicted range for models with minimal tuning, or one extra degree of tuning, respectively, that satisfy the current observational bounds. The thick curve represents the lower-bound for $\Omega_{gw}$ from among the region enclosed by the black curve in Fig. 1. The purple curve is the lower bound for models with $Z_\eta < 6$. The dashed curve ($r = 10^{-3}$) is the lowest prediction among all models shown in Fig. 1. . . . . 16

3.1 The four panels correspond to the four inflationary potentials described in the text. In each panel, the solid black curve shows the primordial tensor power spectrum produced by that potential. The dashed, green, blue, and red curves show the approximate primordial tensor power spectra obtained by using the 0th, 1st, 2nd, and 3rd order approximations, respectively, for the primordial tensor extrapolation function. . . . . 34

3.2 $C_3$ is the transfer function factor that accounts for the damping of the tensor power spectrum due to tensor anisotropic stress. The factor depends on the fraction $\Omega_{free}$ of the background (critical) energy density contained in free-streaming relativistic particles. The figure plots this dependence for $0 < \Omega_{free} < 1$. . . . . 59

3.3 The black solid curve represents the present-day gravitational-wave energy spectrum, $\Omega_{gw}(f, \tau_0)$, for the inflationary model $V(\phi) = (1/2)m^2\phi^2$. The red dotted curve shows the damping effect due to (three ordinary massless species of) free-streaming neutrinos. The green dot-dashed curve shows the damping effect which arises if free-streaming particles make up fifty percent of the background energy density at the time $\tau_{BBO}$ when the modes probed by BBO/DECIGO re-enter the horizon. As shown in the figure, the particles begin free-streaming sometime before $\tau_{BBO}$, and decay sometime after $\tau_{BBO}$, but prior to electroweak symmetry breaking. Finally, the blue dashed curve shows the effect of a conformal anomaly in the early universe that slightly reduces the equation of state from $w = 0.33$ to $w = 0.31$ above the electroweak phase transition. The spectrum will also be modified on comoving scales that re-enter the horizon during the reheating epoch after inflation; but the range of scales affected by reheating is unknown. Finally, note that the correlated BBO interferometer proposal claims a sensitivity that extends beyond the bottom of the figure (down to roughly $\Omega_{gw} \sim 10^{-17}$) in the frequency range from $10^{-1}$ Hz to $10^{0}$ Hz. . . . . 64

4.1 Penrose diagrams for spatially-flat FRW universes with: (a) $0 < \epsilon < 1$, expanding; (b) $1 < \epsilon < \infty$, expanding; (c) $0 < \epsilon < 1$, contracting; (d) $1 < \epsilon < \infty$, contracting. The left edge of each diagram is the world line of a comoving observer at the origin; curved lines represent other comoving world lines and spatial hypersurfaces. The Hubble horizon is a curve connecting the $90^o$ vertex to the lightlike boundary, but the precise curve depends on $\epsilon$. For illustration, we have shown the horizon for $\epsilon = 0$ in (a, c) and for $\epsilon = 2$ in (b, d). In this paper, we focus on cases (a) and (d), in which comoving scales exit the Hubble horizon. . . . . 80

4.2 The dominant and subdominant scalar spectral indices (a) as a function of $w$ and (b) as a function of $\ln\epsilon$. Note especially the symmetry of (b). . . . . 87

5.1 Schematic of cyclic potential with numbers representing the stages described in the text. To the left of $\phi_{end}$, where the scalar kinetic energy dominates, we approximate $V$ with a Heaviside function, jumping to zero as shown by the dashed line. . . . . 106

5.2 A schematic comparison of the dimensionless strain observed today $\Delta h(k, \tau_0)$, as predicted by inflation and the cyclic model. Here $n_T$ is the inflationary tensor spectral index (a small negative number), and $\alpha \ll 1$ in the cyclic model is a small positive number. $k_r$ denotes the mode on the horizon at the start of radiation domination. . . . . 112

5.3 The present-day dimensionless strain, $\Delta h(k, \tau_0)$, predicted by the cyclic model with $T_r = 10^7$ GeV and $V_{end}^{1/4} = 10^{14}$ GeV. These parameters yield a gravity wave density four orders of magnitude below the BBN bound. Some observational bounds and (optimistic) future strain sensitivities are indicated. 115

# List of Tables

4.1 Comparison of the duality presented here with those presented by Wands [175, 155] and by Brustein *et al.* [28]. The first two columns show how the background and perturbation variables transform under each duality. The third column shows the range of $\epsilon$ to which the duality applies. The fourth column indicates the condition under which an expanding background solution and its contracting dual are both stable under small perturbations. 91

# Chapter 1

# Introduction

The universe today is a (glorious) mess: a complicated and random assortment of planets, stars, galaxies, clusters of galaxies, voids of empty space, and other structures — all interacting among themselves, and with other poorly-understood physical components that we label "dark matter" and "dark energy." By contrast, the early universe was a much simpler place — an extremely homogeneous and isotropic soup of particles, in which physical quantities such as the temperature and density were nearly identical from one place to the next, at a given time. This is seen most dramatically in the cosmic microwave background (CMB) radiation, which is observed by microwave telescopes today to have an almost perfectly thermal blackbody spectrum, with almost exactly the same temperature in every direction on the sky. The CMB provides a snapshot of the universe in its infancy, just a few hundred thousand years after the Big Bang, so it should be interpreted as a direct window onto the homogeneous and isotropic conditions at that time.

Although the early universe was *nearly* homogeneous, it was not *perfectly* homogeneous. Instead, physical quantities like the density and temperature differed from one location to the next by roughly one part in 10,000. The statistical properties of these tiny cosmological perturbations encode crucial information about the physical conditions immediately after the Big Bang. In a general sense, this thesis — and much of modern cosmology — is about how to decode this information, and determine what generated the primordial perturbations

in the first place.

The perturbations in the early universe come in three basic types: scalar, vector, and tensor [9]. The scalar perturbations (*e.g.* perturbations in the energy density and the three-dimensional Ricci curvature) are easiest to measure, and we already have detailed information about them — for example, from measurements of the tiny fluctuations in the CMB radiation from point to point on the sky. Primordial vector perturbations may or may not play a significant role in cosmic history, but we will not consider them any further in this thesis.

Instead, this thesis will focus mainly on primordial tensor perturbations — specifically, primordial gravitational waves. Although primordial tensor perturbations have not yet been observed, their existence is predicted by some of the simplest and most compelling models of the early universe (as discussed below). For this reason, a variety of experiments (in various stages of development) are hoping to detect them — either indirectly through their imprint in the polarization of the CMB [19], or directly using satellite laser interferometry [129, 87]. In this thesis, we will examine the prospects for observing primordial gravitational waves, and explore the wealth of new information that they can provide about the early universe if we can detect them.

In chapter 2, we consider inflation — the idea that a burst of accelerated expansion shortly after the Big Bang drove the early universe toward a state of extreme homogeneity, isotropy, and spatial-flatness, and also generated the tiny perturbations in the early universe by stretching quantum fluctuations to cosmological distances. In order to test inflation convincingly, we need a clear idea of its predictions. For starters, it is well known that the simplest inflationary models produce a spatially-flat universe and primordial scalar perturbations that are adiabatic, gaussian, nearly scale invariant, and correlated over distances much larger than the instantaneous Hubble length in the early universe. It is also well known that theorists can dream up (and *have* dreamed up) more baroque inflationary models that can violate any and all of these predictions. Ultimately, though, observations have thus far ruled in favor of the simplest models: every prediction mentioned above has

received striking experimental confirmation, especially over the past several years (see *e.g.* [153] and references therein). It is tempting to see a lesson here: Even though we do not yet know the correct model of inflation, the fact that theorists can dream up inflationary models that predict anything and everything does not mean that all of these predictions are on an equal footing. The predictions of those inflationary models favored by Occam's razor — the simplest and most powerfully predictive models, whose successes outnumber their arbitrary tunings and unnecessarily baroque features — should be taken more seriously.

Unnecessarily baroque features come in many varieties, and one of the goals of chapter 2 is to suggest a simple, general, model-independent, and robust technique for counting such features. We then use this technique to consider two important issues that have not yet been settled by observations — namely the values of two CMB observables: the primordial tensor-to-scalar ratio $r$, and the primordial scalar spectral index $n_s$. We will argue that inflation makes predictions for these two quantities, analogous to the predictions discussed above. That is, while it is possible to find inflationary models corresponding to any point in the $(n_s, r)$ plane, the inflationary models that should be taken most seriously (as discussed above) lie in a much more restricted region of the $(n_s, r)$ plane. This region makes two predictions. First, $n_s$ should satisfy $n_s \lesssim 0.98$. (Note that this prediction, first made in July 2005 [25], is in good agreement with the WMAP 3-year result $n_s = 0.951^{+0.015}_{-0.019}$ which appeared several months later in March 2006 [153].) Second, if $n_s$ lies in the range $0.95 \lesssim n_s \lesssim 0.98$, then $r$ should satisfy $r \gtrsim 10^{-2}$. This is encouraging, since it means that the tensor perturbations generated by inflation are likely to be detectable — both by future CMB polarization experiments such as the proposed CMBPOL satellite mission [19], and by future space-based laser-interferometer (LI) experiments such as the proposed BBO satellite mission [129].

Having argued in chapter 2 that inflationary gravitational waves should be detectable, we turn in chapter 3 to a discussion of what these gravitational waves can teach us about the early universe if we detect them. Most importantly, we will stress that we can obtain significantly more information about the early universe by *combining* CMB and LI

experiments than we can obtain from either type of experiment on its own. Considered separately, CMB and LI detections of the primordial tensor power spectrum each measure just one (very important) number: the energy density during inflation ($\sim 60$ and $\sim 20$ $e$-folds before the end of inflation, respectively). And since these are difficult experiments that only measure a single number, we would be justified in worrying that any claimed detection of the inflationary gravitational wave signal was actually either (a) a detection of a gravitational wave signal of non-inflationary origin, or (b) an unexpected foreground or noise source masquerading as a gravitational wave signal.

Now consider CMB and LI experiments in combination. If we measure the inflationary tensor amplitude on CMB scales, then — as we shall explain in chapter 3 — by combining this with information about the *scalar* spectrum on CMB scales, we can obtain a rather precise prediction for the gravitational wave amplitude on LI scales. By checking this prediction, an LI experiment can then: (a) provide much greater certainty that the signal observed on CMB and LI scales is really the gravitational wave signal from inflation; (b) check the inflationary consistency relations, which were what allowed us to predict the gravitational-wave amplitude on LI scales in the first place; and (c) measure "exotic" deviations in the tensor transfer function on LI scales, imprinted by the physical conditions when LI scales re-entered the Hubble horizon shortly after inflation, while the temperature ($T \sim 10^4$ GeV) was still well above the electroweak phase transition in the early universe.

In chapter 4, we discuss a surprising duality [95, 24] relating different cosmological models that generate precisely the same spectrum of scalar perturbations, as measured by the gauge-invariant Newtonian potential $\Phi$. When the parameter $\epsilon = \frac{3}{2}(1+w)$ is time-independent, the duality is simple to state: the fluctuations $\Phi$ generated in an expanding model characterized by $\epsilon$ precisely match the fluctuations $\widehat{\Phi}$ generated in a contracting model characterized by $\widehat{\epsilon} = 1/\epsilon$, so that every expanding model is paired with a contracting model, and vice versa. A tantalizing puzzle is to find a full generalization of this result that encorporates models with time-varying $\epsilon$, and yet retains the elegance of the $\epsilon \to 1/\epsilon$ prescription. We present some evidence that this may be possible, and some initial steps

toward this goal.

This duality is promoted from a theoretical curiosity to an issue of physical significance by virtue of its relationship to the cyclic model — a surprisingly economical idea for reproducing the successes of the standard inflationary cosmology, while radically altering the conceptual framework and sequence of events (see [159, 158, 58]). According to the cyclic model, each complete cycle of cosmic evolution includes a collision between two orbifold planes along an extra dimension, corresponding (from a four-dimensional viewpoint) to the Big Bang which begins the standard cosmological picture. Although the background geometry and the evolution of cosmological perturbations both naively become singular at this collision, the cyclic model assumes (for reasons explained in [91, 170, 163, 164, 59]) that a fully quantum-gravitational treatment resolves them in such a way that: (a) the orbifold planes pass smoothly through each other — or, equivalently, bounce off each other — so that cosmological evolution continues; and (b) perturbations generated before the bounce continue smoothly after the bounce in a way that is captured approximately (on long wavelengths and from the four-dimensional perspective) by matching $\Phi$ across the bounce. If these two assumptions are correct, then cyclic cosmology would be indistinguishable from inflationary cosmology on the basis of primordial scalar perturbations alone, as a special case of the duality discussed above. But note that the two cyclic-model assumptions are still controversial, and it will be interesting to see how this controversy resolves itself as theorists obtain better control of the physics near the bounce.

It would be even better to resolve this controversy observationally, and the cleanest resolution would come from the observation of a primordial gravitational wave signal. In chapter 5, we compute the complete present-day gravitational wave spectrum predicted by the cyclic model (from CMB scales down to microscopic scales), as a function of the parameters characterizing the scalar field potential in the cyclic model. The gravitational wave energy spectrum is very "blue" (*i.e.* it is a rapidly rising function of frequency), so almost all of its energy density is concentrated in its high-frequency tip. As a result, the strongest observational upper bound on the spectrum comes from the requirement that the

successful predictions of primordial big-bang nucleosynthesis not be spoiled by the extra energy density coming from gravitational waves. This upper bound then implies a very clean (if somewhat disappointing) result: the cyclic model gravitational-wave spectrum is simply too low — by orders of magnitude — to be detected by any known gravitational wave experiment, in any frequency band. Thus, the detection of a primordial gravitational wave signal in any frequency band would rule out the cyclic model.

# Chapter 2

# Observable predictions from inflation

## 2.1 Introduction

This chapter is based in part on the paper [25], a collaboration with Paul Steinhardt and Neil Turok. I have decided to reprint this paper verbatim in section 2.2 of this chapter (whereas the rest of the thesis is largely original, or rewritten relative to my original publications). The reason is that our paper [25] was intended to argue for a set of definite, fixed, and testable inflationary predictions — in contrast to the amorphous moving target that seems to be presented by the inflationary literature as a whole. In order for these tests to be meaningful, it is critical that the assertions in [25] be clearly recognized as *pre*-dictions rather than *post*-dictions. This is especially an especially important concern in light of the fact that, in the year between the completion of [25] (in July 2005) and the writing of this chapter (in July 2006), there have been important experimental developments, notably the WMAP three-year data release [153] in March 2006. Rather than muddying the waters by modifying or updating our paper [25] in the present chapter, I have decided instead to include the unaltered paper in section 2.2. Then, in section 2.3, I discuss the implications of various developments that have occured since the completion of [25], including the WMAP

three-year data release, and add a few additional comments.

## 2.2 Inflationary predictions for scalar and tensor fluctuations reconsidered

**Abstract:** We reconsider the predictions of inflation for the spectral index of scalar (energy density) fluctuations ($n_s$) and the tensor/scalar ratio ($r$) using a discrete, model-independent measure of the degree of fine-tuning required to obtain a given combination of ($n_s$, $r$). We find that, except for cases with numerous unnecessary degrees of fine-tuning, $n_s$ is less than 0.98, measurably different from exact Harrison-Zel'dovich. Furthermore, if $n_s \gtrsim 0.95$, in accord with current measurements, the tensor/scalar ratio satisfies $r \gtrsim 10^{-2}$, a range that should be detectable in proposed cosmic microwave background polarization experiments and direct gravitational wave searches.

Inflation predicts nearly scale-invariant spectra of primordial scalar (energy density) and tensor (gravitational wave) perturbations. What has been less clear is the precise prediction for the scalar spectral index $n_s$ and the tensor/scalar ratio $r$. In particular, is $n_s$ likely to be distinguishable from pure Harrison-Zel'dovich ($n_s = 1$)? And is $r$ likely to be large enough for the tensor perturbations to be detected ($r \gtrsim 10^{-2}$)? One approach for addressing these questions is anecdotal experience based on explicitly constructing inflaton potentials $V(\phi)$ with different combinations of ($n_s$, $r$). A more recent approach is to use the inflationary flow equations to compute $n_s$ and $r$ for random choices of the "Hubble" slow-roll parameters [55, 76, 98, 128], and plot the results as a dot-plot in the ($n_s$, $r$) plane. One problem with these methods is that the sampling does not incorporate a weight based on physical plausibility, so it does not provide a well-motivated measure of the relative likelihood across the ($n_s$, $r$)-plane. It is as if all inflaton potentials are created equal. Another problem in many of these studies is that only some of the minimal requirements for a successful inflaton potential are considered. It is simply assumed that the rest can be satisfied without reducing the attractiveness of the model. Yet, this assumption is often invalid in practice.

In this Letter, we attempt to rectify this situation by considering the complete set of inflationary conditions and introducing a discrete counting scheme for assessing the degrees

of fine-tuning required to obtain a given combination of $n_s$ and $r$. (An alternative approach is based on Bayesian model selection [166].) We find (see Fig. 1) that models with blue or slightly red tilts ($n_s > 0.98$) require significantly more degrees of fine-tuning than models with $n_s < 0.98$. This concurs with the intuitive impression obtained by trying to construct potentials by hand. More importantly, the procedure reveals valuable additional information: (1) a significant gap exists between the inflationary prediction for $n_s$ and pure Harrison-Zel'dovich ($n_s = 1$), a difference that near-future measurements should be able to resolve; and (2) if $n_s \gtrsim 0.95$, as current measurements suggest, $r$ exceeds $10^{-2}$, so tensor fluctuations should be observable in proposed cosmic microwave background (CMB) polarization and gravitational wave interferometer experiments. (Interestingly, independent approaches based on physical field-theoretic arguments have reached similar conclusions [36, 20].)

As we have emphasized, it is important to consider all the conditions necessary for inflation when assessing the degrees of fine-tuning, namely:

1. as $\phi$ evolves over some range $\Delta\phi$, the universe undergoes at least $N > 60$ $e$ folds of inflation in order to become homogeneous, isotropic, spatially flat, and monopole-free;

2. after the field evolves past this range, inflation must halt and the universe must reheat without spoiling the large-scale homogeneity and isotropy;

3. the energy density (scalar) perturbations, which we assume are generated by the quantum fluctuations of the inflaton field, must have amplitude $\sim 10^{-5}$ on scales that left the horizon $\approx 60$ $e$ folds[1] before the end of inflation, to agree with observations;

4. after inflation, the field must evolve smoothly (*i.e.*, without generating unacceptable inhomogeneities) to an analytic minimum with $V \approx 0$;

5. if the minimum is metastable, then it must be long-lived and $V$ must be bounded below.

[1] Although the text states $N_0 \approx 60$, we compute the precise value of $N_0$ model by model in our analysis. Also, we assume instantaneous reheating, but it is straightforward to show that assuming a low reheat temperature (or preheating) does not change our conclusions.

The analyticity condition is to avoid physically questionable terms of the form $|\phi|$ or $\phi^\alpha$ where $\alpha$ is non-integer. Many analyses consider only the first three conditions, but we find that the fourth condition, which is equally essential, imposes a non-linear constraint on $V$ that can significantly affect the degree of fine-tuning required to obtain a given ($n_s$, $r$). (We have stated the conditions above as if the inflaton potential is a function of a single field $\phi$; the generalization to multiple fields is straightforward.)

To quantify the degree of fine-tuning, we count the number of unnecessary features introduced during the last 60 $e$ folds of inflation to achieve a given ($n_s$, $r$). To pose the conditions in a physically motivated and model-independent way, we use the standard slow-roll parameters:

$$\epsilon \equiv (3/2)(1+w) \approx (1/2) d\ln V/dN \tag{2.1}$$

$$\eta \equiv (1/2) d\ln (V')^2/dN, \tag{2.2}$$

where $N$ is the number of $e$ folds remaining before inflation ends and a prime indicates $d/d\phi$ for an inflaton field $\phi$ canonically normalized in Einstein frame. The parameters $\epsilon$ and $\eta$ have a physical interpretation: they represent respectively the fractional rate of change of the Hubble parameter ($\propto V^{\frac{1}{2}}$) and the force on the scalar field ($\propto V'$) per inflationary $e$ fold. In all inflationary models, $\epsilon$ and $\eta$ must increase from small values ($\lesssim 1/60$) when $N \approx 60$ to values of order unity at the end of inflation ($N = 0$).

For minimally tuned models, the simplest being a monomial potential $V = \alpha\phi^n$ with integer $n$ and a single adjustable coefficient $\alpha$, both $\epsilon(N)$ and $\eta(N)$, as well as all of their derivatives (*e.g.*, $d^m\eta/dN^m$) are monotonic and have no zeroes during the last 60 $e$ folds. The range of ($n_s$, $r$) associated with these models lies in the shaded region marked "0" in Fig. 1, which has $n_s < 0.98$ and $r > 10^{-2}$. To move further outside this range requires that more zeroes of $\eta$ and its derivatives occur in the last 60 $e$ folds. The zeroes are independent in the sense that they can be added one at a time by successively adjusting parameters, as shown in Fig. 1.

Our key point is this: As exemplified by the minimally tuned models, no zeroes whatsoever are required during the last 60 $e$ folds to satisfy the five inflationary conditions. Hence, each zero added to the last 60 $e$ folds can be properly construed as representing an extra degree of fine-tuning beyond what is necessary – an extra acceleration, jerk or higher order-shift in the equation of state (for $\epsilon$) or the force (for $\eta$) artificially introduced at nearly the exact moment when the modes currently observed in the cosmic microwave background are exiting the horizon during inflation.

More specifically, the number of zeroes is a conservative (lower bound) measure of how many derivatives of $\epsilon(N)$ and $\eta(N)$ must be finely adjusted to achieve a given $(n_s, r)$. This can be seen by constructing the Taylor expansion about $N_0 \approx 60$ and comparing the higher-order terms to the lower order ones at, say, $N = 10$. As the number of zeroes increases, more higher-order terms contribute non-negligibly before inflation ends, revealing the delicate toggling of the equation of state and the force $V'$ during the last 60 $e$ folds. Proceeding deeper into the gray region in Fig. 1 (many tunings), the terms in the Taylor series grow until the series is no longer absolutely convergent.

Therefore, we propose to quantify the fine-tuning by introducing the integers $Z_{\epsilon,\eta}$ which measure the number of zeroes that $\epsilon$ and $\eta$ and their derivatives undergo within the last 60 $e$ folds of inflation. Fig. 1 is based on zeroes of $\eta$; a similar result occurs for $\epsilon$. We find that these metrics are robust methods for dividing models into those that are simple (few zeroes) and those that are highly tuned (many zeroes). (N.B. The point of our metric is not to rank a model with $Z_\eta = 20$ over a model with $Z_\eta = 1000$; the significance of a difference in $Z_\eta$ when $Z_\eta$ is large for both models is unclear. Rather, $Z_\eta$ is designed to show that both models are vastly more finely adjusted than models with $Z_\eta = 0$ or 1.)

Fig. 1 summarizes our analysis for quartic polynomial potentials $V(\phi)$ that satisfy the five inflationary conditions. The simplest models, the monomial potentials, are represented by a sequence of discrete white circles. Next, we consider more general polynomials combining terms of different order. The cases in which the coefficients all have the same sign lie on the boundary of the shaded region marked "0", along the curve connecting the white

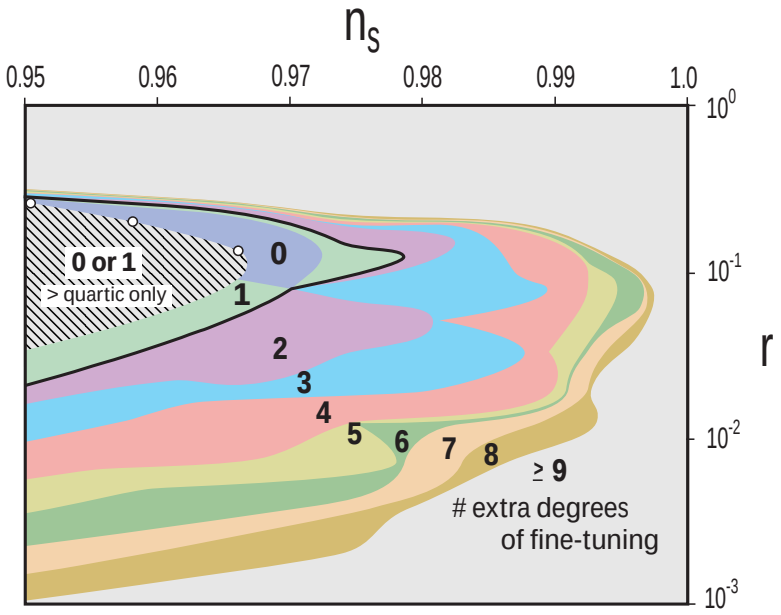



Figure 2.1: Predictions for tensor/scalar ratio $r$ versus spectral tilt $n_s$ for minimal tuning ($Z_\eta = 0$) and for different degrees of extra fine-tuning ($Z_\eta \geq 1$). The small white circles correspond to monomial potentials (from right to left: quadratic, cubic, quartic). The thick curve enclosing all models with zero or one extra degree of fine-tuning has $n_s < 0.98$ and $r > 10^{-2}$; hatched portion has $Z_\eta = 0$ or 1 but is only accessible for polynomials of degree greater than four ($Z_{order} \geq 1$). Nine or more extra degrees of fine-tuning are required to obtain $n_s$ close to 1 or small $r$ (gray).

circles. All of these models are minimally tuned ($Z_\eta = 0$) and have $n_s < 0.97$ and $r > 10\%$.

A special case occurs among models with only one degree of fine-tuning $Z_\eta = 1$: namely, models tuned so that the 60 $e$-fold mark lies very near a maximum of the potential. Simple examples include the Mexican hat potential, $V(\phi) = V_0 - \frac{1}{2}m^2\phi^2 + \lambda\phi^4$ and the axion potential, $V(\phi) = V_0(\,1 + \cos(\phi/f))$. If $\phi$ at the 60 $e$-fold mark lies close to the maximum, then $\eta$ has a zero since the force must have a maximum in these potentials. Although this kind of zero is unnecessary for inflation, it can occur naturally if the action is invariant under certain symmetries, as illustrated by the two examples above. Hence, we include this region within our thick black curve in Fig. 1. As the parameters of $V$ are further adjusted so that $\phi$ lies very close to the maximum at the 60 $e$-fold mark, the allowed range in the ($n_s$, $r$) plane expands to fill out the shaded region marked "1."

Everywhere else in the ($n_s$, $r$) plane is reached by adding sequentially more zeroes of $\eta$ and its derivatives within the last 60 $e$ folds. Increasing the zeroes introduces one or more special features in $V$ (extrema, inflection points, ...), progressively flattens the potential in the vicinity of the feature, and finely tunes $\phi$ at the 60 $e$-fold mark to lie closer and closer to it. Unlike the first ($Z_\eta = 1$) case of tuning discussed above, there is no symmetry principle that dictates any of these additional tunings. Yet, as Fig. 1 shows, many such tunings are necessary to reach low values of $r$ or high values of $n_s$.

Although Fig. 1 is based on quartic (renormalizable) polynomial potentials, a similar plot can be constructed for polynomials of arbitrary order. For polynomials of any order, there is always a wide range of parameters for which $Z_\eta = 0$ or 1, $n_s \lesssim 0.98$ and $r \gtrsim 10^{-2}$. With higher-order polynomials, it is possible to insert more bumps and jerks into the final 60 $e$ folds, even though this is not required for inflation. Introducing them for the purpose of enabling anomalous values of $n_s$ and $r$ should be included in assessing the degrees of fine-tuning. Just as the zeroes are independent and can be added one by one, the space of polynomial functions can be extended order by order. Hence, we suggest amending the degrees of tuning to be $Z_\eta + Z_{order}$, where $Z_{order}$ is the difference between the actual polynomial order and four. One regime that now becomes accessible with $Z_\eta = 0$ or 1 is

the hatched region in Fig. 1; $n_s \gtrsim 0.98$ and/or $r \lesssim 10^{-2}$ (for $n_s \gtrsim 0.95$) still require many degrees of fine-tuning.

Models with more than one field can be treated in a similar way provided the path(s) describing the last 60 $e$ folds of inflation and the passage to the potential minimum can be described by the classical equations of motion for the fields. We shall call these "deterministic." The above analysis may simply be applied to each path individually. As before, each path with many zeroes is related to paths with $Z_\eta = 0$ or 1 by a continuous fine-tuning of parameters. In some special models (like the hybrid model in [109]), the path is non-deterministic. Instead, the classical evolution reaches a critical point in the potential where quantum diffusion is needed to reach the end of inflation, as in the case $V = V_0 + \frac{1}{2}m_\phi^2\phi^2 - \frac{1}{2}m_\psi^2\psi^2 + \frac{1}{2}\gamma\phi^2\psi^2 + \ldots$ (which has a critical point at $\gamma\phi_c^2 = m_\psi^2$ and $\psi = 0$). There is effectively a discontinuous jump in $\epsilon$ and $\eta$ at the critical point, and there is not a unique procedure to relate these cases to models with $Z_\eta = 0$ or 1. As a result, counting zeroes may not be an appropriate way to judge them. We note that these cases include examples with $n_s \gtrsim 0.98$ and $r \lesssim 10^{-2}$; however, compared to the $Z_\eta = 0$ models in Fig. 1, one must add at least one extra field, an exponentially large mass hierarchy between $m_\phi$ and $m_\psi$, and, for some ($n_s$, $r$), another hierarchy between the three dimensionless quartic couplings and/or one or more higher-order couplings [14].

Our conclusion that gravitational waves should be detectable runs contrary to some claims in the literature. A common but flawed argument has been that the amplitude of the tensor power spectrum is highly uncertain because it is proportional to the fourth power of the inflationary energy scale $M_*$, whose value is poorly determined. In actuality, although gravitational waves will allow us to determine $M_*$, their detectability only depends on the tensor/scalar ratio, $r$, which does not depend on $M_*$ at all [157]. ($r \approx 16\epsilon$, so it only depends on the equation of state during inflation.)

A second argument by Lyth [112] and others makes the claim that $r$ must be small in any theory which includes quantum gravity effects. They point out that the slow-roll equations imply the relation $\Delta\phi = (m_{Pl}/8\sqrt{\pi}) \int r^{1/2} dN$, where $m_{Pl} = 1.2 \times 10^{19}$ GeV is the Planck

mass. From this relation, if $r \gtrsim .05$, then $\Delta\phi/m_{Pl}$ should exceed unity over the final 60 $e$ folds of inflation. Lyth argues that once gravitational effects are included: (a) the effective potential $V$ can only be reliably calculated over a domain $\Delta\phi < m_{Pl}$; and (b) inflation is likely to occur only over a range $\Delta\phi < m_{Pl}$. These two claims are disputed in [36, 20], and several inflationary models [97, 47, 53, 88, 89, 7] provide explicit counter-examples. These models have $\Delta\phi/m_{Pl} > 1$ and gravitational corrections are under control. They give values of $n_s$ and $r$ consistent with Fig. 1 and with current data.

Our analysis shows that forcing $\Delta\phi/m_{Pl}$ to be less than unity and maintaining a small number of zeroes requires $n_s \lesssim 0.95$, past the left boundary of Fig. 1 and outside the range favored by current data. To obtain $r < 10^{-3}$ and a small number of zeroes requires much smaller $n_s$. The only alternatives for obtaining small $r$ are to introduce many zeroes or to turn to non-deterministic models. The latter, as we have noted, typically require extra fields and tunings compared to deterministic models with $Z_\eta = 0$ or 1.

The primordial spectrum of gravitational waves in Fig. 1 is related to the observable spectrum today ($\tau_0$) by a "transfer function" $T_t(k, \tau_0)$ [22]:

$$\Omega_{gw}(k, \tau_0) \equiv \frac{1}{\rho_{cr}} \frac{d\rho_{gw}}{d \ln k} = \frac{1}{12} \frac{k^2}{a_0^2 H_0^2} T_t(k, \tau_0) \Delta_t^2(k), \tag{2.3}$$

where $\Delta_t^2(k)$ is the primordial tensor power spectrum, $\rho_{gw}$ is the gravitational-wave energy density, $\rho_{cr}$ is the critical density, and $\Omega_{gw}(k, \tau_0)$ is the ratio of the gravitational wave energy density in a log-interval about $k$ to the critical density. We have used an analytic expression for the transfer function derived in [22] which improves on the accuracy of previous calculations [169, 168] (but see also [148, 11]). The transfer function includes the redshift-suppression after horizon re-entry; the imprint of horizon re-entry itself; the possibility of dark energy with equation-of-state $w(z)$; the damping due to free-streaming relativistic particles (*e.g.*, neutrinos) in the early universe [179]; and several early-universe effects that were not considered in previous treatments.

Fig. 2 shows the inflationary predictions for $\Omega_{gw}$ as a function of frequency $f$ compared to present and future observations, assuming current limits, from [140], on non-inflationary parameters. (For a related figure with the observations shown in more detail, see Fig. 2 in

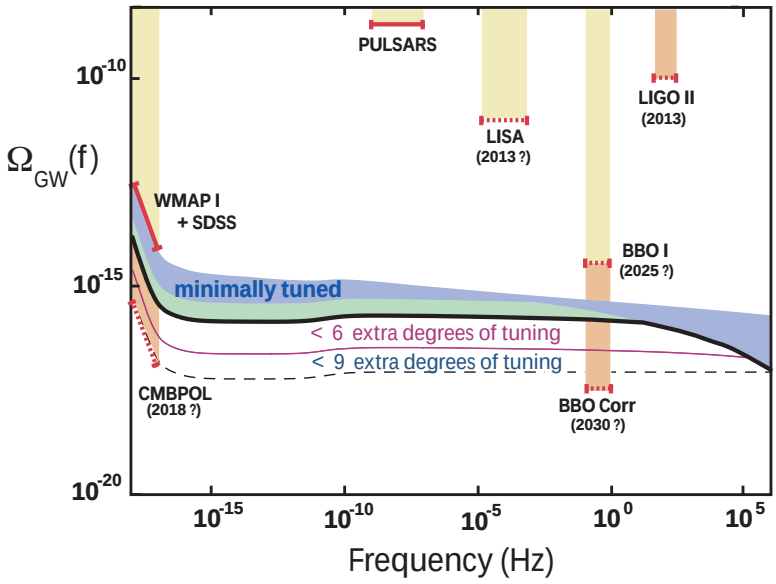



Figure 2.2: Inflationary predictions of $\Omega_{gw}(f)$ vs. $f$ with present (solid bars) and future (dashed bars) observational limits. The solid blue and green bands represent the predicted range for models with minimal tuning, or one extra degree of tuning, respectively, that satisfy the current observational bounds. The thick curve represents the lower-bound for $\Omega_{gw}$ from among the region enclosed by the black curve in Fig. 1. The purple curve is the lower bound for models with $Z_\eta < 6$. The dashed curve ($r = 10^{-3}$) is the lowest prediction among all models shown in Fig. 1.

[148].) The thick solid curve in Fig. 2 represents the lower bound among all models with minimal tuning ($Z_\eta = 0$) or one extra degree of fine-tuning ($Z_\eta = 1$). The thick dotted curve is the lower bound predicted for the entire range of models in Fig. 1; for quartic potentials, at least nine extra degrees of fine-tuning are required to go below it. Also notice the kinks near $f \sim 10^{-11}$ Hz, caused by the onset of neutrino free-streaming, as reflected in the tensor transfer function. Most importantly, the entire range of models discussed here should be accessible to future CMB polarization experiments [32, 173, 19] and space-based gravitational wave detectors, like the Big Bang Observer (BBO) [129, 171].

Hence, we find that, contrary to some suggestions in the literature, all inflationary models are not created equal. The goals of inflation do not require going beyond models with minimal or near-minimal tuning ($Z_\eta \leq 1$). (For skeptical readers who may demur from this conclusion, we pose a challenge: construct a deterministic, complete inflationary model forced by fundamental physics into a parameter region with $Z_\eta \gg 1$.) Furthermore, the minimal and near-minimal models are the most powerfully predictive in the sense that they require the fewest tunings and make the highest number of successful predictions. Based on this analysis, both a red tilt with $n_s < 0.98$ and cosmic gravitational waves with $r \gtrsim 10^{-2}$ are expected and should be detected if inflation is right. A similar analysis should be applied to cyclic models [159, 158], which must satisfy some conditions analogous to the inflationary conditions (and some not), to determine if the same range of $n_s$ is favored.

We thank L. Page and D. Spergel for discussions that inspired this project, and D. Baumann for many insightful comments. This work was supported in part by US Department of Energy grant DE-FG02-91ER40671.

## 2.3 Additional remarks

In this section, we will discuss several relevant issues that were not included in [25], either because they arose after the completion of [25], or else due to space constraints.

Let us first discuss some of the implications of the three-year WMAP data set, which was released in March 2006 (in between the completion of our paper [25] in July 2005, and

the writing of this chapter in July 2006).

When the WMAP three-year data set (alone) is fit by the minimal set of six cosmological parameters ($\Omega_m h^2$, $\Omega_b h^2$, $h$, $\tau$, $n_s$, $\sigma_8$), the marginalized value of the primordial scalar tilt ($n_s = 0.951^{+0.015}_{-0.019}$) is less than unity by roughly three sigma [153]. This appears to be the first significant evidence that the primordial scalar power spectrum is "red" ($n_s < 1$), in agreement with the predictions of [25]. Indeed, the WMAP data is consistent with (and even appears to provide marginal evidence for) the stronger prediction $n_s < 0.98$ from [25].

How seriously should we take the $n_s < 1$ result based on the three-year WMAP data set? First, it has been correctly noted [153] that if we include a seventh fitting parameter (the tensor-to-scalar ratio $r$), then the marginalized value of $n_s$ increases, seemingly decreasing the significance of the $n_s < 1$ result. But one should not attach any importance to this phenomenon for the following reason. The tensor-to-scalar ratio $r$ does not improve the fit to the three-year WMAP data set — it only contributes an extra degeneracy in parameter space, indicating that this parameter is simply too unconstrained by the current data to contribute any real information to the parameter estimations. Instead, including $r$ in the analysis actually contributes a systematic bias that spuriously increases the marginalized value of $n_s$. This is easily understood: there is a well known degeneracy in the $\{n_s, r\}$ plane (in the direction of increasing $n_s$ and increasing $r$), so that when we include $r$ as a seventh fitting parameter, we drive up the marginalized value of $n_s$ — not because of any constraint on $r$ coming from the data, but rather because we only marginalize over $r \geq 0$. But there have been two other challenges to the significance of the $n_s < 1$ result that are worth noting. The first challenge [79] argues that a more careful treatment of the power coming from unresolved microwave point sources in the WMAP data reduces the significance of the $n_s < 1$ result from $\sim 3$ sigma to $\sim 2$ sigma. The second challenge [127] argues for using the Bayesian evidence to determine wheter $n_s < 1$ is favored over $n_s = 1$. Based on the three-year WMAP data alone, they argue that the Bayesian evidence for $n_s < 1$ in the 6-parameter fit is not sufficient to favor it over a five-parameter fit with $n_s = 1$. But the Bayesian evidence for $n_s < 1$ does become significant when other complementary data sets

are also included in the analysis. Taken together, these various analyses seem to suggest that the significance of the $n_s < 1$ result should be currently regarded as marginal, but non-negligible — somewhere in the 2 to 3 sigma range.

Let us add a few words about hybrid inflation and brane inflation. In [25] we note that, since our arguments only apply to inflationary models in which the vacuum expectation value of the inflaton field rolls classically to the bottom of its potential well, they do not apply hybrid inflation models in which the inflaton field makes a quantum jump to the bottom of its potential well after rolling past a critical point on its potential. Until relatively recently, most hybrid inflation models in the literature (those based on traditional field theory) predicted a significantly "blue" scalar spectral tilt ($n_s > 1$) — and, as we have seen, such models are now significantly disfavored by the third-year WMAP data. But more recently, there has been much discussion in the literature of various brane inflation models in string theory (of both hybrid [80] and non-hybrid [51] type) which appear to predict $n_s < 1$. In these models, one must stabilize various moduli (notably the overall volume modulus), and the stabilization mechanisms generate important mass terms (as well as higher self-coupings) in the inflaton potential. If one neglects these stabilization-induced terms in the inflaton potential, then these brane inflation models seem to predict $n_s \approx 0.97$ and negligibly small $r$. As a result, brane inflation has been widely advertised by some as an example of a well-motivated model that agrees with the current data, yet predicts undetectably small $r$. But this is wrong — or at least premature. Indeed, recent work [15, 13] treating this issue more carefully has shown that one gets completely incorrect answers by neglecting the stabilization-induced corrections to the inflaton potential. Indeed, properly including these corrections generally yields a potential that is much less flat (and hence produces a much larger $r$). A computation of the brane-inflation potential, including all known corrections, can be done for a tractable subset of these brane inflation models, and we eagerly anticipate more realistic predictions for $n_s$ and $r$ from these models [13].

Based on the arguments in [25], the three-year WMAP data set should be regarded as providing encouragement for future attempts to detect the inflationary gravitational-wave

spectrum in the following sense. The measured range for the spectral tilt ($n_s = 0.951^{+0.015}_{-0.019}$) agrees surprisingly well with the range predicted by the arguments in [25]. But these same arguments also imply that, if $n_s = 0.951^{+0.015}_{-0.019}$, then then $r$ should be $\gtrsim 10^{-2}$ (see Figure 2.1), so the primordial tensor spectrum should be detectable by future CMB and laser-interferometer satellite experiments.

The range $r \gtrsim 10^{-2}$, suggested by our analysis, corresponds to a rather high energy scale $M_* \sim 10^{16}$ GeV for inflation [105]. Here $M_*$ denotes the fourth root of the energy density as CMB scales leave the horizon during inflation. But it is worth mentioning that there are at least two possible clues about how this this energy scale might fit with our other ideas about physics beyond the standard model. The first point is that the scale of inflation might be related to the physics of grand unification, since it is well known that the three standard-model gauge couplings roughly unify at an energy $\sim 10^{16}$ GeV [67]. Indeed, in the minimal supersymmetric extension of the standard model, they *precisely* unify (to within current experimental/theoretical uncertainties) at $2 \times 10^{16}$ GeV [48]. A second (closely related) point is that the energy scale of inflation might be connected to the physics that generates neutrino mass. The left-handed neutrinos in the standard model are observed to have masses $m_\nu \sim 0.1$ eV. Currently, the leading theoretical explanation for these tiny masses is the "see-saw mechanism" [118, 185, 66, 68], according to which a heavy right-handed neutrino $N$ gives the ordinary left-handed neutrino $\nu$ a see-saw mass $m_\nu \sim \Lambda_{EW}^2/\Lambda_{seesaw}$, where $\Lambda_{EW}$ is an electroweak scale, and $\Lambda_{seesaw}$ is the see-saw scale. If we take a characteristic value $\Lambda_{EW} \sim 10^3$ GeV for the electroweak scale, then we automatically obtain $\Lambda_{seesaw} \sim 10^{16}$ GeV for the see-saw scale. Alternatively, it may be more correct [29] to take $\Lambda_{EW} \sim 174$ GeV, corresponding to the vacuum expectation value of the Higgs, in which case we infer $\Lambda_{seesaw} \sim 3 \times 10^{14}$ GeV. At first sight, this scale might seem low (relative to $10^{16}$ GeV); but we should keep in mind that the inflationary energy scale $M_*$ is generally distinct from the inflaton mass $m$: for example in $V = \frac{1}{2}m^2\phi^2$ inflation, the energy scale is $M_* \approx 10^{16}$ GeV, but the inflaton mass is $m \approx 2 \times 10^{13}$ GeV. In fact, it is interesting to note that this value for $m$ agrees rather well with the second (lower) estimate for $\Lambda_{seesaw}$ given

above, especially considering the relatively large uncertainty in this estimate. Indeed, this fact was noticed, and used to build a rather economical model of inflaton in [124].

In this chapter we have argued that the gravitational waves from inflation are likely to have a detectably large amplitude. In the next chapter, we will explore what these gravitational waves can teach us about the early universe if we successfully measure them.

# Chapter 3

# What can inflationary gravitational waves teach us?

This chapter is based on work done in collaboration with Paul Steinhardt, part of which has already appeared as [22].

## 3.1 Introduction

Inflation [74, 107, 4] generates tensor perturbations (gravitational waves) with a nearly scale-invariant primordial power spectrum [70, 155]. This spectrum extends across a broad range of scales, from laboratory to cosmological wavelengths, and there are two different approaches to detecting it. First, we can look for a gravitational-wave imprint in the cosmic microwave background (CMB) anisotropy — either in the temperature anisotropy [138, 49, 137, 60, 1, 154] or, more promisingly, in the "B-mode" polarization anisotropy [133, 84, 85, 143, 186]. This CMB approach is being actively pursued by a diverse community of experimental groups [19]. Second, we can use space-based laser-interferometer (LI) experiments to search for inflationary gravitational waves directly [168, 148, 139]. This direct-detection approach has received increased attention over the past few years, since it has been realized that space-based laser interferometers operating in the frequency range

0.1 Hz $< f <$ 1 Hz might realistically achieve the sensitivity and foreground subtraction necessary to reach the inflationary signal [129]. In particular, two satellite LI missions have been proposed — NASA's "Big Bang Observer" (BBO) [129] and the Japanese "Deci-hertz Interferometer Gravitational Wave Observatory" (DECIGO) [146, 87] — and are currently being investigated [171, 42, 145, 101, 39], although they would not fly for at least another decade.

The goal of this chapter is to elucidate the relationship between these two different techniques — the CMB approach and the LI approach. Most importantly, we will show that by combining CMB and LI experiments, we have the *realistic* opportunity to learn things about the extremely early universe that are qualitatively different from anything that can be learned by either approach on its own — or by any other known technique, for that matter.

Let us start with a brief overview of the ideas fleshed out in this chapter. Inflation generates a primordial tensor power spectrum $\Delta_t^2(k, \tau_i)$. Here $k$ is a comoving wavenumber; and $\tau_i$ refers to some very early time, shortly after the end of inflation, when all wavenumbers of interest are still outside the Hubble horizon. CMB and LI experiments probe the primordial tensor spectrum in two widely separated wavebands: CMB experiments probe a narrow band near the comoving wavenumber $k_{CMB} = 0.002\,\mathrm{Mpc}^{-1}$, while LI experiments probe a narrow band near the much higher comoving wavenumber $k_{LI} = 2 \times 10^{14}\,\mathrm{Mpc}^{-1} = 10^{17} k_{CMB}$. To clarify the relationship between these two measurements, it is convenient to introduce the function $E_t(k, k_*)$, which we will call the "tensor extrapolation function." It is defined through the equation

$$\Delta_t^2(k, \tau_i) = E_t(k, k_*) \Delta_t^2(k_*, \tau_i) \tag{3.1}$$

and relates the primordial tensor spectrum evaluated at a general wavenumber $k$ to the primordial tensor spectrum evaluated at a reference wavenumber $k_*$.

Of course, neither LI nor CMB experiments directly measure the spectrum at $\tau = \tau_i$. Instead, CMB experiments measure $\Delta_t^2(k, \tau_{CMB})$ at the later time $\tau_{CMB}$ of hydrogen recombination, several hundred thousand years after the big bang; and LI experiments measure

$\Delta_t^2(k,\tau_0)$ at the present time $\tau_0$. To connect the primordial power spectrum $\Delta_t^2(k,\tau_i)$ to the spectrum $\Delta_t^2(k,\tau)$ measured at some later time $\tau$, we must propagate forward in time using the tensor transfer function $T_t(k,\tau)$, defined through the equation

$$\Delta_t^2(k,\tau) = T_t(k,\tau)\Delta_t^2(k,\tau_i). \tag{3.2}$$

In other words, the transfer function $T_t(k,\tau)$ propagates the spectrum in $\tau$ (with $k$ held fixed), while the extrapolation function $E_t(k,k_*)$ propagates the spectrum in $k$ (with $\tau = \tau_i$ held fixed). Combining equations (3.1) and (3.2), we obtain

$$\Delta_t^2(k,\tau) = T_t(k,\tau)E_t(k,k_*)\Delta_t^2(k_*,\tau_i). \tag{3.3}$$

Thus, if CMB measurements manage to infer the primordial tensor power spectrum at the reference wavenumber $k_*$, then (3.3) tells us how to infer the tensor power spectrum at a different wavenumber $k$ and a later time $\tau$. In the remainder of this chapter, we unpack this equation: section 2 derives and examines the extrapolation function $E_t$, section 3 derives and examines the transfer function $T_t$, and section 4 explores the science possibilities and opportunities that arise from combining CMB and LI experiments.

Now let us highlight the key ideas in each section. In section 2, we write the tensor extrapolation function $E_t(k,k_*)$ as an infinite product $E_t(k,k_*) = \prod_{m=1}^{\infty} A_m$ where the individual factors $A_m$ rapidly approach unity as $m$ increases, and the product rapidly converges. Then we use the hierarchy of inflationary "consistency relations" to determine the factors $A_m$, order by order, in terms of CMB observables. Thus, if we measure $r$ with the CMB, then we can determine the leading-order contribution to $A_1$; and if we also measure $n_s$ with the CMB, then we can also determine the next-to-leading order (NLO) contribution to $A_1$ and the leading-order contribution to $A_2$; and if we also measure (or sufficiently tightly constrain) $dn_s/d\ln k$, then we can also determine the NLO contributions to $A_2$ and the leading-order contributions to $A_3$; and so forth. This means that, if we can measure two observables $\{r, n_s\}$ on CMB scales, then we can replace the exact extrapolation function $E_t(k,k_*)$ by an approximate (second-order) extrapolation function $E_t^{(2)}(k,k_*)$ that only depends on these two observables. And if we can measure (or sufficiently constrain) three

observables $\{r, n_s, dn_s/d\ln k\}$ on CMB scales, then we can replace the exact extrapolation function $E_t(k, k_*)$ by an approximate (3rd-order) extrapolation function $E_t^{(3)}(k, k_*)$ that only depends on these three observables. Having obtained these approximate extrapolation functions, we must then consider the two types of extrapolation uncertainty that arise. The first type comes from replacing the exact extrapolation function ($E_t$) by the approximate extrapolation function ($E_t^{(2)}$ or $E_t^{(3)}$). The second type is induced by the error bars on the CMB observables $\{r, n_s, dn_s/d\ln k\}$ themselves. We study how both types of extrapolation uncertainty depend on the measured values of the CMB observables and their corresponding error bars.

We think that the following point is rather important from a conceptual standpoint. Note that we have used the tensor extraplation function to "repackage" the hierarchy of inflationary consistency relations into a form where they relate measurements made at two different wavenumbers. Previously authors have considered the potential relevance and testability of the first consistency relation; but have disregarded the higher consistency relations as impractical and untestable. For example, Cortes and Liddle [40] have given the most lucid discussion of the hierarchy of inflationary consistency relations, and yet they discount their discussion as "an academic one, as there is little prospect of testing any of these relations beyond the first, and even *it* is likely to prove challenging." And, indeed, if one only thinks about experiments that probe a relatively narrow band of wavenumbers, then the higher consistency relations are irrelevant. But since we are interested in combining CMB and LI experiments, which are widely separated in wavenumber, we will see that the higher consistency relations play a non-negligible role.

The key point in section 3 has to do with an important difference between the tensor transfer function on CMB scales ($k \sim k_*$) and LI scales ($k \sim 10^{17} k_*$). To see the difference, let us start by considering a wavenumber $k$ that re-enters the horizon at time $\tau_k$ after inflation. At some later time $\hat{\tau}$ after horizon re-entry, how does $T_t(k, \hat{\tau})$ depend on the physical conditions at earlier times $\tau < \hat{\tau}$? In other words, what does the transfer function at time $\hat{\tau}$ remember about the past history of the universe? The answer is that $T_t(k, \hat{\tau})$ is

completely insensitive to the physical conditions *before* horizon re-entry ($\tau_i < \tau < \tau_k$), but quite sensitive to the physical conditions *after* horizon re-entry ($\tau_k < \tau < \hat{\tau}$) and especially *during* horizon re-entry ($\tau \approx \tau_k$). This means that, since CMB experiments probe long-wavelength modes that re-enter the horizon well after big-bang nucleosynthesis (BBN), the corresponding transfer function is insensitive to the physical conditions prior to BBN. As a result, $T_t(k, \tau)$ on CMB scales is well understood: it does not depend on the details of uncertain high-energy phenomena such as the decay of the inflaton and the "reheating" of the universe after inflation; it has recently received detailed analytical treatment in [11] and [181]; and it can be reliably computed by numerical codes such as CMBFAST [142, 187] and CAMB [104].

By contrast, LI experiments probe short-wavelength modes that re-enter the horizon much earlier than CMB modes. For example, BBO is designed to probe gravitational waves that oscillate with frequencies between 0.1 and 1 Hz today. These modes re-enter the horizon extremely early in cosmic history, when the universe is at a temperature $T_{BBO} \sim 10^4$ TeV — long before the electroweak phase transition (which is expected to take place around $T \sim$ 300 GeV). Note that $T_{BBO}$ is about four orders of magnitude above the electroweak scale, which marks the current frontier in high-energy physics, and the threshold above which our knowledge of both fundamental physics and cosmology becomes highly speculative. This means that, in contrast to the transfer function on CMB scales, the transfer function on LI scales at late times should be expected to retain memory of a highly uncertain epoch in the early universe (when the temperature was in the range 1 TeV $\lesssim T \lesssim 10^4$ TeV).

Thus, in order to properly interpret LI measurements of the inflationary gravitational wave spectrum, we must analyze the various early-universe phenomena that can imprint themselves in the tensor transfer function on LI scales. This job is made much easier by the fact that gravitational waves are extremely linear, and couple to all different types of matter in a universal way. We identify a number of relevant early-universe phenomena, and derive a new and improved tensor transfer function designed to incorporate their effects. Some of these effects can be quite significant and, given our state of ignorance about physics

above $T \sim 1$ TeV, it would be dangerous to ignore them! It is worth emphasizing two other technical but important points about our new transfer function. First, since we formulate it in terms of redshift $z$ rather than cosmic time $t$ or conformal time $\tau$, it is easy to incorporate non-standard background equation-of-state parameters $w(z)$ (for example due to dark energy at late times, or various possible violations of $w = 1/3$ during the radiation era); and also easy to avoid the spurious suppression or enhancement factors that have arisen in some previous derivations (for example, from the incorrect treatment of the transition from matter domination to dark-energy domination). Second, our transfer function includes an important overall suppression factor of 1/2 that has been neglected in previous transfer functions. The factor of 1/2 arises because, at fixed time $\tau_0$, the transfer function on LI scales is a rapidly oscillating function of $k$, proportional to $\sin^2(k\tau_0)$. Since the frequency difference between neighboring peaks in this oscillation ($\Delta f \sim 10^{-17} Hz$) is much too small to be observed by any known technique [6], we must replace the $\sin^2(k\tau_0)$ factor by its wavenumber-averaged value 1/2.

The key point in section 4 is that, by combining CMB and LI experiments, we can test an important null hypothesis — namely, the joint hypothesis that:

- (a) the correct extrapolation function is the fiducial one derived in section 2 from the inflationary consistency relations; and
- (b) the correct transfer function on LI scales is the fiducial one produced by the standard model of particle physics, with no additional corrections coming from "exotic" high-energy effects.

If we assume this null hypothesis, then we can use the fiducial extrapolation function and the fiducial transfer function to convert the CMB observables $\{r, n_s, dn_s/d\ln k\}$ (and their corresponding uncertainties) into a predicted range for the gravitational-wave amplitude on LI scales. We show that if the CMB observables and their corresponding experimental uncertainties turn out to lie in a certain range (that is experimentally allowed [153], realistic [173], and even likely [25]), then a futuristic (but *realistic*) LI experiment like the "standard"

BBO mission proposal being considered by NASA could perform a high-significance test of whether or not the gravitational wave amplitude lies within the predicted range on LI scales. On one hand, if the amplitude is found to lie outside the range predicted by the null hypothesis, it would indicate either (a) the existence of significant "exotic" high-energy effects in the tensor transfer function, or (b) the violation of the inflationary consistency relations. Either interpretation would provide an interesting clue about exotic high-energy physics, but it might require complementary theoretical or experimental progress to sort out which interpretation was correct. On the other hand, if the amplitude is indeed found to lie within the range predicted by the null hypothesis, this would provide strong support for the inflationary consistency conditions, and arguably clinch the case for inflation.

Previously, two strategies for testing the consistency condition have been discussed in the literature: one strategy uses the CMB alone [150], while the other uses CMB and LI experiments together [149]. Unfortunately, as [149] correctly concludes, neither of these two strategies is feasible — that is, neither can successfully test the consistency relation, even with the most optimistic assumptions for the future CMBPOL and BBO satellite missions. In particular, see [149] — especially section 3 and the right-hand panel in Figure 2 of that paper. (Ignore the dotted curve in that figure, since it corresponds to an unrealistic LI experiment.)

As we explain in section 4, our strategy of

- (a) using the extrapolation function to repackage the hierarchy of inflationary consistency relations, and
- (b) combining CMB and LI experiments to test the validity of the extrapolation function, and hence the validity of the inflationary consistency relations,

is completely distinct from the previous two strategies [150, 149], both logically and practically, even if it shares superficial similarities with them. We will argue that our strategy is more feasible than the previous two proposals — not because we make more optimistic assumptions about future CMB and LI experiments, but rather because we are propos-

ing a new approach. Briefly, the relationship between the three strategies may be roughly summarized as follows:

- Song & Knox [150] suggest comparing the ratio $r$ on CMB scales to the tensor tilt $n_t$ on CMB scales, in order to test the first consistency relation $r = -8n_t$. Unfortunately, CMB experiments will only probe the inflationary tensor spectrum over a narrow waveband, so they will yield a very poor constraint on the *local* tensor tilt $n_t$ on CMB scales.

- Smith, Peiris and Cooray [149] suggest comparing the ratio $r$ on CMB scales to the tensor tilt $n_t$ *on LI scales*, in order to test the first consistency relation $r = -8n_t$. Unfortunately — as with the CMB — LI experiments will only probe the inflationary tensor spectrum over a narrow waveband, so they will yield a very poor constraint on the *local* tensor tilt $n_t$ on LI scales (see [145]). Another difficulty is that the equation $r = -8n_t$ is only appropriate when $r$ and $n_t$ are measured at the *same* wavenumber.

- We suggest comparing the ratio $r$ on CMB scales to the band-averaged tensor *amplitude on LI scales*. In other words, instead of attempting to measure $n_t$ locally on either CMB or LI scales, we (roughly speaking) probe the "effective" tensor tilt by measuring the overall drop in tensor amplitude between two different wavenumbers — thus taking advantage of the long lever arm provided by the wide separation between CMB and LI scales. To interpret our test more precisely, we should *not* think about it in terms of the usual consistency relation, $r = -8n_t$, since this is only appropriate for comparing two measurements at the same wavenumber; instead we must recast the usual consistency relations in the tensor extrapolation function, which provides the proper formalism for comparing measurements at two different wavenumbers.

Simply put, our approach is more feasible because CMB and LI experiments both measure the local tensor amplitude much better than the local tensor slope. Furthermore our strategy is designed to take advantage of the large separation between CMB and LI scales, while this large separation works to the disadvantage of the second strategy [149].

## 3.2 The tensor extrapolation function $E_t(k, k_*)$

In this section we introduce the tensor extrapolation function, and note how it conveniently repackages the hierarchy of slow-roll consistency equations. We use the tensor extrapolation function to predict the primordial tensor amplitude (and its uncertainty) on LI scales, as a function of the CMB observables $\{r, n_s, dn_s/d\ln k\}$ and their respective uncertainties.

In what follows, $\Delta_s^2(k, \tau_i)$ is the primordial scalar power spectrum, $\Delta_t^2(k, \tau_i)$ is the primordial tensor power spectrum, and we define the tensor-to-scalar ratio

$$r(k) \equiv \frac{\Delta_t^2(k, \tau_i)}{\Delta_s^2(k, \tau_i)} \qquad (\tilde{r} \equiv \frac{1}{8} r). \tag{3.4}$$

Note that the tensor-to-scalar ratio $r$ agrees with the WMAP conventions, and we have also defined the "reduced" tensor-to-scalar ratio $\tilde{r}$ for later convenience, since it significantly simplifies a number of the formulae below. It will also be very convenient to define the scalar parameters $\alpha_s^{(m)}$ and the tensor parameters $\alpha_t^{(m)}$:

$$\alpha_s^{(m)}(k) \equiv \frac{d^m[\ln \Delta_s^2(k, \tau_i)]}{d[\ln k]^m}, \tag{3.5a}$$

$$\alpha_t^{(m)}(k) \equiv \frac{d^m[\ln \Delta_t^2(k, \tau_i)]}{d[\ln k]^m}, \tag{3.5b}$$

Note that $\alpha_s^{(1)}$ is just the ordinary scalar tilt $(n_s - 1)$; and $\alpha_t^{(1)}$ is the ordinary tensor tilt $n_t$; and $\alpha_s^{(2)}$ is the running of the scalar tilt; and $\alpha_t^{(2)}$ is the running of the tensor tilt; and $\alpha_s^{(3)}$ is the running-of-the-running of the scalar tilt; and $\alpha_t^{(3)}$ is the running-of-the-running of the tensor tilt; and so forth. In terms of these parameters, we can Taylor expand the primordial scalar and tensor spectra as follows:

$$\ln[\Delta_s^2(k, \tau_i)] = \ln[\Delta_s^2(k_*, \tau_i)] + \sum_{m=1}^{\infty} \frac{\alpha_s^{(m)}(k_*)}{m!} \left(\ln \frac{k}{k_*}\right)^m, \tag{3.6a}$$

$$\ln[\Delta_t^2(k, \tau_i)] = \ln[\Delta_t^2(k_*, \tau_i)] + \sum_{m=1}^{\infty} \frac{\alpha_t^{(m)}(k_*)}{m!} \left(\ln \frac{k}{k_*}\right)^m. \tag{3.6b}$$

Then, by exponentiating (3.6) we obtain

$$\Delta_s^2(k, \tau_i) = E_s(k, k_*) \Delta_s^2(k_*, \tau_i), \tag{3.7a}$$

$$\Delta_t^2(k, \tau_i) = E_t(k, k_*) \Delta_t^2(k_*, \tau_i), \tag{3.7b}$$

where we will call

$$E_s(k,k_*) = \prod_{m=1}^{\infty} \exp\left[\frac{\alpha_s^{(m)}(k_*)}{m!}K^m\right] \tag{3.8a}$$

the "scalar extrapolation function," and

$$E_t(k,k_*) = \prod_{m=1}^{\infty} \exp\left[\frac{\alpha_t^{(m)}(k_*)}{m!}K^m\right] \tag{3.8b}$$

the "tensor extrapolation function," and for later convenience we have used the notation

$$K \equiv \ln\frac{k}{k_*}. \tag{3.9}$$

For reference, note that there is a separation $K \approx 36$ between the wavenumber $k = 2 \times 10^{14}$ Mpc$^{-1}$ (which corresponds to a present-day frequency $f \approx 0.3$ Hz in the middle of the BBO frequency band) and the wavenumber $k_* = 0.05$ Mpc$^{-1}$, (which is a common reference wavenumber at which parameters are evaluated in combined CMB/large-scale-structure measurements *e.g.* [140, 141]).

Now we want to use the hierarchy of inflationary consistency relations to put the tensor extrapolation function in a more useful form. For a lucid discussion of the hierarchy of consistency relations, see the recent paper by Cortes and Liddle [40]. Here we add a simple but important interpretational point: By making use of the first $n$ consistency relations, we can always recast the $n$th consistency relation as an expression for the $n$th tensor parameter $\alpha_t^{(n)}$ in terms of the scalar parameters $\alpha_s^{(i)}$ as well as the tensor-to-scalar ratio $r$ [or, equivalently, the "reduced" tensor-to-scalar ratio $\tilde{r}$ defined in equation (3.4)]. In particular, if we want to infer the $n$th tensor parameter $\alpha_t^{(m)}$ at leading order in the slow-roll approximation, it is enough to know $r$ along with the first $n-1$ scalar parameters $\alpha_s^{(1)}, \ldots, \alpha_s^{(n-1)}$; and if we want to infer the NLO contribution to the $n$th tensor parameter $\alpha_t^{(n)}$, it is enough to know $r$ along with the first $n$ scalar parameters $\alpha_s^{(1)}, \ldots, \alpha_s^{(n)}$; and so forth. Let us present the most direct and elegant route to deriving the consistency relations in precisely this form.

We start by writing the first consistency equation as

$$\alpha_t^{(1)} = -\tilde{r} + \tilde{r}\Big[\alpha_s^{(1)} + \frac{1}{2}\tilde{r}\Big] \tag{3.10}$$

(see *e.g.* [40] for a derivation). The right-hand side of (3.10) is valid up to next-to-leading order (NLO) in the slow-roll approximation: the leading order term is $-\tilde{r}$ and the NLO terms are inside the square brackets. Now it is easy to iteratively derive the subsequent consistency relations in the hierarchy. First note that the exact relations

$$\frac{d\tilde{r}}{d\ln k} = \tilde{r}(\alpha_t^{(1)} - \alpha_s^{(1)}) \tag{3.11a}$$

$$\frac{d\alpha_s^{(m)}}{d\ln k} = \alpha_s^{(m+1)} \tag{3.11b}$$

$$\frac{d\alpha_t^{(m)}}{d\ln k} = \alpha_t^{(m+1)} \tag{3.11c}$$

follow immediately from the definitions (3.4) and (3.5) of $r$, $\alpha_s^{(m)}$, and $\alpha_t^{(m)}$. Thus, to derive the $n$th consistency relation, we can first use the relations (3.11) to take $d/d\ln k$ of the $(n-1)$th consistency relation, and then use the first consistency relation (3.10) to eliminate $\alpha_t^{(1)}$ from the right-hand side of the resulting equation. By following this procedure, we will obtain a version of the $n$th consistency relation that is valid up to the same order as the first consistency relation. [In other words, since the first relation (3.10) was valid up to NLO, then the $n$th relation will be as well.] For example, the next few consistency relations in the hierarchy are

$$\alpha_t^{(2)} = \tilde{r}\Big(\alpha_s^{(1)} + \tilde{r}\Big) + \tilde{r}\left[\alpha_s^{(2)} - \alpha_s^{(1)\,2} - 3\alpha_s^{(1)}\tilde{r} - \frac{3}{2}\tilde{r}^2\right], \tag{3.12a}$$

$$\begin{aligned}\alpha_t^{(3)} = {} & \tilde{r}\Big(\alpha_s^{(2)} - \alpha_s^{(1)\,2} - 3\alpha_s^{(1)}\tilde{r} - 2\tilde{r}^2\Big) + \\ & + \tilde{r}\left[\alpha_s^{(3)} - \alpha_s^{(2)}(3\alpha_s^{(1)} + 4\tilde{r}) + \alpha_s^{(1)\,3} + 8\alpha_s^{(1)\,2}\tilde{r} + 13\alpha_s^{(1)}\tilde{r}^2 + \frac{11}{2}\tilde{r}^3\right],\end{aligned} \tag{3.12b}$$

$$\begin{aligned}\alpha_t^{(4)} = {} & \tilde{r}\Big(\alpha_s^{(3)} - \alpha_s^{(2)}(3\alpha_s^{(1)} + 4\tilde{r}) + \alpha_s^{(1)\,3} + 7\alpha_s^{(1)\,2}\tilde{r} + 12\alpha_s^{(1)}\tilde{r}^2 + 6\tilde{r}^3\Big) + \\ & + \tilde{r}\left[\alpha_s^{(4)} - \alpha_s^{(3)}(4\alpha_s^{(1)} + 5\tilde{r}) + \alpha_s^{(2)}(6\alpha_s^{(1)\,2} + 28\alpha_s^{(1)}\tilde{r} + \frac{43}{2}\tilde{r}^2)\right. \\ & \left. - 3\alpha_s^{(2)\,2} - \alpha_s^{(1)\,4} - 18\alpha_s^{(1)\,3}\tilde{r} - \frac{123}{2}\alpha_s^{(1)\,2}\tilde{r}^2 - 70\alpha_s^{(1)}\tilde{r}^3 - 25\tilde{r}^4\right],\end{aligned} \tag{3.12c}$$

where, as before, the NLO terms are inside the square brackets. Note that these results illustrate the structure pointed out above: If we want to compute $\alpha_t^{(n)}$ at leading order, it is enough to know $r$ as well as the first $n-1$ scalar parameters $\alpha_s^{(1)}, \ldots, \alpha_s^{(n-1)}$. And if we also also know $\alpha_s^{(n)}$, then we can also infer the NLO contribution to $\alpha_t^{(n)}$.

In particular, this means that the CMB observables $\{r, n_s\}$ are enough to determine the leading-order and NLO terms in $\alpha_t^{(1)}$; and the leading-order terms in $\alpha_t^{(2)}$. Futhermore, if we can put sufficiently tight constraints on $dn_s/d\ln k$ (where "sufficiently" is clarified below), then we can also determine the NLO contribution to $\alpha_t^{(2)}$, and the leading-order contribution to $\alpha_t^{(3)}$. Based on this observation, let us approximate the primordial tensor power $\Delta_t^2(k, \tau_i)$ on LI scales as follows. First, if $dn_s/d\ln k$ is insufficiently tightly constrained, then we will use the expression

$$\Delta_t^2(k, \tau_i) \approx E_t^{(2)}(k, k_*)\Delta_t^2(k_*, \tau_i), \tag{3.13}$$

where

$$E_t^{(2)}(k, k_*) = \exp\Big[\frac{\alpha_t^{(1)}(k_*)}{1!}K\Big]\exp\Big[\frac{\alpha_t^{(2)}(k_*)}{2!}K^2\Big] \tag{3.14}$$

is the second-order approximation of the extrapolation function, with $\alpha_t^{(1)}$ given by Eq. (3.10), and $\alpha_t^{(2)}$ given by the leading-order terms on the right-hand side of Eq. (3.12a). On the other hand, if $dn_s/d\ln k$ is sufficiently tightly constrained, then we will use the expression

$$\Delta_t^2(k, \tau_i) \approx E_t^{(3)}(k, k_*)\Delta_t^2(k_*, \tau_i), \tag{3.15}$$

where

$$E_t^{(3)}(k, k_*) = \exp\Big[\frac{\alpha_t^{(1)}(k_*)}{1!}K\Big]\exp\Big[\frac{\alpha_t^{(2)}(k_*)}{2!}K^2\Big]\exp\Big[\frac{\alpha_t^{(3)}(k_*)}{3!}K^3\Big] \tag{3.16}$$

is the third-order approximation of the extrapolation function, with $\alpha_t^{(1)}$ given by Eq. (3.10), $\alpha_t^{(2)}$ given by Eq. (3.12a), and $\alpha_t^{(3)}$ given by the leading-order terms on the right-hand side of Eq. (3.12b). Note that the 2nd-order extrapolation function $E_t^{(2)}$ only depends on the CMB observables $\{r, n_s\}$, while the 3rd-order extrapolation function $E_t^{(3)}$ only depends on the CMB observables $\{r, n_s, dn_s/d\ln k\}$. Thus, we have obtained a prediction for the amplitude of the primordial power spectrum at wavenumber $k$ as a function of the observables $\{r, n_s, dn_s/d\ln k\}$ at reference wavenumber $k_*$.

We must now analyze the two qualitatively different uncertainties in this prediction. The first uncertainty comes from replacing the exact extrapolation function $E_t(k, k_*)$ — either with the second-order approximation $E_t^{(2)}(k, k_*)$ on the right-hand side of (3.13),

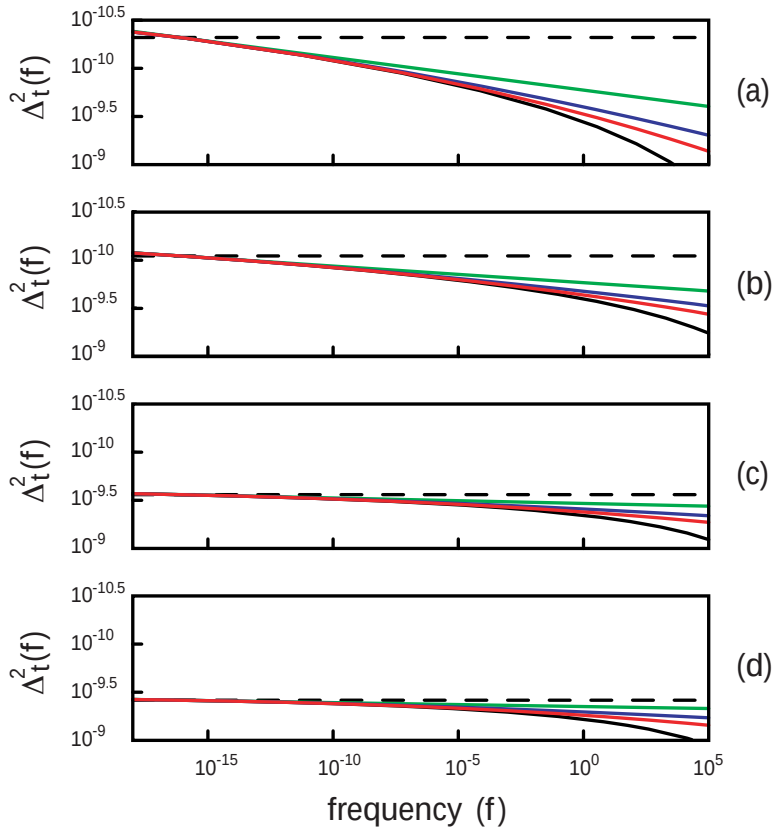



Figure 3.1: The four panels correspond to the four inflationary potentials described in the text. In each panel, the solid black curve shows the primordial tensor power spectrum produced by that potential. The dashed, green, blue, and red curves show the approximate primordial tensor power spectra obtained by using the 0th, 1st, 2nd, and 3rd order approximations, respectively, for the primordial tensor extrapolation function.

or with the third-order approximation $E_t^{(3)}(k, k_*)$ on the right-hand side of (3.15). The second uncertainty comes from the fact that we cannot plug in exact values for the CMB observables on the right-hand side of (3.13) or (3.15), since these CMB observables will have non-vanishing error bars themselves. Let us consider these two issues in turn.

To get a sense of the first type of uncertainty, it is instructive to look at a few examples. The four panels in Figure 3.1 correspond to four different inflationary potentials. In each panel, the solid black curve corresponds to the actual primordial tensor power spectrum produced by the corresponding potential; the dashed horizontal curve comes from using the 0th-order approximation for the tensor extrapolation function $E_t \approx E_t^{(0)} = 1$; the green

curve comes from using the 1st-order approximation $E_t \approx E_t^{(1)}$ which retains the first factor in the product (3.8b); the blue curve comes from the 2nd-order approximation $E_t \approx E_t^{(2)}$ which retains the first two factors in the product (3.8b); and the red curve comes from the 3rd-order approximation $E_t \approx E_t^{(3)}$ which retains the first three factors in the product (3.8b). The basic point is clear from the figure: the second- and third-order extrapolation functions, $E_t^{(2)}$ and $E_t^{(3)}$, continue to give good approximations to the actual primordial tensor spectrum on LI scales ($10^{-1}$ Hz $< f < 10^{0}$ Hz). Looking at the individual panels in more detail:

- Panel (a) corresponds to the quartic potential well $V(\phi) = \frac{1}{4!}\lambda\phi^4$, which produces the CMB observables (evaluated at the pivot wavenumber $k = 0.05$ Mpc$^{-1}$) $r \approx 0.27$, $n_s - 1 \approx -0.05$, and $|dn_s/d\ln k| < 10^{-3}$. If we focus on the present-day frequency $f = 3$ Hz, in the middle of the BBO band, the zeroth-, first-, second-, and third-order extrapolation functions overestimate the primordial tensor power spectrum by factors of 6.70, 1.98, 1.36, and 1.17, respectively.

- Panel (b) corresponds to the quadratic potential well $V(\phi) = \frac{1}{2}m^2\phi^2$ [108], which produces the CMB observables $r \approx 0.14$, $n_s - 1 \approx -0.034$, and $|dn_s/d\ln k| < 10^{-3}$. At $f = 3$ Hz, the zeroth-, first-, second-, and third-order extrapolation functions overestimate the primordial tensor power spectrum by factors of 2.65, 1.42, 1.17, and 1.08, respectively.

- Panel (c) corresponds to the simplest spontaneous symmetry breaking potential $V(\phi) = V_0[1 - (\phi/\mu)^2]^2 = V_0 - m^2\phi^2 + \lambda\phi^4$. After matching the observed scalar amplitude, the potential still has one free parameter, which we choose so that $n_s - 1 = -0.04$, which then fixes $r \approx 0.046$ and $|dn_s/d\ln k| < 10^{-3}$. At $f = 3$ Hz, the zeroth-, first-, second-, and third-order extrapolation functions overestimate the primordial tensor power spectrum by factors of 1.58, 1.29, 1.14, and 1.07, respectively.

- Panel (d) corresponds to the axion-like cosine potential $V = V_0[1 + \cos(\phi/\mu)]$, often called "natural inflation" [2, 61]. Again, after matching the observed scalar amplitude,

we have an additional free parameter. This time we choose it so that $n_s - 1 = -0.05$, which fixes the other CMB observables to be $r \approx 0.033$ and $|dn_s/d\ln k| < 10^{-3}$. At $f = 3$ Hz, the zeroth-, first-, second-, and third-order extrapolation functions overestimate the primordial tensor power spectrum by factors of 1.53, 1.32, 1.17, and 1.08, respectively.

Let us make three comments. The first comment is that, as one can see from Figure 3.1, the extremely shallow slope of the primordial tensor power spectrum is ultimately the reason that $E_t^{(2)}$ and $E_t^{(3)}$ continue to agree well with $E_t$, all the way from CMB to LI scales scales. (Note that, in the figure, the vertical axis covers only 1.5 orders of magnitude, while the horizontal axis covers 23 orders of magnitude!) But this flatness is a very generic property of the inflationary tensor spectrum on scales that leave the horizon deep in the inflationary regime (since $\epsilon \ll 1$ in this regime, and so the corresponding tensor slope is $n_t \approx -2\epsilon \ll 1$). Futhermore, Smith, Kamionkowski, and Cooray have shown [148] that if the tensor amplitude is anywhere close to being large enough to be detectable, then both CMB and BBO scales must have left the horizon deep in the inflationary regime ($N \approx 60$ and $N \approx 20$ e-folds before the end of inflation, respectively). Thus, if we manage to detect $r$ on CMB scales, then we can assume that the tensor spectrum has a shallow slope, all the way from CMB to LI scales, and should be well approximated by $E_t^{(2)}$ or $E_t^{(3)}$ on these scales. The second comment is that, since all of the tensor parameters $\alpha_t^{(m)}$ are proportional to $r$, models with larger $r$ also have larger (although still very small) tensor tilt $n_t$, tensor running $dn_t/d\ln k$, and so on. Thus, we should expect the approximate extrapolation function at a given order to be less accurate for models with larger $r$; and this expectation is roughly born out (at each order: zeroth, first, second, and third) in the examples that we have tried, including the four examples shown in Figure 3.1. In particular, since $\lambda\phi^4$ produces roughly the largest tensor-to-scalar ratio ($r \approx 0.3$) allowed by current observational constraints, we can use the corresponding extrapolation errors for this model to conservatively estimate that, for models with $r \lesssim 0.3$, the second-order and third-order extrapolation functions, $E_t^{(2)}$ and $E_t^{(3)}$, should be accurate on BBO scales to within factors

of $\sim 1.4$ and $\sim 1.2$, respectively. Furthermore, these extrapolation uncertainties decrease as we tighten the upper bound on $r$. For example, if $r \lesssim 0.15$, then $E_t^{(2)}$ and $E_t^{(3)}$ should be accurate on BBO scales to within factors of $\sim 1.2$ and $\sim 1.1$, respectively. The third comment is that, although all current lines of evidence (including the observational upper limits on the magnitudes of $r$, $n_s - 1$ and $dn_s/d\ln k$) indicate that the inflaton potential $V(\phi)$ is a slowly varying function of the inflaton field $\phi$, it is nevertheless possible to imagine inserting a sharp feature, like a kink or step, in the inflaton potential betweeen $N \approx 60$ and $N \approx 20$ e-folds before the end of inflation, in such a way that this feature doesn't halt inflation, and is invisible to CMB and other large-scale structure experiments. In such a case, our approximate extrapolation functions, $E_t^{(2)}$ and $E_t^{(3)}$, based on information gathered on CMB scales, could provide a poor approximation on LI scales. We will ignore this unnecessarily baroque possibility here. One might say that the absence of such a kink is another part of the null hypothesis that we are hoping to test.

Before moving on to discuss the second type of extrapolation uncertainty, let us mention that it is, in fact, possible to replace $E_t^{(2)}$ and $E_t^{(3)}$ with improved extrapolation functions $\widehat{E}_t^{(2)}$ and $\widehat{E}_t^{(3)}$, respectively, that also just depend on the CMB observables $\{r, n_s\}$ and $\{r, n_s, dn_s/d\ln k\}$, respectively. We use the extrapolation functions $E_t^{(2)}$ and $E_t^{(3)}$ in this chapter since the improved extrapolation functions $\widehat{E}_t^{(2)}$ and $\widehat{E}_t^{(3)}$ are still work in progress. Nevertheless, we mention them here to clarify that the extrapolation errors estimated in the previous paragraph should not be regarded as final irreducible uncertainties. Note that the possibility of significantly improving on the extrapolation functions $E_t^{(2)}$ and $E_t^{(3)}$ is strongly hinted at by the fact that every approximate spectrum shown in every panel of Figure 3.1 systematically *overestimates* the actual spectrum. The basic idea behind the relationship between the "ordinary" second-order extrapolation function $E_t^{(2)}$ and the "improved" extrapolation function $\widehat{E}_t^{(2)}$ is that, even if we can only measure the CMB observables $\{r, n_s\}$, and hence can only determine the first two tensor parameters $\{\alpha_t^{(1)}, \alpha_t^{(2)}\}$, we can nevertheless use the hierarchy of inflationary consistency relations to *partially* determine each of the higher-order tensor parameters $\{\alpha_t^{(3)}, \alpha_t^{(4)}, \ldots\}$. The part of $\alpha_t^{(m)}$ that we can determine this

way may be called the "reduced" tensor parameter $\tilde{\alpha}_t^{(m)}$. We obtain the improved extrapolation function $\widehat{E}_t^{(2)}$ by replacing the higher tensor parameters $\{\alpha_t^{(3)}, \alpha_t^{(4)}, \ldots\}$ in the full extrapolation function (3.8b) by their "reduced" counterparts $\{\tilde{\alpha}_t^{(3)}, \tilde{\alpha}_t^{(4)}, \ldots\}$ — instead of setting these higher tensor parameters to zero, as we do in defining $E_t^{(2)}$. The relationship between the ordinary third-order extrapolation function $E_t^{(3)}$ and the improved third-order extrapolation function $\widehat{E}_t^{(3)}$ is similar. For the present chapter, though, it is sufficient to use the ordinary extrapolation functions $E_t^{(2)}$ and $E_t^{(3)}$, and leave further improvements to future work.

To estimate the second type of uncertainty, note that the amplitude of the primordial tensor power spectrum $\Delta_t^2(k, \tau_i)$ is proportional to $r(k_*)E_t(k, k_*)$, so we can estimate its fractional uncertainty on LI scales through the equation

$$\frac{\delta[\Delta_t^2(k, \tau_i)]}{\Delta_t^2(k, \tau_i)} \approx \frac{\delta[r(k_*)E_t^{(3)}(k, k_*)]}{r(k_*)E_t^{(3)}(k, k_*)}. \tag{3.17}$$

To examine this situation, let us first suppose that (by combining future CMB and other large-scale-structure experiments) we can measure each of the CMB observables $\{r, n_s, dn_s/d\ln k\}$ with the realistically-achivable one-sigma error bars $\pm 0.005$. What would be the induced uncertainty on LI scales?

- As a first example, suppose that on CMB scales we measure $r \approx 0.27$, $(n_s - 1) \approx -0.05$, and $dn_s/d\ln k \approx 0$, corresponding to model (a) in Figure 3.1. Then, on LI scales, the uncertainty in $r$ induces a very small "error factor" ($\sim 1.005$); the uncertainty in $n_s$ induces a modest error factor ($\sim 1.11$); and the uncertainty in $dn_s/d\ln k$ induces a large error factor ($\sim 4$). Note that when we refer to an "error factor" of 1.2, say, we mean that, at one-sigma, our prediction for $\Delta_t^2(k, \tau_i)$ on BBO scales can be pushed up or down by a factor of 1.2. Note that error factors (or, equivalently, logarithmic error bars) are more appropriate than linear error bars here because the dominant uncertainty on LI scales is induced by the CMB uncertainties appearing in the exponent of the tensor extrapolation function.

- As a second example, suppose that on CMB scales we measure $r \approx 0.14$, $(n_s -$

1) $\approx -0.034$, and $dn_s/d\ln k \approx 0$, corresponding to model (b) in Figure 3.1. Then, on LI scales, the uncertainty in $r$ again induces a small error factor ($\sim 1.015$); the uncertainty in $n_s$ again induces a modest error factor ($\sim 1.07$); and the uncertainty in $dn_s/d\ln k$ again induces a large error factor ($\sim 2$).

- As a third example, suppose that on CMB scales we measure $r \approx 0.046$, $(n_s - 1) \approx -0.04$, and $dn_s/d\ln k \approx 0$, corresponding to model (c) in Figure 3.1. Now, on LI scales, the situation is different: the uncertainty in $r$ induces a modest error factor ($\sim 1.07$); the uncertainty in $n_s$ induces a small error factor ($\sim 1.035$); and the uncertainty in $dn_s/d\ln k$ now contributes a significantly smaller error factor ($\sim 1.27$).

- As a final example, suppose that on CMB scales we measure $r \approx 0.033$, $(n_s - 1) \approx -0.05$, and $dn_s/d\ln k \approx 0$, corresponding to model (d) in Figure 3.1. Then, on LI scales, the uncertainty in $r$ induces a modest error factor ($\sim 1.10$); the uncertainty in $n_s$ induces a small error factor ($\sim 1.03$); and the uncertainty in $dn_s/d\ln k$ induces a modest error factor ($\sim 1.18$).

Again, let us make three comments. The first comment is that the extrapolation error factor induced by each CMB observable will obviously increase or decrease as the error bar on that CMB observable increases or decreases. In the above calculations, we have assumed error bars of $\pm 0.005$ on each of the CMB observables $\{r, n_s, dn_s/d\ln k\}$. These are realistic goals — but it may be possible to do even better. Indeed, there have even been a number of parameter-forecast papers in the literature (see [86, 151]) suggesting that the CMBPOL satellite — especially in combination with other complementary large-scale-structure measurements — might be able to achieve error bars that are significantly smaller than the ones we have assumed here, and this would decrease the extrapolation uncertainties accordingly. The second comment is that the relative sizes of the various extrapolation uncertainties can depend rather sensitively on the values of the CMB observables themselves. In particular, for large $r$, the uncertainty induced by $r$ is insignificant, the uncertainty induced by $n_s$ is modest, and the uncertainty induced by $dn_s/d\ln k$ is large. On the other

hand, for small $r$ ($r \lesssim 0.05$, say), the uncertainty induced by $r$ can be more important than the uncertainty induced by $n_s$ (which can become quite small), and the uncertainty induced by $dn_s/d\ln k$ is much reduced. The third comment is that the uncertainty induced by $dn_s/d\ln k$ can be so large (especially for large values of $r$) that we would have esimated a smaller uncertainty by simply neglecting $dn_s/d\ln k$ altogether and using the 2nd-order extrapolation function $E_t^{(2)}$ instead of the 3rd-order extrapolation function $E_t^{(3)}$. This is simply telling us that we should, indeed, work with $E_t^{(2)}$ instead of $E_t^{(3)}$ in these cases, since the running $dn_s/d\ln k$ is not well enough constrained to provide additional useful information in approximating the extrapolation function.

Although the extrapolation uncertainty induced by the error bar on each CMB observable may be most directly and accurately obtained by simply substituting the CMB observables (best $\pm$ uncertainty) into the approximate extrapolation function $E_t^{(3)}$, as we have done above, it is also useful to derive simple approximate expressions so we can see at a glance how the fractional uncertainty in the primordial tensor power spectrum on LI scales depends on the CMB observables and their uncertainties. Starting from Eq. (3.17), we can approximate the fractional uncertainties $\{\delta_r, \delta_1, \delta_2\}$ induced by the observables $\{r, \alpha_s^{(1)}, \alpha_s^{(2)}\}$, respectively, as

$$\delta_r \approx \frac{\partial[r(k_*)E_t^{(3)}(k,k_*)]}{\partial[r(k_*)]}\frac{\delta r(k_*)}{r(k_*)E_t^{(3)}(k,k_*)}, \tag{3.18a}$$

$$\delta_1 \approx \frac{\partial[r(k_*)E_t^{(3)}(k,k_*)]}{\partial[\alpha_s^{(1)}(k_*)]}\frac{\delta\alpha_s^{(1)}(k_*)}{r(k_*)E_t^{(3)}(k,k_*)}, \tag{3.18b}$$

$$\delta_2 \approx \frac{\partial[r(k_*)E_t^{(3)}(k,k_*)]}{\partial[\alpha_s^{(2)}(k_*)]}\frac{\delta\alpha_s^{(2)}(k_*)}{r(k_*)E_t^{(3)}(k,k_*)}, \tag{3.18c}$$

and then use the leading-order terms on the right-hand side of Eqs. (3.10,3.12a, 3.12b) to

rewrite these as

$$\delta_r = \delta r\Big[\frac{1}{r} - \Big(\frac{1}{8}\Big)K + \Big(\frac{\alpha_s^{(1)}}{16} + \frac{r}{64}\Big)K^2 + \qquad (3.19a)$$
$$+\Big(\frac{\alpha_s^{(2)}}{48} - \frac{\alpha_s^{(1)2}}{48} - \frac{\alpha_s^{(1)}r}{64} - \frac{r^2}{512}\Big)K^3\Big],$$
$$\delta_1 = \delta\alpha_s^{(1)}\Big[\Big(\frac{r}{16}\Big)K^2 - \Big(\frac{\alpha_s^{(1)}r}{24} + \frac{r^2}{128}\Big)K^3\Big], \qquad (3.19b)$$
$$\delta_2 = \delta\alpha_s^{(2)}\Big[\Big(\frac{r}{48}\Big)K^3\Big]. \qquad (3.19c)$$

One can check that these formulae agree very well with the induced error factors computed above directly from the 3rd-order extrapolation function $E_t^{(3)}$ (except, of course, when $\delta_2$ grows to order unity). To get numbers out of these formulae, it is useful to recall that CMB/large-scale-structure scales and LI scales are separated by roughly $K \approx 36$.

## 3.3 The tensor transfer function $T_t(k, \tau)$

In this section we derive a general expression for the gravitational-wave transfer function — applicable to those modes that have re-entered the horizon after inflation and, in particular, to those modes probed by laser-interferometer (LI) gravitational-wave experiments. We pay special attention to early-universe effects (from the "primordial dark age" between the end of inflation and the time of big bang nucleosynthesis) that may imprint themselves in the transfer function on LI scales.

The gravitational wave spectrum generated by inflation carries important information about the conditions *during* inflation. But the spectrum also receives corrections, both large and small, from the subsequent evolution and matter content of the universe *after* inflation. In this section, we identify various post-inflationary physical effects which modify the gravitational-wave background (GWB), and show how they may be encoded in the gravitational-wave transfer function that relates the primordial tensor power spectrum to the gravitational-wave spectrum at a later point in cosmic history. It is necessary to properly understand and disentangle the post-inflationary effects in order to optimally extract the inflationary information in the GWB. But these modifications are also interesting in their

own right, since they offer a rare window onto the physical properties of the high-energy universe during the "primordial dark age" between the end of inflation and the electroweak phase transition.

The same features that make the inflationary GWB difficult to detect — namely its small amplitude and gravitational-strength coupling to matter — also make it a clean cosmological probe. First, because of their tiny amplitude, the gravitational waves obey linear equations of motion, so that their evolution may be predicted analytically with high precision. (By contrast, density perturbations grow after horizon entry, and perturbations shorter than $\sim 10$ Mpc have already grown non-linear; so it is impossible to study their evolution analytically, and even numerical analysis is challenging.) Second, because of their ultra-weak interactions with matter, the gravitational waves have been free streaming since the end of inflation — in contrast to neutrinos (which began streaming roughly a second later) and photons (which began streaming several hundred thousand years later). The gravitational waves carry unsullied information from the early universe, and subsequent modifications of the gravitational-wave spectrum are not washed out by thermal effects (since the gravitons are thermally decoupled).

The gravitational-wave spectrum near a given wavenumber $k$ is primarily sensitive to two "moments" in cosmic history: (1) the moment when that mode "left the horizon" (*i.e.*, became longer than the instantaneous Hubble radius during inflation), and (2) the moment when the mode "re-entered the horizon" (*i.e.*, became shorter than the instantaneous Hubble radius once again, after the end of inflation). The first moment imprints information about inflation itself, while the second moment imprints information about post-inflationary conditions. The CMB is sensitive to long-wavelength modes that re-entered at relatively low temperatures (well after BBN), corresponding to relatively well-understood physics. By contrast, laser interferometers are sensitive to shorter wavelengths that entered the horizon at high temperatures ($T \sim 10^7$ GeV), well above the electroweak phase transition. The physical conditions at such high energies, which are considerably beyond the reach of particle accelerators, are a mystery, so that any information about this epoch from the GWB

would be very valuable.

The outline of this section is as follows. In subsection 3.3.1, we review the basics of inflationary gravitational waves. We take special care to clearly establish our conventions, and to point out where conventions diverge and become confused in the literature. In subsection 3.3.2 we present our approach to computing the gravitational-wave transfer function by separating it into three factors: $T_t = C_1 C_2 C_3$. Each factor has a distinct physical meaning, and we will include several effects that were absent in previous calculations [148, 169, 11, 125, 126].

It is worth noting that various papers in the literature use the phrases "gravitational-wave transfer function" and "tensor transfer function" to mean various slightly different things. For example, whereas some authors use these phrases to refer to the factor that propagates tensor fluctuations $h_k(\tau)$ forward in time, we use it to refer to the factor that propagates the tensor *power spectrum* $\Delta_t^2(k,\tau)$ forward in time. For three definitions that are slightly different from (and more CMB specific than) ours, see [181, 11, 169]. Thus, we should clarify that the tensor transfer function computed in this paper may be more specifically described as the "tensor-power-spectrum transfer function," and is defined in a simple way by the equations (3.49) and (3.50). From now on, when we refer to the transfer function, this is the object we shall mean.

In subsection 3.3.3 we derive the first factor, which accounts for the redshift-suppression of the gravitational-wave amplitude after horizon re-entry. Among other things, this factor accomodates a dark energy component with a time-varying equation-of-state parameter $w(z)$. In subsection 3.3.4 we compute the second factor, which captures the behavior of the background equation-of-state parameter $w$ near the time of horizon re-entry. In subsection 3.3.5 we compute the third factor, which accounts for the damping of tensor modes due to tensor anisotropic stress from free-streaming relativistic particles in the early universe. This damping effect was recently pointed out by Weinberg [179], and the damping on CMB scales due to free-streaming neutrinos has been studied in [179, 134, 11, 46]. In this chapter, we point out that it is also necessary to consider this damping effect on laser-interferometer

scales, which re-enter the horizon when free-streaming particles were an unknown fraction $\Omega_{free}$ of the background energy density. We present accurate expressions for the damping effect as a function of $\Omega_{free}$. We also observe that Weinberg's analysis, which originally focused on a single fermionic particle (the neutrino), extends in a simple way to the more general case of a mixture of free-streaming bosons and fermions with different temperatures and decoupling times.

In subsection 3.3.6 we identify six physical effects which can modify the relic GWB by causing the equation-of-state parameter $w$ to deviate from its standard value ($w = 1/3$) during the radiation-dominated epoch. Furthermore, although it is often treated as a stationary random process, the inflationary GWB is actually highly non-stationary (as emphasized by Grishchuk [71, 72]; also see [3, 132, 6]). Thus, our transfer function keeps track of the coherent phase information that it contains, and we emphasize a related factor-of-2 suppression of the transfer function that has been neglected in previous treatments (see the discussion in subsection 3.3.2).

### 3.3.1 Tensor perturbations: fundamentals and conventions

In this subsection, we derive some basic facts about inflationary gravitational waves. These derivations are not new, but are intended to make our discussion of the transfer function more pedagogical and self-contained. They also establish our conventions explicitly, and provide a brief guide to conventions used by other authors. Such a guide is necessary since there are many slightly different conventions floating around in the inflationary gravitational-wave literature and, as a result, erroneous factors of 2 and $\pi$ are ubiquitous. We have been careful to highlight each spot where convention choices crop up, and to state which convention we have chosen.

Tensor perturbations in a spatially-flat Friedmann-Robertson-Walker (FRW) universe are described by the line element

$$\mathrm{d}s^2 = a^2[-\mathrm{d}\tau^2 + (\delta_{ij} + h_{ij})\mathrm{d}x^i\mathrm{d}x^j], \tag{3.20}$$

where $\tau$ is the conformal time, $x^i$ are comoving spatial coordinates, and $h_{ij}$ is the gauge-

invariant tensor metric perturbation. Note that, although our definition of $h_{ij}$ is perhaps the most common (see *e.g.* [179, 123, 114]), some authors define the tensor perturbation with an extra factor of 2 (see *e.g.* [9, 99]) so that the spatial metric is written $\delta_{ij}+2h_{ij}$. Throughout this appendix, if a term contains a repeated index, that index should be summed: from 1 to 3 for latin indices and from 0 to 3 for greek indices. The perturbation $h_{ij}$ is symmetric ($h_{ij}=h_{ji}$), traceless ($h_{ii}=0$), and transverse ($h_{ij,j}=0$), and therefore contains $6-1-3=2$ independent modes (corresponding to the "+" and "×" gravitational-wave polarizations).

One can think of $h_{ij}(\tau,\mathbf{x})$ as a quantum field in an unperturbed FRW background metric $\bar{g}_{\mu\nu}=diag\{-a^2,a^2,a^2,a^2\}$. At quadratic order in $h_{ij}$ (which is adequate, since $h_{ij}$ is tiny), tensor perturbations are governed by the action (see *e.g.* [123, 114])

$$S=\int d\tau d\mathbf{x}\sqrt{-\bar{g}}\left[\frac{-\bar{g}^{\mu\nu}}{64\pi G}\partial_\mu h_{ij}\partial_\nu h_{ij}+\frac{1}{2}\Pi_{ij}h_{ij}\right]. \tag{3.21}$$

where $\bar{g}^{\mu\nu}$ and $\bar{g}$ are the inverse and determinant of $\bar{g}_{\mu\nu}$, respectively, and $G$ is Newton's constant. The tensor part of the anisotropic stress

$$\Pi_{ij}=T^i_j-p\delta^i_j \quad (p=\text{unperturbed pressure}) \tag{3.22}$$

satisfies $\Pi_{ii}=0$ and $\partial_i\Pi_{ij}=0$, and couples to $h_{ij}$ like an external source in (3.21). By varying $h_{ij}$ in (3.21), we obtain the equation of motion

$$h''_{ij}+2\frac{a'(\tau)}{a(\tau)}h'_{ij}-\nabla^2 h_{ij}=16\pi G a^2(\tau)\Pi_{ij}(\tau,\mathbf{x}), \tag{3.23}$$

where a prime ($'$) indicates a conformal time derivative $d/d\tau$. Next, it is convenient to Fourier transform as follows:

$$h_{ij}(\tau,\mathbf{x}) \;=\; \sum_r \sqrt{16\pi G}\int\frac{d\mathbf{k}}{(2\pi)^{3/2}}\epsilon^r_{ij}(\mathbf{k})h^r_{\mathbf{k}}(\tau)\mathrm{e}^{i\mathbf{kx}}, \tag{3.24a}$$

$$\Pi_{ij}(\tau,\mathbf{x}) \;=\; \sum_r \sqrt{16\pi G}\int\frac{d\mathbf{k}}{(2\pi)^{3/2}}\epsilon^r_{ij}(\mathbf{k})\Pi^r_{\mathbf{k}}(\tau)\mathrm{e}^{i\mathbf{kx}}, \tag{3.24b}$$

where $r=$("+" or "×") labels the polarization state, and the polarization tensors are symmetric [$\epsilon^r_{ij}(\mathbf{k})=\epsilon^r_{ji}(\mathbf{k})$], traceless [$\epsilon^r_{ii}(\mathbf{k})=0$], and transverse [$k_i\epsilon^r_{ij}(\mathbf{k})=0$]. We also choose a circular-polarization basis in which $\epsilon^r_{ij}(\mathbf{k})=(\epsilon^r_{ij}(\text{-}\mathbf{k}))^*$, and normalize the polarization

basis as follows:

$$\sum_{i,j} \epsilon^r_{ij}(\mathbf{k})(\epsilon^s_{ij}(\mathbf{k}))^* = 2\delta^{rs}. \tag{3.25}$$

Although our normalization convention (3.25) is the most standard one, other conventions — *i.e.* different numerical constants on the right-hand side of (3.25) — exist in the literature. Substituting (3.24) into (3.21) then yields

$$S = \sum_r \int d\tau d\mathbf{k} \frac{a^2}{2} \Big[ h^r_{\mathbf{k}}{}' h^r_{-\mathbf{k}}{}' - k^2 h^r_{\mathbf{k}} h^r_{-\mathbf{k}} + 32\pi G a^2 \Pi^r_{\mathbf{k}} h^r_{-\mathbf{k}} \Big]. \tag{3.26}$$

Now we can canonically quantize by promoting $h^r_{\mathbf{k}}$ and its conjugate momentum

$$\pi^r_{\mathbf{k}}(\tau) = a^2(\tau) h^r_{-\mathbf{k}}{}'(\tau) \tag{3.27}$$

to operators, $\hat{h}^r_{\mathbf{k}}$ and $\hat{\pi}^r_{\mathbf{k}}$, satisfying the equal-time commutation relations

$$\left[ \hat{h}^r_{\mathbf{k}}(\tau), \hat{\pi}^s_{\mathbf{k}'}(\tau) \right] = i\delta^{rs}\delta^{(3)}(\mathbf{k} - \mathbf{k}'), \tag{3.28a}$$

$$\left[ \hat{h}^r_{\mathbf{k}}(\tau), \hat{h}^s_{\mathbf{k}'}(\tau) \right] = [\hat{\pi}^r_{\mathbf{k}}(\tau), \hat{\pi}^s_{\mathbf{k}'}(\tau)] = 0. \tag{3.28b}$$

Since $\hat{h}_{ij}(\tau, \mathbf{x})$ is hermitian, its fourier components satisfy $\hat{h}^r_{\mathbf{k}} = \hat{h}^{r\dagger}_{-\mathbf{k}}$, and we write them as

$$\hat{h}^r_{\mathbf{k}}(\tau) = h_k(\tau)\hat{a}^r_{\mathbf{k}} + h^*_k(\tau)\hat{a}^{r\dagger}_{-\mathbf{k}}, \tag{3.29}$$

where the creation and annihilation operators, $\hat{a}^{r\dagger}_{\mathbf{k}}$ and $\hat{a}^r_{\mathbf{k}}$, satisfy standard commutation relations

$$\left[ \hat{a}^r_{\mathbf{k}}, \hat{a}^{s\dagger}_{\mathbf{k}'} \right] = \delta^{rs}\delta^{(3)}(\mathbf{k} - \mathbf{k}'), \tag{3.30a}$$

$$\left[ \hat{a}^r_{\mathbf{k}}, \hat{a}^s_{\mathbf{k}'} \right] = \left[ \hat{a}^{r\dagger}_{\mathbf{k}}, \hat{a}^{s\dagger}_{\mathbf{k}'} \right] = 0, \tag{3.30b}$$

while the ($c$-number) mode functions $h_k(\tau)$ and $h^*_k(\tau)$ are linearly-independent solutions of the fourier-transformed equation of motion

$$h''_k + 2\frac{a'(\tau)}{a(\tau)} h'_k + k^2 h_k = 16\pi G a^2(\tau)\Pi_k(\tau). \tag{3.31}$$

Eq. (3.29) makes use of the fact that, by isotropy, the mode functions $h_k(\tau)$ will depend on the time ($\tau$) and the wavenumber ($k = |\mathbf{k}|$), but not on the direction ($\hat{\mathbf{k}}$) or the polarization

$(r)$. Note that consistency between the two sets of commutation relations, (3.28) and (3.30), requires that the mode functions satisfy the Wronskian normalization condition

$$h_k(\tau)h_k^{*\prime}(\tau) - h_k^*(\tau)h_k'(\tau) = \frac{i}{a^2(\tau)} \tag{3.32}$$

in the past. In particular, the standard initial condition for the mode function in the far past (when the mode $k$ was still far inside the horizon during inflation):

$$h_k(\tau) \to \frac{\exp(-ik\tau)}{a(\tau)\sqrt{2k}} \qquad (\text{as } \tau \to -\infty), \tag{3.33}$$

satisfies (3.32) — but it is not the unique initial condition which does so. This is a manifestation of the well known vacuum ambiguity that is responsible for particle production in cosmological spacetimes (see [17]).

Now that we have discussed the quantization of tensor perturbations, let us turn to the three different spectra that are commonly used to describe the stochastic GWB: the tensor power spectrum $\Delta_t^2(k,\tau)$, the chirp amplitude $h_c(k,\tau)$, and the energy spectrum $\Omega_{gw}(k,\tau)$.

In the early universe, the GWB is usually characterized by the tensor power spectrum $\Delta_t^2(k,\tau)$. With the formalism developed thus far, one can check that

$$\langle 0|\hat{h}_{ij}(\tau,\mathbf{x})\hat{h}_{ij}(\tau,\mathbf{x})|0\rangle = \int_0^\infty 64\pi G\frac{k^3}{2\pi^2}|h_k(\tau)|^2\frac{dk}{k}, \tag{3.34}$$

so that the tensor power spectrum is given by

$$\Delta_t^2(k,\tau) \equiv \frac{d\langle 0|\hat{h}_{ij}^2|0\rangle}{d\ln k} = 64\pi G\frac{k^3}{2\pi^2}\left|h_k(\tau)\right|^2. \tag{3.35}$$

Note that our definition (3.35) agrees with the WMAP convention (see [128]) — this will be clearer when we present the approximate slow-roll form of the spectrum below. Although the WMAP convention seems to be becoming the standard one, several other definitions of the tensor power spectrum exist in the literature, and differ from (3.35) by an overall numerical constant. Also, since (3.35) defines the tensor power spectrum in terms of the full tensor perturbation $h_{ij}$, the normalization of the power spectrum is independent of the normalization (3.25) of the polarization basis. By contrast, some authors define the tensor power spectrum in terms of the polarization components of $h_{ij}$, so that the normalization

of the spectrum is linked to the convention-dependent coefficient on the right-hand side of (3.25).

The present-day GWB is usually characterized either by its chirp amplitude $h_c(k,\tau)$, or by its energy spectrum $\Omega_{gw}(k,\tau)$. First, the chirp amplitude represents the rms dimensionless strain ($\sim \delta L/L$ in a gravitational wave antenna) per logarithmic wavenumber interval (or logarithmic frequency interval), and is related to the tensor power spectrum as follows: $h_c(k,\tau) = \sqrt{\Delta_t^2(k,\tau)/2}$ (see [161] for more details). The energy spectrum

$$\Omega_{gw}(k,\tau) \equiv \frac{1}{\rho_{crit}(\tau)} \frac{d\langle 0|\hat{\rho}_{gw}(\tau)|0\rangle}{d\ln k} \tag{3.36}$$

represents the gravitational-wave energy density ($\rho_{gw}$) per logarithmic wavenumber interval, in units of the "critical density"

$$\rho_{crit}(\tau) = \frac{3H^2(\tau)}{8\pi G}. \tag{3.37}$$

To compute $\Omega_{gw}(k,\tau)$, we can again think of the tensor perturbation $h_{ij}$ as a field in an unperturbed FRW background metric $\bar{g}_{\mu\nu}$, and then use its action (3.21) to compute its stress-energy tensor

$$T_{\alpha\beta} = -2\frac{\delta L}{\delta \bar{g}^{\alpha\beta}} + \bar{g}_{\alpha\beta} L, \tag{3.38}$$

where $L$ is the Lagrangian function in (3.21) — see, for example, Ch. 21.3 in [119]. In particular, the gravitational-wave energy density (ignoring the anisotropic stress coupling) is

$$\rho_{gw} = -T_0^0 = \frac{1}{64\pi G} \frac{(h'_{ij})^2 + (\vec{\nabla} h_{ij})^2}{a^2}, \tag{3.39}$$

which has vacuum expectation value

$$\langle 0|\rho_{gw}|0\rangle = \int_0^\infty \frac{k^3}{2\pi^2} \frac{|h'_k|^2 + k^2 |h_k|^2}{a^2} \frac{dk}{k}, \tag{3.40}$$

so that the energy spectrum is given by

$$\Omega_{gw}(k,\tau) = \frac{8\pi G}{3H^2(\tau)} \frac{k^3}{2\pi^2} \frac{|h'_k(\tau)|^2 + k^2 |h_k(\tau)|^2}{a^2(\tau)}. \tag{3.41}$$

We have shown how each of the three spectra ($\Delta_t^2$, $h_c$, and $\Omega_{gw}$) may be written in terms of the mode functions $h_k(\tau)$. Also note that, once the mode $k$ re-enters the horizon after

inflation, the corresponding mode function obeys $|h_k'(\tau)|^2 = k^2|h_k(\tau)|^2$, as explained in the next section, so that the three spectra may be related to one another in a simple way:

$$\Omega_{gw}(k,\tau) = \frac{1}{12}\frac{k^2\Delta_t^2(k,\tau)}{a^2(\tau)H^2(\tau)} = \frac{1}{6}\frac{k^2 h_c^2(k,\tau)}{a^2(\tau)H^2(\tau)}. \tag{3.42}$$

Next, let us "derive" the slow-roll expression for the primordial tensor power spectrum. As long as $k$ remains inside the Hubble horizon ($k \gg aH$) during inflation, the mode function $h_k^{(\mathrm{in})}(\tau)$ is given by (3.33); and once $k$ is outside the horizon ($k \ll aH$), the mode function $h_k^{(\mathrm{out})}$ is independent of $\tau$. Then, by simply matching $|h_k^{(\mathrm{in})}|$ to $|h_k^{(\mathrm{out})}|$ at the moment of horizon-exit ($k = aH$), one finds $h_k^{(\mathrm{out})} = 1/(a_*\sqrt{2k})$, where an asterisk ($*$) denotes that a quantity is evaluated at horizon-exit ($k = a_*H_*$). Thus, the primordial tensor power spectrum is given by

$$\Delta_t^2(k) \equiv 64\pi G\frac{k^3}{2\pi^2}\left|h_k^{(\mathrm{out})}\right|^2 \approx 8\left(\frac{H_*}{2\pi M_{pl}}\right)^2, \tag{3.43}$$

where, in this equation (and in the remainder of this section) "$\equiv$" denotes a definition, "$\approx$" indicates that the slow-roll approximation has been used, and $M_{pl} = (8\pi G)^{-1/2}$ is the "reduced Planck mass." Note that the primordial power spectrum is time-independent, since (by definition) it is evaluated when the mode $k$ is far outside the horizon. Although our derivation of (3.43) may seem crude, it is well known that (3.43) provides a very good approximation to the inflationary tensor spectrum. We will not reproduce it here, but a closely analogous derivation leads to the slow-roll expression for the primordial *scalar* power spectrum:

$$\Delta_s^2(k) \approx \frac{1}{2\epsilon_*}\left(\frac{H_*}{2\pi M_{pl}}\right)^2. \tag{3.44}$$

It is also useful to define the tensor/scalar ratio

$$r(k) \equiv \Delta_t^2(k)/\Delta_s^2(k) \approx 16\epsilon_*, \tag{3.45}$$

along with the scalar and tensor spectral indices

$$n_s(k) - 1 \equiv \frac{d\ln\Delta_s^2}{d\ln k} \approx -6\epsilon_* + 2\eta_*, \tag{3.46a}$$

$$n_t(k) \equiv \frac{d\ln\Delta_t^2}{d\ln k} \approx -2\epsilon_*. \tag{3.46b}$$

In the slow-roll formulae above, we have introduced the usual "potential" slow roll parameters

$$\epsilon \equiv \frac{1}{2} M_{pl}^2 \left( \frac{V'(\phi)}{V(\phi)} \right)^2, \qquad \eta \equiv M_{pl}^2 \frac{V''(\phi)}{V(\phi)}, \tag{3.47}$$

where $V(\phi)$ is the inflaton potential. Note that Eqs. (3.45) and (3.46b) together imply the well known consistency relation for inflation

$$r = -8 n_t. \tag{3.48}$$

### 3.3.2 Organizing the calculation

Since cosmological tensor perturbations are tiny, they are well described by linear perturbation theory, so that each fourier mode evolves independently. Thus, we see from Eq. (3.35) that the primordial tensor power spectrum — defined at some conformal time $\tau_i$ shortly after the end of inflation, when all modes of interest have already left the horizon, but have not yet re-entered — is related to the tensor power spectrum at a later time $\tau$ by a multiplicative transfer function

$$\Delta_t^2(k, \tau) = T_t(k, \tau) \Delta_t^2(k, \tau_i), \tag{3.49}$$

where

$$T_t(k, \tau) = \left| \frac{h_k(\tau)}{h_k(\tau_i)} \right|^2. \tag{3.50}$$

Note that we will not necessarily want to evaluate $T_t(k, \tau)$ at the present time ($\tau = \tau_0$), since different experiments probe the gravitational-wave spectrum at different redshifts. For example, while laser interferometers measure $T_t$ today, CMB experiments measure it near the redshift of recombination. As long as a mode remains outside the horizon ($k \ll aH$), the corresponding perturbation does not vary with time [$h_k(\tau) = h_k(\tau_i)$], so that the transfer function is very well approximated by $T_t(k, \tau) = 1$. (For a general proof, even in the presence of anisotropic stress, see the appendix in [179].) Thus, the rest of this paper will focus on $T_t(k, \tau)$ for modes that have already re-entered the horizon prior to time $\tau$.

It is very convenient to split the transfer function (3.50) into three factors as follows:

$$T_t(k,\tau) = \left| \frac{\bar{h}_k(\tau)}{h_k(\tau_i)} \frac{\widetilde{h}_k(\tau)}{\bar{h}_k(\tau)} \frac{h_k(\tau)}{\widetilde{h}_k(\tau)} \right|^2 = C_1 C_2 C_3. \tag{3.51}$$

Here $h_k(\tau)$, $\widetilde{h}_k(\tau)$ and $\bar{h}_k(\tau)$ represent three different solutions of the tensor mode equation (3.31). In particular, $h_k(\tau)$ is the true (exact) solution of (3.31); $\widetilde{h}_k(\tau)$ is an approximate solution obtained by ignoring the tensor anisotropic stress $\Pi_k$ on the right-hand-side of (3.31); and $\bar{h}_k(\tau)$ is an even cruder approximation obtained by first ignoring $\Pi_k$ and then using the "thin-horizon" approximation, described in subsection 3.3.3, to solve (3.31). [Briefly, the thin-horizon approximation treats horizon re-entry as a "sudden" or instantaneous event. In this approximation, $\bar{h}_k(\tau)$ is frozen outside the Hubble horizon, redshifts as $1/a(\tau)$ inside the Hubble horizon, and has a sharp transition between these two behaviors at the moment when the mode re-enters the Hubble horizon ($k = aH$).]

These three factors each represent a different physical effect. The first factor,

$$C_1 = \left| \frac{\bar{h}_k(\tau)}{h_k(\tau_i)} \right|^2, \tag{3.52}$$

represents the redshift-suppression of the gravitational-wave amplitude after the mode $k$ re-enters the horizon. The second factor,

$$C_2 = \left| \frac{\widetilde{h}_k(\tau)}{\bar{h}_k(\tau)} \right|^2, \tag{3.53}$$

captures the behavior of the background equation-of-state parameter $w(\tau) = p(\tau)/\rho(\tau)$ around the time of horizon re-entry. The third factor,

$$C_3 = \left| \frac{h_k(\tau)}{\widetilde{h}_k(\tau)} \right|^2, \tag{3.54}$$

measures the damping of the gravitational-wave spectrum due to tensor anisotropic stress. Note that $C_1$ by itself is $\ll 1$ and provides a rough approximation to the full transfer function $T_t$. The other two factors, $C_2$ and $C_3$, are typically of order unity, and are primarily sensitive to the physical conditions near the time that the mode $k$ re-entered the Hubble horizon. In subsections 3.3.3, 3.3.4, and 3.3.5, we derive expressions for $C_1$, $C_2$ and $C_3$,

respectively, and flesh out the physical interpretations given above. In subsection 3.3.6, we discuss various physical effects that cause deviations $\delta w$ from the usual equation-of-state parameter ($w = 1/3$) in the early universe, and explain how these effects modify the gravitational-wave transfer function.

### 3.3.3 The redshift-suppression factor, $C_1$

The mode function $h_k(\tau)$ behaves simply in two regimes: far outside the horizon ($k \ll aH$), and far inside the horizon ($k \gg aH$). Far outside, $h_k(\tau)$ is $\tau$-independent (as we have seen). Far inside, after horizon re-entry, $h_k(\tau)$ oscillates with a decaying envelope

$$h_k(\tau) = \frac{A_k}{a(\tau)} \cos[k(\tau - \tau_k) + \phi_k], \tag{3.55}$$

as we shall see in the next two subsections, where $A_k$ and $\phi_k$ are constants representing the amplitude and phase shift of the oscillation, and $\tau_k$ is the conformal time at horizon re-entry ($k = aH$). These two simple regimes are separated by an intermediate period (horizon-crossing) when $k \sim aH$.

In the thin-horizon approximation, we neglect this intermediate regime. That is, we assume that $\bar{h}_k(\tau) = h_k(\tau_i)$ when $k < aH$; and that $\bar{h}_k(\tau)$ is given by Eq. (3.55) for $k > aH$; and that the outside amplitude is connected to the inside envelope via the matching condition $h_k(\tau_i) = A_k/a(\tau_k)$. Ignoring the phase shift $\phi_k$, which is really an asymptotic quantity, this matching condition simply imposes continuity of the inside and outside amplitudes at $k = aH$. Combining the matching condition with Eq. (3.52), we see that

$$C_1 = \left(\frac{1+z}{1+z_k}\right)^2 \cos^2[k(\tau - \tau_k) + \phi_k], \tag{3.56}$$

where $1 + z = a_0/a(\tau)$ is the redshift at which the spectrum is to be probed, and $1 + z_k = a_0/a_k$ is the redshift at which the mode re-entered the Hubble horizon ($k = aH$).

The relic GWB from inflation is often treated as a "quasi-stationary" process (which means that its statistical properties only vary on cosmological time scales — much longer than the timescales in a terrestrial experiment). But the $\cos^2[\ldots]$ factor in Eq. (3.56) implies that the background is actually highly *non-stationary* — its power spectrum oscillates as

a function of both wavenumber $k$ and time $\tau$. This $\cos^2[\ldots]$ factor represents a genuine feature, and is *not* a spurious byproduct of our thin-horizon approximation. Physically (as observed in [71, 72]) the inflationary GWB consists of gravitational *standing waves* with random *spatial* phases, and coherent *temporal* phases. All modes $\vec{k}$ at fixed wavenumber $k = |\vec{k}|$ re-enter the Hubble horizon simultaneously, and subsequently oscillate in phase with one another — even until the present day. Classically, the modes are synchronized by inflation; quantum mechanically, they are squeezed [71, 72]. Thus, at a fixed wavenumber, $T_t(k,\tau)$ is sinusoidal in $\tau$, with oscillation frequency $k$, and a phase shift $\phi_k$ (computed in the next subsection). Alternatively, at fixed time, $T_t(k,\tau)$ oscillates rapidly in wavenumber. For the direct detection experiments we are considering, which cannot resolve these oscillations, the factor $\cos^2[\ldots]$ should be averaged, and replaced by $1/2$ in Eq. (3.56).

In the remainder of this subsection, we derive an accurate expression for $(1+z_k)$, the horizon-crossing redshift. To start, let us write the background energy density $\rho$ as a sum of several components. The $i$th component has energy density $\rho_i$, pressure $p_i$, equation-of-state parameter $w_i \equiv p_i/\rho_i$, and obeys a continuity equation

$$\frac{d\rho_i}{\rho_i} = 3[1+w_i(z)]\frac{dz}{1+z}. \tag{3.57}$$

Integrating this equation yields

$$\rho_i(z)/\rho_i^{(0)} = (1+z)^3 \exp\left[3\int_0^z \frac{w_i(\tilde{z})}{1+\tilde{z}}d\tilde{z}\right], \tag{3.58}$$

where $\rho_i^{(0)}$ is the present value. Then the Friedmann equation

$$H^2(z) = \frac{8\pi G}{3}\sum_i \rho_i(z) \tag{3.59}$$

may be rewritten as

$$\frac{a^2H^2}{a_0^2H_0^2} = \sum_i \Omega_i^{(0)}(1+z)\exp\left[3\int_0^z \frac{w_i(\tilde{z})}{1+\tilde{z}}d\tilde{z}\right], \tag{3.60}$$

where $H_0$ is the present Hubble rate, and the density parameter $\Omega_i^{(0)} \equiv \rho_i^{(0)}/\rho_{cr}^{(0)}$ represents the $i$th component's present energy density in units of the present critical density $\rho_{cr}^{(0)} =$

$3H_0^2/8\pi G$. Hence $z_k$ is obtained by solving the equation

$$(k/k_0)^2 = F(z_k) \tag{3.61}$$

where

$$F(z_k) = \sum_i \Omega_i^{(0)}(1+z_k)\exp\left[3\int_0^{z_k}\frac{w_i(z)}{1+z}dz\right], \tag{3.62}$$

and $k_0 = a_0 H_0 = h \times 3.24 \times 10^{-18}\,\mathrm{Hz}$ is today's horizon wavenumber.

Before solving this equation properly, let us pause to extract a few familiar approximate scalings from our formalism. Since the primordial inflationary power spectrum $\Delta_t^2(k,\tau_i)$ is roughly scale invariant [$\propto (k/k_0)^0$], the current power spectrum $\Delta_t^2(k,\tau_0)$ is roughly $\propto C_1$, and hence $\propto (1+z_k)^{-2}$. From Eq. (3.42) we have $h_c(k,\tau_0) \propto (1+z_k)^{-1}$ and $\Omega_{gw}(k,\tau_0) \propto (k/k_0)^2(1+z_k)^{-2}$. For modes that re-enter the horizon during radiation domination, when the $w_r = 1/3$ term dominates the sum (3.62), we solve (3.61) to find $(1+z_k) \propto (k/k_0)$, which implies the approximate scalings $h_c(k,\tau_0) \propto (k/k_0)^{-1}$ and $\Omega_{gw}(k,\tau_0) \propto (k/k_0)^0$. For modes that re-enter during matter domination, when the $w_m = 0$ term dominates the sum (3.62), we find $(1+z_k) \propto (k/k_0)^2$, which implies that $h_c(k,\tau_0)$ and $\Omega_{gw}(k,\tau_0)$ are both $\propto (k/k_0)^{-2}$ in this regime.

For a more proper analysis, consider a universe with 4 components: matter ($w_m = 0$), curvature ($w_K = -1/3$), dark energy ($w_{de}(z)$), and radiation ($w_r(z) = 1/3 + \delta w_r(z)$). Note that, although one often assumes $w_r = 1/3$ during radiation domination, we have allowed for corrections $\delta w_r(z)$ due to early-universe effects discussed in subsection 3.3.6. Then we can write

$$F(z) = \hat{F}(z) + \delta F(z), \tag{3.63}$$

where

$$\hat{F}(z) = \Omega_r^{(0)}(1+z)^2 + \Omega_m^{(0)}(1+z) + \Omega_K^{(0)}, \tag{3.64}$$

and

$$\begin{aligned}\delta F(z) &= \Omega_{de}^{(0)}(1+z)\exp\Big[3\int_0^z d\tilde{z}\frac{w_{de}(\tilde{z})}{1+\tilde{z}}\Big] \\ &+ \Omega_r^{(0)}(1+z)^2\Big\{\exp\Big[3\int_0^z d\tilde{z}\frac{\delta w_r(\tilde{z})}{1+\tilde{z}}\Big]-1\Big\}.\end{aligned} \tag{3.65}$$

Here $\hat{F}$ represents a universe with spatial curvature, matter, and "standard" ($w_r = 1/3$) radiation; and $\delta F$ contains the modifications due to dark energy ($w_{de}$) and equation-of-state corrections ($\delta w_r$).

If we neglect these modifications [by setting $\Omega_{de}^{(0)} = 0 = \delta w_r(z)$ so that $\delta F = 0$], Eq. (3.61) has the exact solution

$$1 + \hat{z}_k \equiv \frac{1 + z_{eq}}{2} \left[ -1 + \sqrt{1 + \frac{4[(k/k_0)^2 - \Omega_K^{(0)}]}{(1 + z_{eq})\Omega_m^{(0)}}} \right], \tag{3.66}$$

where $1 + z_{eq} \equiv \Omega_m^{(0)}/\Omega_r^{(0)}$ is the redshift of matter-radiation equality. Then, including both modifications, the solution becomes

$$(1 + z_k) = (1 + \hat{z}_k) + \delta z_k \tag{3.67}$$

where $\hat{z}_k$ is defined by Eq. (3.66) and

$$\delta z_k = \frac{F'(\hat{z}_k)}{F''(\hat{z}_k)} \left[ -1 + \sqrt{1 - 2\frac{F''(\hat{z}_k)\delta F(\hat{z}_k)}{[F'(\hat{z}_k)]^2}} \right], \tag{3.68}$$

This solution is obtained by Taylor expanding $F(z_k)$ around $\hat{z}_k$ (to 2nd order in $\delta z_k = z_k - \hat{z}_k$), and then solving the equation $F(z_k) = \hat{F}(\hat{z}_k)$ for $\delta z_k$. It is extremely accurate for a wide range of $\Omega_{de}^{(0)}$, $w_{de}(z)$, and $\delta w_r(z)$. Indeed, the simpler 1st-order expression

$$\delta z_k = -\frac{\delta F(\hat{z}_k)}{F'(\hat{z}_k)} \tag{3.69}$$

is often sufficiently accurate.

### 3.3.4 The horizon-crossing factor, $C_2$

In the previous subsection, we treated horizon re-entry as a sudden event that occurs when $k = aH$. In reality, the "outside" behavior ($h_k = $ constant) only holds when $k \ll aH$, and the "inside" behavior ($h_k \propto a^{-1}\cos[k\tau + \text{phase}]$) only holds when $k \gg aH$. In between, when $k \sim aH$, neither behavior holds — *i.e.*, the horizon has a non-zero "thickness."

The behavior of the background equation-of-state $w(\tau) = p(\tau)/\rho(\tau)$ during the period of horizon re-entry is imprinted in the factor $C_2$. For example, let us compute $C_2$ for a mode $k$

that re-enters the Hubble horizon when $w(\tau)$ is varying slowly relative to the instantaneous Hubble rate. Then we can write

$$a = a_0(\tau/\tau_0)^\alpha \quad \text{with} \quad \alpha = \frac{2}{1+3w}, \tag{3.70}$$

so that the equation of motion for $\widetilde{h}_k$

$$\widetilde{h}_k'' + 2(a'/a)\widetilde{h}_k' + k^2\widetilde{h}_k = 0 \tag{3.71}$$

has the solution

$$\widetilde{h}_k(\tau) = h_k(\tau_i)\Gamma(\alpha+1/2)[k\tau/2]^{1/2-\alpha}J_{\alpha-1/2}(k\tau), \tag{3.72}$$

where we have used $k\tau_i \ll 1$ and $\widetilde{h}_k'(\tau_i) = 0$. (Early in the radiation era, the relevant modes were far outside the horizon, and the corresponding mode functions were $\tau$-independent.) We have neglected the spatial curvature, $K$, because the two conditions $K \ll a_0^2H_0^2$ (current observations indicate that the spatial curvature is small) and $k > a_0H_0$ (we are only interested in modes that are already inside the horizon) imply that $K$ produces a negligible correction to the equation of motion for $h_k$.

Once the modes are well inside the horizon ($k\tau \gg 1$), we can use the asymptotic Bessel formula

$$J_{\alpha-1/2}(k\tau) \to \sqrt{\frac{2}{\pi k\tau}}\cos(k\tau - \alpha\pi/2) \tag{3.73}$$

to find

$$\frac{\widetilde{h}_k^2(\tau)}{h_k^2(\tau_i)} = \frac{\Gamma^2(\alpha+1/2)}{\pi}[k\tau/2]^{-2\alpha}\cos^2(k\tau - \alpha\pi/2). \tag{3.74}$$

On the other hand, since a mode $k$ re-enters the horizon ($k = aH = a'/a$) at time $\tau_k = \alpha/k$, we can rewrite Eq. (3.56) for $C_1$ as

$$C_1 = [\tau/\tau_k]^{2\alpha}\cos^2[k(\tau-\tau_k)+\phi_k], \tag{3.75a}$$

$$= [k\tau/\alpha]^{-2\alpha}\cos^2(k\tau - \alpha + \phi_k). \tag{3.75b}$$

Comparing Eqs. (3.52), (3.53), (3.74) and (3.75b), we see that the phase shift $\phi_k$ in Eq. (3.56) is given by

$$\phi_k = [1-\pi/2]\alpha, \tag{3.76}$$

and that $C_2$ is given by

$$C_2(k) = \frac{\Gamma^2(\alpha + 1/2)}{\pi}[2/\alpha]^{2\alpha}, \tag{3.77}$$

where $\alpha$ should be evaluated at horizon re-entry ($k = aH$). In particular, note that

$$\begin{aligned} w = 0 \quad &\Rightarrow \quad C_2(k) = \tfrac{9}{16} \quad \text{and} \quad \phi_k = 2 - \pi, \\ w = \tfrac{1}{3} \quad &\Rightarrow \quad C_2(k) = \ 1 \quad \text{and} \quad \phi_k = 1 - \pi/2. \end{aligned} \tag{3.78}$$

### 3.3.5 The anisotropic-stress damping factor, $C_3$

In this subsection, we will include the effects of the anisotropic stress term $\Pi_k$ on the right-hand side of the tensor mode equation (3.31). A non-negligible tensor anisotropic stress $\Pi_k$ is most naturally generated by relativistic particles free-streaming along geodesics that are perturbed by the presence of tensor metric perturbations $h_k$. In this situation, Weinberg [179] has recently shown that the tensor mode equation (3.31) may be rewritten as a fairly simple integro-differential equation for $h_k$ — see Eq. (18) in [179].

Let us focus on a particularly interesting case: a radiation-dominated universe in which the free-streaming particles constitute a nearly-constant fraction $\Omega_{free}$ of the background (critical) energy density. (Physically, if the free-streaming particles are stable, or long-lived relative to the instantaneous Hubble time at re-entry, then $\Omega_{free}$ will indeed be nearly-constant, as required.) In this case, following an approach that is essentially identical to the one outlined in [46], we write the solution in the form

$$h_k(\tau) = h_k(\tau_i) \sum_{n=0}^{\infty} a_n j_n(k\tau), \tag{3.79}$$

where $j_n(k\tau)$ are spherical Bessel functions, and find the first five non-vanishing coefficients

to be given by

$$a_0 = 1, \tag{3.80a}$$
$$a_2 = \frac{10f}{(15+4f)}, \tag{3.80b}$$
$$a_4 = \frac{18f(3f+5)}{(15+4f)(50+4f)}, \tag{3.80c}$$
$$a_6 = \frac{\frac{130}{7}f(14f^2+92f+35)}{(15+4f)(50+4f)(105+4f)}, \tag{3.80d}$$
$$a_8 = \frac{\frac{85}{343}f(4802f^3+78266f^2+161525f-29400)}{(15+4f)(50+4f)(105+4f)(180+4f)}. \tag{3.80e}$$

$$a_0 = 1, \tag{3.81a}$$
$$a_2 = \frac{10\Omega_{free}}{(15+4\Omega_{free})}, \tag{3.81b}$$
$$a_4 = \frac{18\Omega_{free}(3\Omega_{free}+5)}{(15+4\Omega_{free})(50+4\Omega_{free})}, \tag{3.81c}$$
$$a_6 = \frac{\frac{130}{7}\Omega_{free}(14\Omega_{free}^2+92\Omega_{free}+35)}{(15+4\Omega_{free})(50+4\Omega_{free})(105+4\Omega_{free})}, \tag{3.81d}$$
$$a_8 = \frac{\frac{85}{343}\Omega_{free}(4802\Omega_{free}^3+78266\Omega_{free}^2+161525\Omega_{free}-29400)}{(15+4\Omega_{free})(50+4\Omega_{free})(105+4\Omega_{free})(180+4\Omega_{free})}. \tag{3.81e}$$

The odd coefficients all vanish: $a_{2n+1} = 0$. Keeping these first five non-vanishing terms yields a solution for $h_k(\tau)$ that is accurate to within 0.1% for all values $0<\Omega_{free}<1$. Next, as observed in [46], we can use the asymptotic expression

$$j_{2n}(k\tau) \to (-1)^n \frac{\sin(k\tau)}{k\tau} \quad \text{as} \quad k\tau \to \infty, \tag{3.82}$$

along with the $\Omega_{free} = 0$ solution $\widetilde{h}_k(\tau) = h_k(\tau_i) j_0(k\tau)$ to infer that the tensor anisotropic stress $\Pi_k$ induces no additional phase shift in $h_k$, so that our earlier expression (3.76) for $\phi_k$ is unchanged. (See [11] for a complementary explanation of this null result, based on causality.) In this way, one also sees that $\Pi_k$ damps the tensor power spectrum by the asymptotic factor

$$C_3 = |A|^2, \tag{3.83}$$

where

$$A = \sum_{n=0}^{\infty} (-1)^n a_{2n}. \tag{3.84}$$

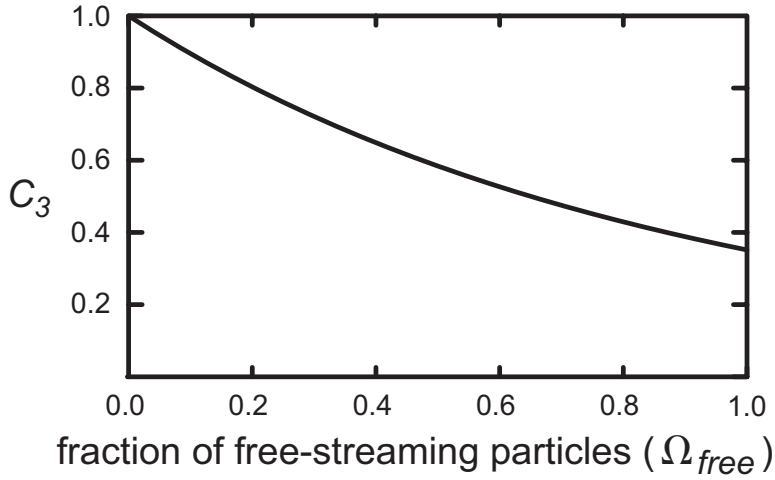



Figure 3.2: $C_3$ is the transfer function factor that accounts for the damping of the tensor power spectrum due to tensor anisotropic stress. The factor depends on the fraction $\Omega_{free}$ of the background (critical) energy density contained in free-streaming relativistic particles. The figure plots this dependence for $0 < \Omega_{free} < 1$.

For example, keeping the first 4 terms in this sum, we find an approximate expression for $A$:

$$\frac{-\frac{10}{7}(98\Omega_{free}^3 - 589\Omega_{free}^2 + 9380\Omega_{free} - 55125)}{(15+4\Omega_{free})(50+4\Omega_{free})(105+4\Omega_{free})}, \tag{3.85}$$

which is accurate to within 1% for all values $0<\Omega_{free}<1$. If we keep the first 5 terms in the sum, we find an even better approximation for $A$:

$$\frac{15(14406\Omega_{free}^4 - 55770\Omega_{free}^3 + 3152975\Omega_{free}^2 - 48118000\Omega_{free} + 324135000)}{343(15+4\Omega_{free})(50+4\Omega_{free})(105+4\Omega_{free})(180+4\Omega_{free})} \tag{3.86}$$

which is accurate to within 0.1% for all values $0<\Omega_{free}<1$. These calculations improve on the accuracy of previous calculations [11, 134].

The exact dependence of $C_3$ on $\Omega_{free}$ is shown in Fig. 3.2. Note that, as $\Omega_{free}$ varies between 0 and 1, the damping factor $C_3$ varies between 1.0 and 0.35. In particular, if we substitute $\Omega_{free} = 0.4052$, corresponding to 3 standard neutrino species,the damping factor 0.80313 agrees with the results of numerical integrations [179, 134]. When the modes probed by the CMB re-enter the horizon, the temperature is relatively low (corresponding to atomic-physics energies), so we are fairly confident that neutrinos are the only free-

streaming relativistic particles. But when the modes probed by laser interferometers re-enter the horizon, the temperature is much higher (above the electroweak phase transition, $T \sim 10^7$ GeV), so that the physics (and, in particular, the instantaneous free-streaming fraction $\Omega_{free}$) is much more uncertain. Thus, laser interferometers offer the possibility of learning about the free-streaming fraction $\Omega_{free}$ in the very early universe at temperatures between the inflationary and electroweak symmetry breaking scales.

Finally, although Weinberg and subsequent authors have concentrated on the tensor anisotropic stress due to a single fermionic species (the neutrino), it is straightforward to generalize the analysis to include a combination of species which (*i*) may each decouple at a different time and temperature, and (*ii*) may be an arbitrary mixture of bosons and fermions. We find that, as long as all of these free-streaming species decouple well before the modes of interest re-enter the horizon, then all of the results presented in this section are completely unchanged. In other words, in order to determine the behavior of the tensor mode function, one only needs to know one number — the total fraction $\Omega_{free}$ of the critical density contained in free-streaming particles — even if the particles are a mixture of fermionic and bosonic species with different temperatures and decoupling times.

### 3.3.6 Equation-of-state corrections, $\delta w_r$

In this subsection, we consider various physical effects that cause the equation of state $w_r(z)$ to deviate from 1/3 during the radiation-dominated epoch, and the corresponding modifications that these effects induce in the GWB transfer function. Some of these effects have been discussed previously by Seto and Yokoyama [147]. The deviations

$$\delta w_r(z) = w_r(z) - 1/3 \tag{3.87}$$

primarily modify the transfer function through the redshift factor $(1 + z_k)$ that appears in $C_1$ [see Eqs. (3.68) and (3.69)]; through the horizon-crossing factor $C_2(k)$ [see Eq. (3.77)]; and through the phase shift $\phi_k$ [see Eq. (3.76)]. We consider here six physical effects which can produce these kinds of modifications of the transfer function.

First, deviations can be caused by mass thresholds in the early universe. Suppose that all particle species are described by equilibrium distribution functions. Then we can write $\rho$ and $p$ as

$$\rho = \frac{1}{2\pi^2}\sum_i g_i T_i^4 \int_{x_i}^{\infty} \frac{(u^2-x_i^2)^{1/2}}{\exp[u-y_i]\pm 1} u^2 du, \tag{3.88a}$$

$$p = \frac{1}{6\pi^2}\sum_i g_i T_i^4 \int_{x_i}^{\infty} \frac{(u^2-x_i^2)^{3/2}}{\exp[u-y_i]\pm 1} du, \tag{3.88b}$$

where the $i$th species (with mass $m_i$, and $g_i$ internal degrees of freedom) is described by temperature $T_i$ and chemical potential $\mu_i$, and we have defined the dimensionless quantities $x_i \equiv m_i/T_i$ and $y_i \equiv \mu_i/T_i$ [100]. In the denominator, the + and − signs are for fermions and bosons, respectively. Then the deviation $\delta w_r$ is given by the exact expression

$$\delta w_r = \sum_i \delta w_r^{(i)} \tag{3.89}$$

where

$$\delta w_r^{(i)} = -\frac{5}{\pi^4}\frac{g_i}{g_{*\rho}}\frac{T_i^4}{T^4} f(x_i, y_i) \tag{3.90}$$

represents the contribution from the $i$th species,

$$g_{*\rho} \equiv \sum_i g_i \frac{T_i^4}{T^4}\frac{15}{\pi^4}\int_{x_i}^{\infty} \frac{(u^2-x_i^2)^{1/2}u^2}{\exp[u-y_i]\pm 1} du \tag{3.91}$$

represents the effective number of relativistic degrees of freedom, $T$ is conventionally chosen to be the photon temperature, and we have defined the function

$$f(x,y) \equiv x^2 \int_x^{\infty} \frac{(u^2-x^2)^{1/2}}{\exp[u-y]\pm 1} du. \tag{3.92}$$

For fixed $y_i$, note that $f(x_i, y_i)$ vanishes as $x_i$ goes to 0 or $\infty$; and in between it has a fairly broad peak, with a maximum located at $x_i^{peak}$, and a peak value $f_i^{peak} = f(x_i^{peak}, y_i)$. In particular, when $y_i = 0$, then the ordered pair $(x_i^{peak}, f_i^{peak})$ is $(2.303, 1.196)$ for bosons and $(2.454, 1.125)$ for fermions. This makes sense: we expect $\delta w_r^{(i)}$ to vanish when $x_i \ll 1$ (since the species is relativistic) and when $x_i \gg 1$ (since the species is non-relativistic, and makes a negligible contribution to the energy density). In between, when $x_i \sim x_{i,peak}$, the

$i$th species is cold enough to exhibit non-relativistic behavior, yet hot enough to contribute non-negligibly to the energy density.

Using the above equations, we can compute $\delta w_r(z)$ once we know $T_i(z)$ and $\mu_i(z)$. But let us estimate the size of the effect. As a species becomes non-relativistic, it produces a maximum equation-of-state deviation

$$\delta w_r^{(i)} = -\frac{5 f_i^{peak}}{\pi^4} \frac{g_i}{g_{*\rho}} \frac{T_i^4}{T^4} \tag{3.93}$$

in the background equation of state. Furthermore, if $N_s$ different species (with the same temperature and similar masses) become non-relativistic at the same time, then (roughly speaking) the effect is multiplied by $N_s$ (since their $\delta w_r^{(i)}$'s add). Ultimately, the fractional correction $\delta w_r / w_r$ is model-dependent, but it can conceivably be as large as a few percent.

Second, deviations can be produced by a trace anomaly in the early universe. During the radiation-dominated epoch, the universe is dominated by highly relativistic particles whose masses may be neglected. Thus, each species is governed by a classical action that is conformally invariant at the classical level, leading to the usual conclusion that the stress-energy tensor is traceless and $w_r = 1/3$. But conformal invariance is broken at the quantum level by interactions among the particles, so that $T^\mu_\mu \neq 0$. For example, for a quark-gluon plasma governed by SU($N_c$) gauge theory, with $N_f$ flavors, and gauge-coupling $g$, the equation of state correction is given (up to $O(g^5)$ corrections) by [81, 43]

$$\delta w_r = \frac{5}{18\pi^2} \frac{g^4}{(4\pi)^2} \frac{(N_c + \frac{5}{4}N_f)(\frac{2}{3}N_f - \frac{11}{3}N_c)}{2 + \frac{7}{2}[N_c N_f/(N_c^2 - 1)]}. \tag{3.94}$$

Note that this effect can be non-negligible: for large gauge groups (*i.e.* large $N_c$) in the early universe (prior to the electroweak phase transition), the equation of state $w_r$ may easily be reduced from 1/3 by several percent, or more.

Third, deviations can be produced if the early universe behaves like a slightly imperfect fluid. The stress-energy tensor for an imperfect fluid contains (in addition to the usual perfect-fluid terms) three extra terms whose coefficients ($\chi$, $\eta$, and $\zeta$) represent heat conduction, shear viscosity, and bulk viscosity (see Weinberg [176], Ch. 2.11). Of these dissipative

effects, only the bulk viscosity term

$$\Delta T^{\mu\nu} = -\zeta(g^{\mu\nu} + U^{\mu}U^{\nu})U^{\lambda}_{;\lambda} \tag{3.95}$$

can contribute to the background evolution in an FRW universe (see Weinberg [176], Chs. 15.10-15.11). This term modifies the continuity equation

$$\dot{\rho} = -3H(\rho + p) + 9\zeta H^2 = -3H\rho\Big[1 + w - \frac{8\pi G\zeta}{H}\Big] \tag{3.96}$$

so that, as far as gravitational waves are concerned, the effective equation is corrected by

$$\delta w_r = -\frac{8\pi G\zeta}{H}. \tag{3.97}$$

Whereas the three effects discussed thus far produce small corrections to the equation of state, it is worth mentioning three other effects that can produce much larger deviations. The first example is a massive particle species that decouples from the thermal plasma before its abundance becomes negligible. Since its energy density falls as $a^{-3}$ (more slowly than the radiation density, which falls as $a^{-4}$), it can come to dominate the energy density of the universe before it decays (if its lifetime $\tau_{decay}$ is sufficiently long). In this case, $w$ drops to zero when the particle becomes dominant, and rises back to $w = 1/3$ over a timescale given by the decay lifetime $\tau_{decay}$. Second, extra-dimensional physics typically modifies the effective 4-dimensional Friedmann equation. Such modifications which, from the standpoint of the GWB, can in some cases look like corrections to the effective radiation equation of state, are primarily constrained by the requirement that the Friedmann equation becomes sufficiently similar to the ordinary 4-dimensional Friedmann equation (with ordinary matter) by the time of BBN. Third, due to the weak coupling of the inflaton, the temperature at the start of the radiation dominated epoch (the reheat temperature) can be much lower than the energy scale at the end of inflation. In this case, laser interferometer scales might actually re-enter the horizon during the reheating epoch (before the start of radiation domination), when the equation of state was probably quite different from $w = 1/3$. The actual equation-of-state depends on the details of the reheating process, but a commonly-discussed value is $w = 0$, or some value in the range $0 < w < 1/3$ [131]. If $w = 0$ during reheating,

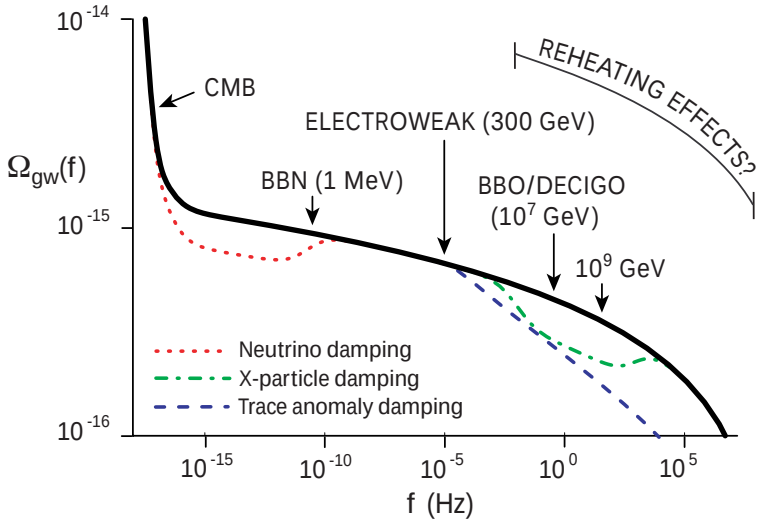



Figure 3.3: The black solid curve represents the present-day gravitational-wave energy spectrum, $\Omega_{gw}(f, \tau_0)$, for the inflationary model $V(\phi) = (1/2)m^2\phi^2$. The red dotted curve shows the damping effect due to (three ordinary massless species of) free-streaming neutrinos. The green dot-dashed curve shows the damping effect which arises if free-streaming particles make up fifty percent of the background energy density at the time $\tau_{BBO}$ when the modes probed by BBO/DECIGO re-enter the horizon. As shown in the figure, the particles begin free-streaming sometime before $\tau_{BBO}$, and decay sometime after $\tau_{BBO}$, but prior to electroweak symmetry breaking. Finally, the blue dashed curve shows the effect of a conformal anomaly in the early universe that slightly reduces the equation of state from $w = 0.33$ to $w = 0.31$ above the electroweak phase transition. The spectrum will also be modified on comoving scales that re-enter the horizon during the reheating epoch after inflation; but the range of scales affected by reheating is unknown. Finally, note that the correlated BBO interferometer proposal claims a sensitivity that extends beyond the bottom of the figure (down to roughly $\Omega_{gw} \sim 10^{-17}$) in the frequency range from $10^{-1}$ Hz to $10^{0}$ Hz.

the corresponding modification of the GWB might be similar to the modification due to the long-lived massive relic discussed above.

Note that the first of these six processes can be expressed as a modification of $g_*$, and the effects on the transfer function can be computed using the methods discussed in Ref. [148]. However, the other five cannot.

### 3.3.7 Transfer function summary

Fig. 3.3 illustrates some of the transfer-function effects discussed in this section. In this figure, the solid black curve represents the present-day energy spectrum, $\Omega_{gw}(f, \tau_0)$, generated by a particular inflationary model — namely, a quadratic potential $V(\phi) = (1/2)m^2\phi^2$. The red dotted curve illustrates the damping due to tensor anisotropic stress from free-streaming neutrinos. We have assumed that the free-streaming fraction is $\Omega_{free} = 0.4052$, which is the $\Omega_{free}$ value for three standard neutrino species which decouple around the time of BBN. The green dot-dashed curve represents the damping due to tensor anisotropic stress from various particle species ($X$ particles) which begin free-streaming before the scales detected by BBO/DECIGO re-enter the horizon and then decay after the scales re-enter, but prior to the electroweak phase transition. As an example, we have assumed that the free-streaming fraction is $\Omega_{free} = 0.5$. Finally, the blue dashed curve represents damping due to a trace anomaly that is present above the electroweak scale. For illustration, we have assumed that this anomaly, through Eq. (3.94), reduces the equation of state from $w_r = 1/3$ by $\delta w_r = -0.02$. This reduction may be achieved by various combinations of the number of colors $N_c$, the number of flavors $N_f$, and the gauge coupling $g$; but the point is that we have not chosen an unreasonable large value for $\delta w_r$, given the large gauge groups that are often theorized to be present at high energies.

The key point conveyed by Fig. 3.3 is that there are a variety of plausible post-inflationary effects that can produce rather large modifications of the gravitational-wave spectrum on laser-interferometer scales, without modifying the spectrum on CMB scales. This is tantalizing, since the modifications on laser-interferometer scales reflect the primordial dark age between the end of inflation and the electroweak phase transition, at energies beyond the reach of terrestrial particle accelerators.

## 3.4 Discussion

In this section, we will explore various issues and opportunities that arise from combining a CMB experiment like CMBPOL and a laser-interferometer experiment like BBO. We will argue that, if gravitational waves are detected in the CMB, then the science case for a laser-interferometer satellite mission like BBO becomes strong, since it is then likely to yield (in combination with the CMB) qualitatively new information about the very early universe that cannot be obtained by other known techniques.

This section has two parts. In the first part, we will consider the scientific possibilities that arise from combining CMBPOL and BBO, assuming we can accurately compute the tensor transfer function on BBO scales — *i.e.* assuming that we can neglect as "exotic" the various high-energy effects discussed in the previous section that can leave an imprint in the tensor transfer function on LI scales at late times. Then, in the second part, we will consider the implications of these uncertainties in the transfer function.

To be concrete, let us imagine a future CMB experiment (which we will call CMBPOL) that can achieve one-sigma error bars of $\delta n_s = 0.005$ for the scalar spectral tilt and $\delta r = 0.005$ for the tensor-to-scalar ratio (so that, roughly speaking, it can detect $r = 0.01$ at 95% confidence). Let us also imagine a future LI experiment that can reach the BBO projections [129] for both experimental sensitivity and astrophysical foreground subtraction. In fact, the BBO mission proposal comes in three different increasingly sensitive (and increasingly expensive) flavors — "BBO-lite," "BBO-standard," and "BBO-grand." These three versions would detect a stochastic gravitational-wave background with signal-to-noise ratio (SNR) [145]

$$\mathrm{SNR} = 0.103 \left(\frac{\Omega_{gw}}{10^{-16}}\right) \left(\frac{T_{\mathrm{obs}}}{10\mathrm{yr}}\right)^{1/2} \qquad (\mathrm{BBO-lite}), \tag{3.98a}$$

$$\mathrm{SNR} = 25.12 \left(\frac{\Omega_{gw}}{10^{-16}}\right) \left(\frac{T_{\mathrm{obs}}}{10\mathrm{yr}}\right)^{1/2} \qquad (\mathrm{BBO-standard}), \tag{3.98b}$$

$$\mathrm{SNR} = 251.2 \left(\frac{\Omega_{gw}}{10^{-16}}\right) \left(\frac{T_{\mathrm{obs}}}{10\mathrm{yr}}\right)^{1/2} \qquad (\mathrm{BBO-grand}), \tag{3.98c}$$

where $\Omega_{gw}$ is the amplitude of the gravitational-wave energy spectrum (from inflation) in

the frequency band 0.1 Hz $< f <$ 1 Hz. Unfortunately, the cheapest version, BBO-lite, is simply not sensitive enough to detect the inflationary gravitational wave background, so we will focus on BBO-standard, and assume that the observation period is $T_{\rm obs} = 5$yr. With these basic parameters in mind, let us consider the science that these experiments can be expected to accomplish.

The first important (although rather pedestrian) motivation for BBO-standard is that it can provide a cross-check on any claimed gravitational-wave detection in the CMB. After all, CMBPOL and BBO-standard are both expected to be difficult experiments. If we only had one experiment or the other, we might worry that any claimed detection of gravitational-waves (especially at low significance) was really a systematic error — an unexpected or exceptionally difficult foreground or noise source that is not treated correctly in the analysis. But, as we have seen in section 2, once we manage to detect a gravitational-wave signal in the CMB, we can immediately obtain a rough estimate for the amplitude of the gravitational-wave spectrum on LI scales. And if BBO measures the amplitude, and confirms this rough expectation, we will be much more confident that both experiments are actually seeing the primordial signal from inflation.

The second motivation for BBO-standard is that, in combination with CMBPOL, it can demonstrate at high significance that $n_t$ is non-zero and negative, thus providing qualitative confirmation of the first inflationary consistency relation. (Recall that the first inflationary consistency relation, $n_t = -r/8$, implies that if the tensor amplitude is non-negligible, then the tensor tilt should also be non-negligible and negative.) To see that this is true, let us return to the four examples studied in section 2.

- If the inflaton potential is $V(\phi) = \frac{1}{4!}\lambda\phi^4$ — a model that is still in agreement, at least marginally, with the current WMAP data [153] — then CMBPOL will measure $r = 0.268 \pm 0.005$ at $k_*$. If the primordial tensor power spectrum were perfectly scale invariant, this would imply $\Omega_{gw} = (23.5 \pm 0.44) \times 10^{-16}$ on BBO scales. But, instead, BBO-standard will measure $\Omega_{gw} = (3.51 \pm 0.06) \times 10^{-16}$, where the error bar here is determined by the formula (3.98b) for the signal-to-noise ratio for BBO-standard.

In other words, the observed value will be below the scale-invariant prediction by 45-sigma! — a very strong demonstration that the primordial tensor power spectrum has a non-vanishing negative tilt $n_t$.

- If the inflaton potential is $V(\phi) = \frac{1}{2}m^2\phi^2$ — a model in good agreement with current WMAP data — then CMBPOL will measure $r = 0.138 \pm 0.005$ at $k_*$. If the primordial tensor power spectrum were perfectly scale invariant, this would imply $\Omega_{gw} = (12.1 \pm 0.44) \times 10^{-16}$ on BBO scales. But, instead, BBO-standard will measure $\Omega_{gw} = (4.56 \pm 0.06) \times 10^{-16}$. The observed value will thus be below the scale-invariant value by more 17-sigma — again, a strong demonstration that the tensor spectrum has a non-vanishing negative tilt.

- If the inflaton potential is of the symmetry-breaking form $V(\phi) = V_0[1 - (\phi/\mu)^2]^2 = V_0 - m^2\phi^2 + \lambda\phi^4$, with $n_s = 0.96$ — another model in good agreement with the current WMAP data — then CMBPOL will measure $r = 0.046 \pm 0.005$ at $k_*$. If the primordial tensor power spectrum were perfectly scale invariant, this would imply $\Omega_{gw} = (3.99 \pm 0.44) \times 10^{-16}$ on BBO scales. But, instead, BBO-standard will measure $\Omega_{gw} = (2.52 \pm 0.06) \times 10^{-16}$. The observed value will thus be below the scale-invariant value by 3.3-sigma. Although much less strong than the previous two examples, this would still constitute moderately significant evidence that the tensor spectrum has a non-vanishing negative tilt.

- If the inflaton potential is of the axion-like cosine (or "natural inflation") form $V(\phi) = V_0[1 + \cos(\phi/\mu)]$ with $n_s = 0.95$ — yet another model in good agreement with the current WMAP data — then CMBPOL will measure $r = 0.033 \pm 0.005$ at $k = k_*$. If the primordial tensor power spectrum were perfectly scale invariant, this would imply $\Omega_{gw} = (2.92 \pm 0.44) \times 10^{-16}$ on BBO scales. But, instead, BBO-standard would measure $\Omega_{gw} = (1.91 \pm 0.06) \times 10^{-16}$. The observed value would thus be below the scale-invariant value by only about 2.3-sigma, corresponding to marginal (but not really significant) evidence for a negative tensor tilt.

Note that these examples illustrate the easily-understood point that, when $r$ on CMB scales is larger, the combination of CMBPOL and BBO-standard is able to show $n_t < 0$ with much higher significance. In particular, if $r \gtrsim 0.05$ on CMB scales, then the combination of BBO-standard and CMBPOL (with the experimental parameters described above) should be able to demonstrate convincingly that the tensor tilt $n_t$ is non-zero and negative, in agreement with the first inflationary consistency relation. This would be a major scientific achievement — analogous to, but independent from, the demonstration that the scalar tilt $n_s - 1$ is non-zero and negative (which is another inflationary prediction that has recently been confirmed, at marginal significance, by WMAP [153]). On the other hand, if $r \lesssim 0.05$ on CMB scales, then the combination of BBO-standard and CMBPOL described above would only be expected to provide marginal or insignificant evidence for $n_t < 0$. Notice, though, that this conclusion is ultimately determined by the size of the CMB error bar $\delta r$. Thus, if it is possible to make the CMBPOL one-sigma error bar significantly smaller than $\delta r = 0.005$, then it will also be possible to demonstrate non-vanishing negative tensor tilt even when $r$ is significantly smaller than $r \approx 0.05$ on CMB scales.

The third motivation for BBO-standard is that, in combination with CMBPOL, it can check at high significance whether or not the the gravitational-wave signal on BBO scales agrees with the quantitative prediction of the inflationary consistency relations. To see this, let us return to our four examples. Suppose that CMBPOL measures $\{n_s - 1 = -0.05 \pm 0.005, r = 0.268 \pm 0.005\}$, or $\{n_s - 1 = -0.034 \pm 0.005, r = 0.138 \pm 0.005\}$, or $\{n_s - 1 = -0.04 \pm 0.005, r = 0.046 \pm 0.005\}$, or $\{n_s - 1 = -0.05 \pm 0.005, r = 0.033 \pm 0.005\}$, corresponding to the predictions of our four inflaton potentials. Then, using the extrapolation function $E_t^{(2)}$ from section 2 (which ultimately is derived from the inflationary consistency relations), we would estimate $\Omega_{gw}$ on BBO scales to be $\approx 4.8 \times 10^{-16}$, or $\approx 5.3 \times 10^{-16}$, or $\approx 2.9 \times 10^{-16}$, or $\approx 2.2 \times 10^{-16}$, respectively, where the first prediction is uncertain up to a factor of $\sim 1.5$, while the subsequent predictions are uncertain up to a factor of $\sim 1.3$. (Here we have included both types of extrapolation uncertainty discussed in section 2 and, as discussed above, neglected any uncertainties coming from high-energy

effects in the tensor transfer function.) If $\Omega_{gw}$ turns out to lie in the predicted range, then BBO-standard will detect it with high signal-to-noise ratio (SNR $\approx 85$, or $\approx 94$, or $\approx 51$, or $\approx 39$, respectively), in support of the consistency relations. On the other hand, BBO-standard could clearly reject the consistency relation by finding that $\Omega_{gw}$ lies either above or below the predicted value by a factor significantly larger than the uncertainty factor. (Note that, if we take the SNR $= 3$ as the detection threshhold, then BBO-standard can detect $\Omega_{gw}$ smaller than the prediction of the consistency relations by up to a factor of $\approx 27$, or $\approx 29$, or $\approx 16$, or $\approx 12$, respectively.) This shows that if $r$ is sufficiently large on CMB scales, then by combining CMBPOL and BBO-standard, we can check at high significance whether or not $\Omega_{gw}$ lies in the range predicted by the inflationary consistency relations on BBO scales. And, as pointed out above, if $\Omega_{gw}$ turns out to lie in this range, it will be distinguishable from the "trivial" prediction of a purely scale-invariant primordial tensor power spectrum. Taken together, this means that CMBPOL and BBO-standard offer the possibility of subjecting the inflationary consistency relations to a highly non-trivial test. Also, note that most of the extrapolation uncertainty in the cases considered above is of the "first type" (in the terminology of section 2). As mentioned in section 2, it seems that it will be possible to further reduce this first type of extrapolation uncertainty by deriving an improved extrapolation function, and this would make for an even sharper test of the consistency relations — but this is still work in progress.

Thus far in this section, we have neglected uncertainties coming from the imprint of high-energy (early-universe) physics in the transfer function at late times. Let us now consider the implications of the additional uncertainties for our interpretation of the experiments. Assuming the null hypothesis described in the introduction of this chapter (a fiducial "standard" tensor transfer function plus a standard extrapolation function derived from the inflationary consistency relations in section 2) we have argued that we can convert CMB observations of $r$ and $n_s$ (and perhaps $dn_s/d\ln k$) into a predicted range for $\Omega_{gw}$ on BBO scales. And we have also seen that if $r$ is large enough to be detected by CMBPOL, then the null-hypothesis prediction for $\Omega_{gw}$ is large enough to be detected by BBO-standard.

Now let us consider four possible outcomes of the BBO-standard experiment, and discuss their likely interpretations (including the uncertainties in the tensor transfer function).

First, suppose that BBO-standard detects $\Omega_{gw}$ in the expected range. It would be most natural to interpret this as evidence in favor of the inflationary consistency relations and the fiducial tensor transfer function. Or, more correctly, it would suggest upper bounds on the size of any violations of the consistency relations, or on the size of any high-energy effects modifying the tensor transfer function on BBO scales. Of course, it is theoretically possible that large violations of the consistency relations and large effects in the transfer function happen to cancel to give the same answer as the null prediction, but this interpretation would be less plausible.

Second, suppose that BBO detects the gravitational-wave amplitude significantly *below* its expected value. Then the interpretation is less straighforward. On the one hand, we could interpret the suppression as being due to transfer-function effects, such as those discussed in the previous section. This interpretation would be interesting, since it would imply a rare opportunity to measure physical properties of the early universe, at temperatures above the electroweak scale, when the relevant modes re-entered the horizon. On the other hand, we could interpret the suppression to mean that inflation is more complicated than we expected: perhaps there is a feature in the inflaton potential which suppresses the primordial tensor spectrum on small scales relative to our expectations, or perhaps the consistency relations are violated. Additional information would be necessary to distinguish these two possibilities. But, in any case, this scenario clearly corresponds to a qualitatively new clue about the early universe – even if it is one that cannot be uniquely interpreted on its own.

Third, suppose that BBO detects the gravitational-wave amplitude significantly *above* its expected value. At this point it is worth noting that virtually all of the transfer-function effects mentioned in section 3.3.2 (including tensor anisotropic stress, the late decay of a massive relic species, bulk viscosity, a conformal anomaly, or standard reheating with $w \leq 1/3$) *suppress* the gravitational-wave spectrum on BBO scales. If we want to use the transfer function to *enhance* the spectrum on laser-interferometer scales, we must invoke an even

more exotic effect, such as a reheating epoch with a anomalously low reheat temperature and an unusual equation of state ($w > 1/3$), or perhaps some sort of extra-dimensional physics in the early universe. But perhaps a more plausible explanation for a higher-than-expected gravitational-wave signal is that we are detecting gravitational waves produced after inflation, rather than the inflationary background itself. One possible post-inflationary source would be cosmic strings of some sort. Another source would be bubble collisions after a first order phase transition — perhaps the electroweak phase transition. But note that a first-order electroweak phase transition is already ruled out in the standard model, and close to being ruled out in the MSSM, from experimental bounds on the Higgs mass. Yet another possible source would be gravitational waves produced during preheating after inflation [90, 52] — but note that for the inflationary and preheating models studied thusfar, the gravitational waves from preheating are at higher frequencies, and do not fall into the BBO waveband.

Fourth, suppose that BBO detects nothing at all (after a detection of $r > 0.01$ by CMBPOL). This would most likely indicate a fundamental problem with inflation itself, since an period of accelerating expansion produces a broad and nearly-flat spectrum of primordial tensor perturbations quite generically (regardless of the particular model that drives inflation, and regardless of whether the consistency relations are quantitatively satisfied). One way out of this conclusion would be to invoke an extreme suppression effect in the transfer function on BBO scales — *e.g.* due to a reheating epoch with equation of state $w = 0$ and a reheat temperature well below $10^7$ GeV. In this case, determining which interpretation is correct would require more information — perhaps by digging deeper, with more sensitive experiments on BBO scales, or perhaps from some complementary experimental technique.

Finally, we should keep in mind that certain observations could give us even more information than we have assumed above, and could thereby break some of the interpretational degeneracies discussed in the preceding few paragraphs. For example, note that the quadratic potential well $V(\phi) = \frac{1}{2}m^2\phi^2 + \ldots$, in addition to being one of the simplest

models of inflation, is also the generic form of an inflaton potential near the bottom of its potential well. Thus, if CMBPOL measured $n_s = -0.034 \pm 0.005$ and $r = 0.138 \pm 0.005$, in perfect agreement with the predictions of $V(\phi) = \frac{1}{2}m^2\phi^2$ inflation, it would be very tempting to interpret this to mean that, during the final 60 e-folds, $\phi$ is close enough to its potential minimum that $V(\phi)$ may be approximated by $\frac{1}{2}m^2\phi^2$. In this case, we wouldn't need to use the extrapolation function, since we could compute the primordial tensor power spectrum directly from the potential. Then, since there would be virtually no extrapolation uncertainty, if $\Omega_{gw}$ was below its expected value on BBO scales, we could interpret this as telling us direct information about the tensor transfer function, and hence about the early universe at $T \sim 10^4$ TeV.

# Chapter 4

# An expansion/contraction duality in ordinary FRW cosmology

This chapter is based in part on the paper [24], a collaboration with Paul Steinhardt and Neil Turok. I would like to thank Alexei Starobinsky for pointing out helpful references.

## 4.1 Introduction

In inflationary cosmology [74, 107, 4], a nearly scale-invariant spectrum of gauge-invariant Newtonian potential perturbations is produced as comoving scales leave the Hubble horizon during an early burst of accelerated expansion [122, 73, 75, 156, 10]. In cyclic cosmology [93, 91, 159, 158], the *same* spectrum of Newtonian potential perturbations is produced as comoving scales leave the Hubble horizon during a period of slow decelerated contraction [94, 163, 164]. This agreement between two physically dissimilar models is unexpected, but not coincidental. As we shall show, the relationship between inflation and the cyclic model may be viewed as a special case of a surprisingly simple and general duality between expanding and contracting cosmologies.

A Friedmann-Robertson-Walker (FRW) universe with a single scalar field $\phi$ and potential $V(\phi)$ is a simple yet important system. In particular, it is the canonical 4d effective

theory used to model the production of density perturbations in both inflationary and cyclic cosmology. Recent results hint at a connection between two apparently unrelated regimes of this model: $(i)$ expanding, in the $\epsilon \to 0$ limit, and $(ii)$ contracting, in the $\epsilon \to \infty$ limit. Here $\epsilon \equiv 3(1+w)/2$, where $w \equiv p/\rho$ denotes the ratio of pressure to energy density. These two regimes, $(i)$ and $(ii)$, both generate a nearly scale invariant power spectrum of fluctuations for the gauge invariant Newtonian potential $\Phi$ [69, 130] in the long wavelength limit. Furthermore, the slight tilt $n_s^{(i)}$ of the long-wavelength $\Phi$-spectrum produced by a model in regime $(i)$ is equal to the slight tilt $n_s^{(ii)}$ produced by a model in regime $(ii)$, provided the corresponding $\epsilon$-parameters satisfy the simple relation $\epsilon^{(i)} = 1/\epsilon^{(ii)}$ [95].

The relationship noted in [69, 95] between expanding $\epsilon \ll 1$ models and contracting $\epsilon \gg 1$ models, turns out to be a special case of a general and exact duality relating expanding and contracting models with identical perturbation spectra. In this chapter, we derive this duality, initially focusing on the case where $\epsilon$ (or $w$) is nearly constant. Later, we discuss how the duality may be generalized to include backgrounds with time-dependent $\epsilon$. When $\epsilon$ is constant, or varies sufficiently slowly, the duality is simple: an expanding universe characterized by $\epsilon$ produces *exactly* the same scalar perturbations as a contracting universe characterized by $\widehat{\epsilon} = 1/\epsilon$. This duality applies in arbitrary spacetime dimension (not just 3+1); it applies for all $w$ (not just the $w \to -1$ and $w \to \infty$ limits discussed in [69, 95]); it applies to all wavelengths (not just the long-wavelength limit); it applies to *both* the dominant scalar perturbation mode *and* a subdominant remainder (which are related to the growing and decaying modes, respectively).

This duality is of general theoretical interest since it provides a new relationship between expanding and contracting universes, and exposes an unexpected symmetry of cosmological perturbation theory. It is also relevant to cosmological models, like the ekpyrotic and cyclic [93, 91, 159, 158] scenarios, in which perturbations produced during a period of contraction are proposed to propagate through a bounce into a subsequent expanding phase. These models require that the growing-mode long-wavelength perturbation spectrum is preserved across the bounce. This has been a controversial matter. At first, some authors argued

that growing-mode perturbations produced in a contracting phase must match to pure decaying-mode perturbations, and vice versa, as one follows them across a bounce into an expanding phase [26, 116, 117, 115]. At heart, this conclusion followed from an assumption that the bounce corresponds to a comoving or constant-energy-density slice across which the curvature perturbation is conserved. The curvature perturbation is blue and decaying before the bounce and, under this assumption, would be blue and growing after the bounce. This situation, which applies to some non-singular bouncing models discussed in [33, 65], is grossly inconsistent with observations. However, recent calculations [163, 164, 12] indicate that comoving or constant-energy-density slices are inappropriate for matching if the bounce event corresponds to a collision between orbifold planes along an extra spatial dimension, as in the ekpyrotic and cyclic models, because the collision is not synchronous on comoving slices. When the matching is performed on the appropriate collision-synchronous slices instead, the curvature perturbation is not conserved across the bounce but, instead, matches to a linear combination of the incoming (blue) curvature perturbation and the nearly scale-invariant Newtonian potential fluctuation. At long wavelengths, the nearly scale-invariant Newtonian potential contribution dominates. Hence, the shape of the curvature perturbation spectrum in the expanding phase (as measured by fluctuations in the cosmic microwave background, for example) is simply equal to the shape of the Newtonian potential spectrum in the contracting phase analyzed in this paper. Other aspects of the cyclic and ekpyrotic models have been criticized [82, 83] (see [92] for replies), and some have argued that a bounce is impossible altogether [77, 110, 111, 113]. While the consistency of a bounce remains to be proven, recent work has shown that the traditional hazard of chaotic mixmaster behavior is strongly suppressed in the contracting phase of the cyclic model [59, 183]. The metric perturbations exhibit ultralocal behavior in which anisotropies remain small, right up to a few Planck times before the bounce. In this situation, causality suggests that the bounce should not disturb correlations on macroscopic scales over which there can be no communication in this finite time interval. Under these conditions, works by several groups [163, 164, 12, 50, 41, 34] suggest that perturbations generated during

the contracting phase may pass into the expanding phase. The subject continues to be an area of active research. If the latter suggestions are made rigorous, it would give added significance to the results presented herein.

Other dualities have been identified in the literature [175, 155, 28, 172, 167, 144, 37, 38, 106, 27, 30, 31] that relate cosmological solutions and perturbations. See Section 4.6 for a comparison. The duality presented here has the distinctive property that it relates two solutions that are stable under perturbations. Hence, both solutions can plausibly play a role in realistic cosmological models.

The layout of this chapter is as follows. In section 4.2, we highlight a few points about the background (unperturbed) model: a spatially-flat FRW model with scalar field $\phi$, potential $V(\phi)$, and constant $\epsilon$. In section 4.3, we highlight relevant aspects of scalar perturbation theory, before deriving exact solutions for Mukhanov's $u$ and $v$ variables. We note that $u$ is invariant under $\epsilon \to 1/\epsilon$. This invariance is independent of our vacuum choice for the fluctuations, and provides our first glimpse of the duality. In section 4.4 we use $u$ and $v$ to separate scalar perturbations into pieces that are dominant and subdominant at long wavelengths, and show that each piece is *independently invariant* under $\epsilon \to 1/\epsilon$. We show how the dominant and subdominant pieces relate to the scalar perturbation growing and decaying modes. In section 4.5 we consider tensor perturbations. In section 4.6, we contrast our duality with other kinds of cosmological dualities that have been studied in the literature [175, 155, 28, 172, 167, 144, 37, 38, 106, 27, 30, 31]. In section 4.7, we interpret the duality geometrically, as a relation between the scale factor and the Hubble parameter. In section 4.8, we generalize the duality to encompass spacetimes with $d$ dimensions. We also briefly discuss the generalization for FRW models with arbitrary spatial curvature ($K = -1, 0, +1$). In section 4.9, we discuss the generalization to include situations where $\epsilon$ is time-dependent. This last issue is an intruiging puzzle that is not yet fully solved, although we describe several partial solutions. These partial solutions are again surprisingly simple, leading us to hope that they may be clues along the path to a more general formulation of the duality that retains the elegance of the $\epsilon \to 1/\epsilon$ prescription.

## 4.2 Background model

A spatially-flat Friedmann-Robertson-Walker (FRW) universe with scalar field $\phi$ and potential $V(\phi)$ is described by the metric

$$\mathrm{d}s^2 = a(\tau)^2 \left[-\mathrm{d}\tau^2 + \mathrm{d}\vec{x}^2\right]. \tag{4.1}$$

The unperturbed scalar field $\phi_0(\tau)$ and scale factor $a(\tau)$ obey the Friedmann equations

$$6(a'^2/a^4) = 2\rho \tag{4.2a}$$

$$6(a''/a^3) = \rho - 3p \tag{4.2b}$$

where we have chosen units such that $c = \hbar = 8\pi G = 1$, a prime ($'$) denotes a conformal time derivative $\mathrm{d}/\mathrm{d}\tau$, and the energy density and pressure are given by

$$\rho = (1/2)a^{-2}\phi_0'^2 + V(\phi_0) \tag{4.3a}$$

$$p = (1/2)a^{-2}\phi_0'^2 - V(\phi_0). \tag{4.3b}$$

Instead of the usual variable $w \equiv p/\rho$, it will be more convenient to use

$$\epsilon \equiv 3(1+w)/2. \tag{4.4}$$

to parameterize the equation of state. Equations (4.2,4.3) imply $-1 \le w < \infty$ (or equivalently $0 \le \epsilon < \infty$). If $w$ is near $-1$, then $\epsilon \ll 1$ is the usual slow-roll parameter; but we make *no slow-roll approximation* in this paper, and $\epsilon$ may be arbitrarily large.

From now on, we shall assume that $\epsilon$ is constant, not equal to unity. (The self-dual case $\epsilon = 1$ possesses special behavior which we shall not study here.) Then the solution of equations (4.2,4.3) is:

$$a(\tau) = |\tau|^{1/(\epsilon-1)} \tag{4.5a}$$

$$\phi_0(\tau) = \pm\frac{(2\epsilon)^{1/2}}{\epsilon-1}\ln|\tau| \tag{4.5b}$$

$$V(\phi) = \frac{3-\epsilon}{(\epsilon-1)^2}\exp\Big[\mp(2\epsilon)^{1/2}\phi\Big] \tag{4.5c}$$

where, to fix integration constants we have, without loss of generality, chosen the origin of conformal time so that $a(0) = 0$, normalized the scale factor so that $a(1) = 1$, and redefined $\phi_0 \to \phi_0$+constant so that $\phi_0(1) = 0$.

This solution separates into 4 cases:

- (a) expanding, $0 \leq \epsilon < 1$, $-\infty < \tau < 0$;
- (b) expanding, $1 < \epsilon < \infty$, $0 < \tau < \infty$;
- (c) contracting, $0 \leq \epsilon < 1$, $0 < \tau < \infty$;
- (d) contracting, $1 < \epsilon < \infty$, $-\infty < \tau < 0$.

The Penrose diagrams for these 4 possibilities are shown in Figure 4.1. Case (b) corresponds to an ordinary expanding FRW model with matter or radiation domination, $\epsilon = 3/2$ or 2, respectively. We are interested in the two cases, (a) and (d), in which $\tau$ runs from $-\infty \to 0$ since, in these two cases, comoving length scales start *inside* the Hubble horizon at early times, and end *outside* the horizon at late times. Since we wish to study the amplification of perturbations as modes leave the horizon, these are the two relevant cases. Thus, we will always assume $\epsilon < 1 \Leftrightarrow$ expanding and $\epsilon > 1 \Leftrightarrow$ contracting. The duality discussed below pairs solutions of type (a) with solutions of type (d). Figure (4.1) emphasizes that (a) and (d) have similar causal structure, but different singularities.

## 4.3 Scalar perturbations

In this section, we introduce relevant aspects of gauge-invariant scalar perturbation theory in 4 dimensions, and catch our first glimpse of the duality discussed in section 4.4. For a thorough introduction to gauge-invariant perturbations [9], see [99] and [123]. We work in Fourier space throughout, so every perturbation variable carries an implicit subscript $\vec{k}$ which, for brevity, is not shown explicitly. Write the perturbed metric

$$\begin{aligned} \mathrm{d}s^2/a^2 &= -(1+2AY)\mathrm{d}\tau^2 - 2BY_i\mathrm{d}\tau\mathrm{d}x^i \\ &\quad + \left[(1+2H_LY)\delta_{ij} + 2H_TY_{ij}\right]\mathrm{d}x^i\mathrm{d}x^j \end{aligned} \tag{4.6a}$$

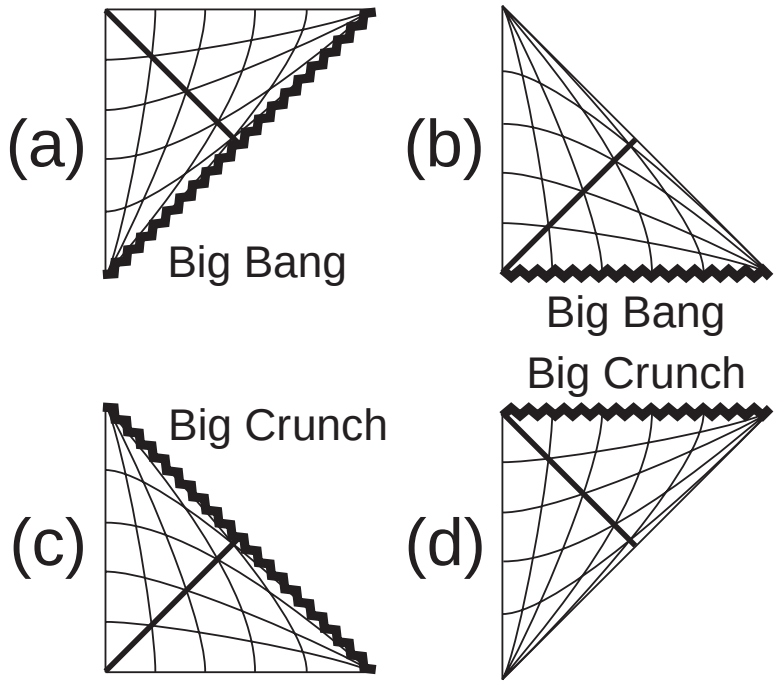



Figure 4.1: Penrose diagrams for spatially-flat FRW universes with: (a) $0 < \epsilon < 1$, expanding; (b) $1 < \epsilon < \infty$, expanding; (c) $0 < \epsilon < 1$, contracting; (d) $1 < \epsilon < \infty$, contracting. The left edge of each diagram is the world line of a comoving observer at the origin; curved lines represent other comoving world lines and spatial hypersurfaces. The Hubble horizon is a curve connecting the $90^o$ vertex to the lightlike boundary, but the precise curve depends on $\epsilon$. For illustration, we have shown the horizon for $\epsilon = 0$ in (a, c) and for $\epsilon = 2$ in (b, d). In this paper, we focus on cases (a) and (d), in which comoving scales exit the Hubble horizon.

and perturbed scalar field

$$\phi = \phi_0(\tau) + \delta\phi(\tau)Y \tag{4.6b}$$

where $Y(\vec{x})$, $Y_i(\vec{x})$, and $Y_{ij}(\vec{x})$ are scalar harmonics (see Appendix C in [99]). The corresponding perturbations of the Einstein tensor and energy-momentum tensor (see Appendices D and F in [99]) are related to one another through the perturbed Einstein equations, $\delta G^{\mu}{}_{\nu} = \delta T^{\mu}{}_{\nu}$.

It is well known that scalar perturbations in a spatially-flat FRW universe with scalar field $\phi$ and potential $V(\phi)$ are completely characterized by a single gauge-invariant variable. But the choice of this variable is neither unique nor standard; two of the most familiar options are the "Newtonian potential," $\Phi$, and the "curvature perturbation," $\zeta$.

The gauge-invariant Newtonian potential $\Phi$ is most easily understood in "Newtonian gauge" ($B = H_T = 0$), where it is related to the metric perturbations in a simple way: $\Phi = A = -H_L$. It obeys the equation of motion

$$\Phi'' + 2\left[\frac{a'}{a} - \frac{\phi_0''}{\phi_0'}\right]\Phi' + \left[k^2 + 2\mathcal{H}' - 2\mathcal{H}\frac{\phi_0''}{\phi_0'}\right]\Phi = 0 \tag{4.7}$$

where $k = |\vec{k}|$ is the magnitude of the (comoving) Fourier 3-vector. On the other hand, the gauge-invariant perturbation variable $\zeta$ is most easily understood in "comoving gauge" ($H_T = \delta T^0{}_i = 0$), where it represents the curvature perturbation on spatial-hypersurfaces, and is related to the spatial metric perturbation in a simple way: $\zeta = -H_L$. The condition $\delta T^0{}_i = 0$ also implies that $\delta\phi = 0$ in this gauge. $\zeta$ obeys the equation of motion

$$\zeta'' + 2(z'/z)\zeta' + k^2\zeta = 0 \tag{4.8}$$

where $z \equiv a^2\phi_0'/a'$. $\Phi$ and $\zeta$ are related to each other by

$$\zeta = \Phi + \frac{1}{\epsilon}\left[(a/a')\Phi' + \Phi\right] \tag{4.9a}$$

$$\Phi = -\epsilon(a'/a)k^{-2}\zeta' \,. \tag{4.9b}$$

Note that our definitions for $\Phi$ and $\zeta$ agree with those in [123]. But beware: in [99], the gauge-invariant Newtonian potential is denoted $\Psi$, while $\Phi$ denotes a different (though closely related) variable.

It is convenient to introduce new variables, $u$ and $v$ [120, 121, 123], by multiplying $\Phi$ and $\zeta$ by $k$-independent functions of $\tau$:

$$u \equiv (a/\phi_0')\Phi \qquad v \equiv z\zeta. \tag{4.10}$$

Note that $u$ and $v$ have the same $k$-dependence as $\Phi$ and $\zeta$, respectively, and may serve as surrogates for $\Phi$ and $\zeta$. If we also define background quantities, $\theta$ and $z$:

$$\theta \equiv 1/z \equiv \frac{a'}{a^2\phi_0'}, \tag{4.11}$$

then $u$ and $v$ obey simple equations of motion

$$u'' + (k^2 - \theta''/\theta)u = 0 \tag{4.12a}$$

$$v'' + (k^2 - z''/z)v = 0 \tag{4.12b}$$

and are related to each other by

$$kv = 2k[u' + (z'/z)u] \tag{4.13a}$$

$$-ku = \frac{1}{2k}[v' + (\theta'/\theta)v]. \tag{4.13b}$$

We must choose a vacuum state for the fluctuations, which corresponds to specifying appropriate boundary conditions for $u$ and $v$ (see Ch.3 in Birrel&Davies [17]). The standard choice is the Minkowski vacuum of a comoving observer in the far past (when all comoving scales were far inside the Hubble horizon), corresponding to the boundary conditions

$$u \to i(2k)^{-3/2}\mathrm{e}^{-ik\tau} \tag{4.14a}$$

$$v \to (2k)^{-1/2}\mathrm{e}^{-ik\tau} \tag{4.14b}$$

as $\tau \to -\infty$. Using (4.13), it is easy to check that these two boundary conditions are equivalent.

When $\epsilon$ is time-independent, we can use (4.5) to find

$$\theta''/\theta = \frac{\epsilon}{(\epsilon-1)^2\tau^2} \tag{4.15a}$$

$$z''/z = \frac{2-\epsilon}{(\epsilon-1)^2\tau^2} \tag{4.15b}$$

Then we may solve (4.12) to obtain

$$u(x) = x^{1/2}\left[A^{(1)}H_\alpha^{(1)}(x) + A^{(2)}H_\alpha^{(2)}(x)\right] \tag{4.16a}$$

$$v(x) = x^{1/2}\left[B^{(1)}H_\beta^{(1)}(x) + B^{(2)}H_\beta^{(2)}(x)\right] \tag{4.16b}$$

where $x \equiv k|\tau|$ is a dimensionless time variable, $A^{(1,2)}$ and $B^{(1,2)}$ are constants, $H_\rho^{(1,2)}(x)$ are Hankel functions, and we have defined

$$\alpha \equiv \sqrt{(\theta''/\theta)\tau^2 + 1/4} = \frac{1}{2}\left|\frac{\epsilon+1}{\epsilon-1}\right| \tag{4.17a}$$

$$\beta \equiv \sqrt{(z''/z)\tau^2 + 1/4} = \frac{1}{2}\left|\frac{\epsilon-3}{\epsilon-1}\right| \tag{4.17b}$$

In the far past ($x \to \infty$) we use the asymptotic Hankel expression,

$$H_s^{(1,2)}(x) \to \sqrt{\frac{2}{\pi x}}\exp\left[\pm\, i\left(x - \frac{s\,\pi}{2} - \frac{\pi}{4}\right)\right] \tag{4.18}$$

so the boundary conditions (4.14) imply

$$u = \frac{\mathcal{P}_1}{2k}(\pi x/4k)^{1/2}H_\alpha^{(1)}(x) \tag{4.19a}$$

$$v = \mathcal{P}_2\,(\pi x/4k)^{1/2}H_\beta^{(1)}(x) \tag{4.19b}$$

where

$$\mathcal{P}_1 = \exp[i(2\alpha+3)\pi/4] \tag{4.20a}$$

$$\mathcal{P}_2 = \exp[i(2\beta+1)\pi/4] \tag{4.20b}$$

are $k$-independent complex phase factors.

Note from (4.15a) that the equation of motion for $u$, (4.12a), is invariant under $\epsilon \to 1/\epsilon$, while the boundary condition, (4.14a), is independent of $\epsilon$. As a result, our expressions (4.72a) for $\alpha$ and (4.19a) for $u$ are invariant under $\epsilon \to 1/\epsilon$. This is our first glimpse of the duality discussed below.

We stress that this result does not depend on the particular vacuum choice (4.14a). Any boundary condition that is independent of $\epsilon$ (or, more generally, invariant under $\epsilon \to 1/\epsilon$) will work. And it is natural to expect the boundary condition to be independent of $\epsilon$, since it is imposed in the far past, when comoving scales are far inside the Hubble horizon.

## 4.4 Dominant and subdominant modes

In this section, we show that $u$ and $v$ can be decomposed into pieces that are dominant and subdominant at long wavelengths such that each is invariant under the transformation $\epsilon \to 1/\epsilon$. The dominant and subdominant parts are closely related to growing and decaying modes over the range of $w$ relevant to cosmological model-building, as we explain below.

For this analysis, it is convenient to scale $u$ and $v$ by appropriate powers of $|\tau|$, so that they only depend on $k$ and $\tau$ through the dimensionless combination $x = k|\tau|$. Thus, using (4.19a, 4.19b), define

$$\bar{u} \equiv |\tau|^{-3/2} u = \frac{\mathcal{P}_1}{2x} (\pi/4)^{1/2} H_\alpha^{(1)}(x) \tag{4.21a}$$

$$\bar{v} \equiv |\tau|^{-1/2} v = \mathcal{P}_2 \, (\pi/4)^{1/2} H_\beta^{(1)}(x). \tag{4.21b}$$

Note that $\bar{u}$ and $\bar{v}$ have the same $k$-dependence as $u$ and $v$, respectively, or $\Phi$ and $\zeta$, respectively, and may be used in place of these more standard variables.

To make the meaning of "dominant" and "subdominant" precise, consider two linearly independent functions $f_1(x)$ and $f_2(x)$. If $\lim_{x\to 0} f_2(x)/f_1(x)$ exists, then $f_1(x)$ and $f_2(x)$ can be related by a linear transformation to two new functions, $f_{\rm dom}(x)$ and $f_{\rm sub}(x)$, satisfying

$$\lim_{x\to 0} f_{\rm sub}(x)/f_{\rm dom}(x) = 0. \tag{4.22}$$

So the subdominant piece, $f_{\rm sub}(x)$, becomes negligible relative to the dominant piece, $f_{\rm dom}(x)$, for small $x$ (*i.e.* far outside the horizon). The condition (4.22) uniquely determines the subdominant piece (up to an overall normalization constant) to be

$$f_{\rm sub}(x) = f_2(x) - \left[\lim_{y\to 0} \frac{f_2(y)}{f_1(y)}\right] f_1(x), \tag{4.23}$$

but does not uniquely fix the dominant piece. Rather, $f_{\rm dom}(x)$ may be any linear combination of $f_1(x)$ and $f_2(x)$ that is linearly independent of $f_{\rm sub}(x)$. We can now choose $f_1(x) = \bar{u}(x)$ and $f_2(x) = \bar{v}(x)$, and find the corresponding dominant and subdominant functions.

Let us first calculate $f_{\rm sub}(x)$. To compute $\lim_{x\to 0} f_2(x)/f_1(x)$, we use the Hankel identity

$$H_s^{(1)}(x) \to -i[\Gamma(s)/\pi](x/2)^{-s} \quad \text{as} \quad x \to 0, \tag{4.24}$$

where $s > 0$ and $\Gamma(s)$ is the Euler gamma function. Then from (4.21a,4.21b) we find

$$\lim_{x\to 0} \frac{f_2(x)}{f_1(x)} = \begin{cases} (\mathcal{P}_2/\mathcal{P}_1)4\alpha & 0 \le \epsilon < 1 \\ 0 & 1 < \epsilon < \infty \end{cases} \tag{4.25}$$

where we have used the fact that $\alpha + 1 = \beta$ when $\epsilon < 1$, while $\alpha + 1 > \beta$ when $\epsilon > 1$. Now, substituting (4.21a), (4.21b) and (4.25) into (4.23), using the Hankel identities

$$H_{s-1}^{(1)}(x) + H_{s+1}^{(1)}(x) = \tfrac{2s}{x} H_s^{(1)}(x) \tag{4.26a}$$

$$H_{-s}^{(1)}(x) = \mathrm{e}^{i\pi s} H_s^{(1)}(x), \tag{4.26b}$$

and paying careful attention to absolute value signs, we find

$$f_{\rm sub}(x) = \mathcal{P}_3(\pi/4)^{1/2} H_\gamma^{(1)}(x), \tag{4.27}$$

where we have defined

$$\gamma \equiv |\alpha - 1|, \tag{4.28}$$

and $\mathcal{P}_3 = \exp[i(2\gamma + 1)\pi/4]$ is a $k$-independent complex phase factor.

Now let us turn to $f_{\rm dom}(x)$. The fact that $\lim_{x\to 0} f_{\rm sub}(x)/\bar{v}(x) = 0$ when $\epsilon < 1$ shows that the dominant mode contributes to $\bar{v}$ in an expanding universe. Since $f_{\rm sub}(x) = \bar{v}(x)$ when $\epsilon > 1$, $\bar{v}$ is purely subdominant (*i.e.* contains no dominant-mode contribution) in a contracting universe. By contrast, $\bar{u}$ and $f_s(x)$ are always linearly independent, and hence $\lim_{x\to 0} f_{\rm sub}(x)/\bar{u}(x) = 0$ for all $\epsilon$. Thus, we can use the freedom in defining $f_{\rm dom}$ to choose

$$f_{\rm dom}(x) = \bar{u}(x) = \frac{\mathcal{P}_1}{2x}(\pi/4)^{1/2} H_\alpha^{(1)}(x) \tag{4.29}$$

for the dominant piece.

Notice that the expressions (4.27) for $f_{\rm sub}(x)$ and (4.29) for $f_{\rm dom}(x)$ are *both* invariant under $\epsilon \to 1/\epsilon$, because $\alpha$ is invariant. Thus, we see that the subdominant mode *automatically* possesses this symmetry, since $f_{\rm sub}(x)$ is uniquely determined (up to a normalization

factor) by the condition (4.22). Furthermore, we have shown that we can choose a linear combination of $\bar{u}(x)$ and $\bar{v}(x)$ such that $f_{\rm dom}(x)$ displays the same symmetry. (If we had made the wrong choice for $f_{\rm dom}(x)$, then the exact $\epsilon \to 1/\epsilon$ symmetry of the dominant mode would be hidden, and would only re-appear in the long-wavelength limit.)

In the $x \to 0$ limit, define dominant and subdominant scalar spectral indices, $n_{\rm dom}$ and $n_{\rm sub}$, which satisfy

$$x^3 |f_{\rm dom}|^2 \propto x^{n_{\rm dom}-1} \tag{4.30a}$$

$$x^3 |f_{\rm sub}|^2 \propto x^{n_{\rm sub}-1} \tag{4.30b}$$

Using (4.27) and (4.29), along with the Hankel identity (4.24), we find

$$n_{\rm dom} - 1 = 1 - 2\alpha = 1 - \left|\frac{\epsilon+1}{\epsilon-1}\right| \tag{4.31a}$$

$$n_{\rm sub} - 1 = 3 - 2\gamma = 3 - \left|\left|\frac{\epsilon+1}{\epsilon-1}\right| - 2\right| \tag{4.31b}$$

These spectral indices are plotted in Fig. 4.2a (as a function of $w$), and in Fig. 4.2b (as a function of $\epsilon$). Again, notice that $n_{\rm dom}$ and $n_{\rm sub}$ are *both* invariant under $\epsilon \to 1/\epsilon$. This symmetry is manifest in Fig. 4.2b.

Since $\epsilon$ lies in the range $0 \leq \epsilon < \infty$, this duality formally pairs every expanding ($\epsilon < 1$) universe with a contracting ($\widehat{\epsilon} > 1$) universe, and *vice versa*. However, the background solution (4.5) is only stable against small perturbations in two cases: (*i*) expanding with $\epsilon < 1$ or (*ii*) contracting with $\epsilon > 3$ [69, 59]. Thus, an expanding model and its contracting dual are *both* stable when $\epsilon < 1/3$ ($w < -7/9$) and $\widehat{\epsilon} > 3$ ($\widehat{w} > 1$). Also note, in agreement with [69], that there are only two limits in which an approximately scale-invariant ($n_{\rm grow} - 1 \approx 0$) spectrum of scalar perturbations is produced: (*i*) when $\epsilon \to 0$ ($w \to -1$), corresponding to the inflationary regime and (*ii*) when $\epsilon \to \infty$ ($w \to \infty$) corresponding to the cyclic/ekpyrotic regime.

The dominant and subdominant pieces, $f_{\rm dom}(x)$ and $f_{\rm sub}(x)$ are related to the growing and decaying modes of $\bar{u}$ and $\bar{v}$, which may simply be obtained by replacing the Hankel

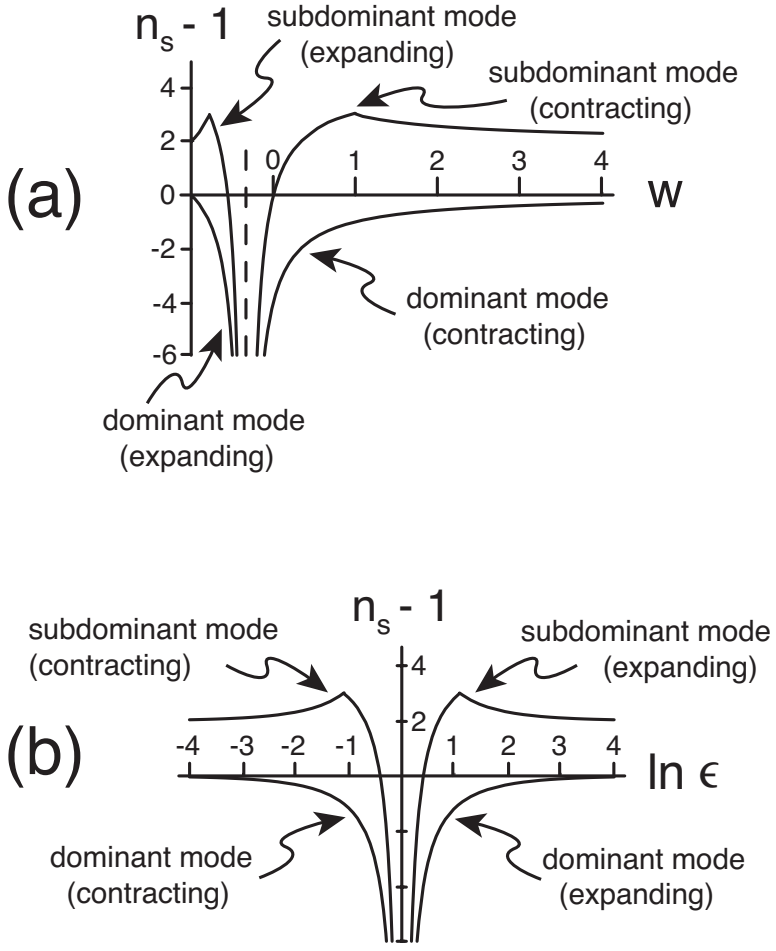



Figure 4.2: The dominant and subdominant scalar spectral indices (a) as a function of $w$ and (b) as a function of $\ln\epsilon$. Note especially the symmetry of (b).

functions in (4.21a,4.21b) with the corresponding Neumann functions or Bessel functions:

$$\bar{u}_{\rm grow} \propto x^{-1} Y_\alpha(x) \qquad \bar{u}_{\rm decay} \propto x^{-1} J_\alpha(x) \tag{4.32a}$$

$$\bar{v}_{\rm grow} \propto Y_\beta(x) \qquad \bar{v}_{\rm decay} \propto J_\beta(x). \tag{4.32b}$$

We may define growing-mode and decaying-mode spectral indices for $\bar{u}$ and $\bar{v}$, just as we did for $f_{\rm sub}$ and $f_{\rm dom}$ in (4.30). Now restrict attention to the ranges $\epsilon < 1/3$ and $\widehat{\epsilon} > 3$ (*i.e.* the range over which the duality pairs stable expanding models to stable contracting models). Then, using (4.32) along with the (4.24), we find

- The growing-mode and decaying-mode spectral indices associated with $\bar{u}$ are simply equal to $n_{\rm dom}$ and $n_{\rm sub}$, respectively.
- The growing-mode spectral index associated with $\bar{v}$ is equal to $n_{\rm dom}$ in an expanding ($\epsilon < 1/3$) universe, but is equal to $n_{\rm sub}$ in a contracting ($\widehat{\epsilon} > 3$) universe.

## 4.5 Tensor perturbations

Tensor perturbations are much simpler than scalar perturbations. The perturbed metric is

$$\mathrm{d}s^2/a^2 = -\mathrm{d}\tau^2 + [\delta_{ij} + 2h_T Y^{(2)}_{ij}]\mathrm{d}x^i \mathrm{d}x^j \tag{4.33}$$

where $Y^{(2)}_{ij}$ is a tensor harmonic (see Appendix C in [99]). The tensor perturbation $h_T$ is gauge-invariant and obeys

$$h_T'' + 2(a'/a)h_T' + k^2 h_T = 0. \tag{4.34}$$

It is useful to define a new variable $f_T \equiv a h_T$ which obeys

$$f_T'' + (k^2 - a''/a) f_T = 0. \tag{4.35}$$

Again, the standard vacuum choice is the Minkowski vacuum of a comoving observer in the far past, corresponding to the boundary condition

$$f_T \to (2k)^{-1/2} e^{-ik\tau} \qquad \text{as} \quad \tau \to -\infty. \tag{4.36}$$

We can now solve for $f_T$, just as in the scalar case. But it is simpler to notice that (4.5) and (4.11) imply $z(\tau) \propto a(\tau)$, and hence $z''/z = a''/a$ when $\epsilon$ is constant. Thus, since $v$ and $f_T$ obey identical equations of motion (compare (4.12b) with (5.9)) and boundary conditions (compare (4.14b) with (4.36)), we find

$$f_T = v = \mathcal{P}_2(\pi x/4k)^{1/2} H_\beta^{(1)}(x). \tag{4.37}$$

The tensor spectral index is defined in the long-wavelength limit by $k^3|f_T|^2 \propto k^{n_T}$. Using (4.24) and (4.37) we find

$$n_T = 3 - 2\beta = 3 - \left|\frac{\epsilon - 3}{\epsilon - 1}\right|. \tag{4.38}$$

Note, in particular, that this expression is *not* invariant under $\epsilon \to 1/\epsilon$. An expanding universe with equation of state $\epsilon$ produces a tensor spectrum which is much redder than the tensor spectrum produced in a contracting universe with $\widehat{\epsilon} = 1/\epsilon$: $n_T \leq \widehat{n}_T - 2$.

## 4.6 Other dualities

It is interesting to contrast our duality with other cosmological dualities that have been discussed in the literature.

One duality, due to Wands [175] (see also [155]), pairs models that share the same "$v$" perturbations. By contrast, our duality pairs models that share the same "$u$" perturbations. (Note: the variable called "$u$" in [175] is called "$v$" in our paper, in agreement with Mukhanov's convention [120, 121, 123].) Thus, whereas our duality connects expanding and contracting models through the substitution $\epsilon \to \widehat{\epsilon} = 1/\epsilon$ (which leaves $\theta''/\theta$, and hence $u$, invariant), Wands' duality instead uses the substitution $\epsilon \to \widehat{\epsilon} = (2\epsilon - 3)/(\epsilon - 2)$ (which leaves $z''/z$, and hence $v$, invariant). For example, his duality pairs an expanding inflationary solution ($\epsilon = 0$, $w = -1$) with a contracting dustlike solution ($\widehat{\epsilon} = 3/2$, $\widehat{w} = 0$). Another difference between our duality and Wands' stems from the fact that $v$ is purely subdominant (*i.e.* contains no dominant-mode contribution) in a contracting universe (see section 4.4). Thus, whereas our duality maps the expanding-phase *dominant*

mode to the contracting-phase *dominant* mode, and the expanding-phase subdominant mode to the contracting-phase subdominant mode, Wands' duality instead associates the expanding-phase *dominant* mode with the contracting-phase *subdominant* mode.

A second interesting duality, discussed by Brustein *et al.* [28], applies to a broad class of cosmological perturbations. Associated with each type of perturbation is a "pump"—a particular function of the background fields. The Hamiltonian governing a given perturbation is invariant under a transformation that swaps the perturbation with its conjugate momentum, and simultaneously inverts the appropriate pump function [28]. It is instructive to apply this duality to the $u$ and $v$ variables considered in this paper. In this case, the Brustein *et al.* duality associates an expanding solution characterized by $\theta$ (or $\epsilon$) to a contracting universe characterized by $\widehat{\theta} = 1/\theta = z$ (or $\widehat{\epsilon} = 2 - \epsilon$). The transformation $\epsilon \to \widehat{\epsilon} = 2 - \epsilon$ effectively swaps the variables $u$ and $v$

$$u \to \widehat{u} = (i/2k)v \tag{4.39a}$$

$$v \to \widehat{v} = (2k/i)u \tag{4.39b}$$

as may be verified from (4.19a,4.19b).

Recall that $\epsilon$ lies in the range $0 \leq \epsilon < \infty$. Thus, our duality formally pairs every expanding ($0 \leq \epsilon < 1$) solution with a contracting ($1 < \widehat{\epsilon} < \infty$) solution, and *vice versa*. By contrast, Wands' duality relates every expanding solution to a contracting solution with $1 < \widehat{\epsilon} \leq 3/2$; but contracting solutions with $\widehat{\epsilon} > 3/2$ have no expanding dual. Similarly, Brustein *et al.*'s duality pairs every expanding solution with a contracting solution in the range $1 < \widehat{\epsilon} \leq 2$; but contracting solutions with $\widehat{\epsilon} > 2$ have no expanding dual.

The constant-$\epsilon$ background solutions (4.5) are only practically useful if they are dynamically stable. Recall that the contracting solutions are stable if and only if $\epsilon > 3$ ($w > 1$) [69, 59]. Thus, Wands' and Brustein *et al.*'s dualities relate every expanding solution to an *unstable* contracting solution. By contrast, our duality relates *stable* expanding solutions with $0 \leq \epsilon < 1/3$ ($-1 \leq w < -7/9$) to *stable* contracting solutions with $3 < \widehat{\epsilon} < \infty$ ($1 < \widehat{w} < \infty$). In terms of the spectral index, this means that $n_s > 0$ may be produced *either* by a stable expanding phase *or* by a stable contracting phase (provided that cos-

| Transformation of background | Transformation of perturbations | Range of $\epsilon$ | Maps stable to stable? | Ref. |
|---|---|---|---|---|
| $\epsilon \to 1/\epsilon$ | $u \to u$ | $[0, \infty)$ | Yes, if $n_s > 0$ | - |
| $\epsilon \to \frac{2\epsilon-3}{\epsilon-2}$ | $v \to v$ | $[0, 3/2]$ | never | [175, 155] |
| $\epsilon \to 2 - \epsilon$ | $u \leftrightarrow (i/2k)v$ | $[0, 2]$ | never | [28] |

Table 4.1: Comparison of the duality presented here with those presented by Wands [175, 155] and by Brustein *et al.* [28]. The first two columns show how the background and perturbation variables transform under each duality. The third column shows the range of $\epsilon$ to which the duality applies. The fourth column indicates the condition under which an expanding background solution and its contracting dual are both stable under small perturbations.

mological models exist in which the Newtonian potential perturbations produced during the contracting phase successfully propagate into the expanding phase, as discussed in the introduction). In the real universe, the condition $n_s > 0$ is satisfied (experiments favor $n_s \approx 1$), so that our duality is of practical relevance in cosmological model building. Some properties of the three different duality relations are summarized in Table I.

Finally, a number of authors have discussed cosmological symmetries of the low-energy string effective action. If one neglects all fields in this action besides the dilaton and the metric, then there is a well–known "scale–factor duality" [172, 167, 144]: starting with any cosmological solution, one can use this duality to generate new solutions. If, in addition to the dilaton, one includes other fields (axions, moduli,...), then the cosmological solutions may display more general dualities, and the resulting perturbation spectra may be invariant under these dualities [37, 38, 106, 27, 30, 31]. But note that these symmetries typically relate different solutions of a single effective action. By contrast, the dualities in Table I relate two different cosmological background solutions corresponding to *two different Lagrangians*: an expanding universe, with potential $V(\phi)$, is dual to a contracting universe, with a *different* potential $\widehat{V}(\phi)$.

## 4.7 Geometric interpretation

Using the background solutions (4.5), we may think of the duality as relating two different scale factors, $a(\tau)$ and $\widehat{a}(\tau)$, or two different scalar potentials, $V(\phi)$ and $\widehat{V}(\phi)$. Alternatively, recall that $a \propto |t|^{1/\epsilon}$ and $H^{-1} \equiv a^2/a' \propto t$, where $t$ is the proper time of a comoving observer and $H^{-1}$ is the "Hubble scale." Thus, two dual universes are related by

$$\mathrm{d}\ln H/\mathrm{d}\ln a = \mathrm{d}\ln\widehat{a}/\mathrm{d}\ln\widehat{H}. \tag{4.40}$$

So the duality effectively swaps the scale factor with the Hubble scale, and simultaneously swaps expansion and contraction. For example, in the $\epsilon \to 0$ ($w \to -1$) limit, the scale factor grows rapidly while the Hubble length grows slowly; whereas in the $\epsilon \to \infty$ ($w \to \infty$) limit, the Hubble length shrinks rapidly while the scale factor shrinks slowly. Expanding models in which modes exit the horizon most rapidly ($w \to -1$) and most slowly ($w \to -1/3$), are associated with contracting models in which modes exit most rapidly ($w \to \infty$) and most slowly ($w \to -1/3$), respectively.

## 4.8 Generalization to arbitrary spacetime dimension

Thus far, we have discussed the $\epsilon \to 1/\epsilon$ duality in the restricted situation that seems likely to be of greatest physical interest: a background FRW universe with $3+1$ dimensions, vanishing spatial curvature, and constant or slowly-varying $\epsilon$. It is also interesting to see how it generalizes when each of these restrictions is removed.

In this section, we generalize the discussion from 4 spacetime dimensions to $d$ spacetime dimensions, with $d \geq 4$, while continuing to assume that $\epsilon$ is constant or slowly varying. At the end of the section, we briefly discuss the generalization to FRW models with arbitrary spatial curvature: open ($K = -1$), closed ($K = +1$), or flat ($K = 0$).

Gauge-invariant cosmological perturbation theory for FRW models with $d$ spacetime dimensions and arbitrary spatial curvature $K$ is treated thoroughly in [99] (see especially the appendices therein).

The background metric (4.1) obeys the Friedmann equations

$$(d-1)(d-2)a'^2/a^4 = 2\rho \tag{4.41a}$$

$$(d-1)(d-2)\,a''/a^3 = (5-d)\rho + (1-d)p \tag{4.41b}$$

where $\rho$ and $p$ are given by (4.3). Instead of $w \equiv p/\rho$, parameterize the equation of state with

$$\epsilon \equiv \frac{(2d-5)+(d-1)w}{d-2}. \tag{4.42}$$

From eqs. (4.3, 4.41) we find $-1 \le w < \infty$ or $\frac{d-4}{d-2} \le \epsilon < \infty$. For constant $\epsilon$, the solution of (4.3, 4.41) is

$$a(\tau) = |\tau|^{2/[(d-2)(\epsilon-1)]} \tag{4.43a}$$

$$\phi_0(\tau) = \pm\frac{1}{\epsilon-1}\sqrt{2\left(\epsilon - \frac{d-4}{d-2}\right)}\,\ln|\tau| \tag{4.43b}$$

$$V(\phi) = \frac{3-\epsilon}{(\epsilon-1)^2}\exp\left[\mp\sqrt{2\left(\epsilon-\frac{d-4}{d-2}\right)}\,\phi\right] \tag{4.43c}$$

where we have chosen $a(0) = 0$, $a(1) = 1$ and $\phi_0(1) = 0$.

As shown in [99], perturbations in $d$ spacetime dimensions may be decomposed into scalars, vectors and tensors, just as in 4 dimensions, and gauge-invariant variables may be defined. In particular, we can again introduce scalar perturbations through equations (4.6a, 4.6b), and describe these perturbations with a single gauge-invariant variable. The gauge-invariant Newtonian potential, $\Phi$, is most easily understood in "Newtonian gauge" ($B = H_L = 0$), where it is related to the metric perturbations in a simple way: $\Phi = A = -(d-3)H_L$. The gauge-invariant variable $\zeta$ is most easily understood in "comoving gauge" ($H_T = \delta T^0_i = 0$), where it is related to the spatial metric perturbation in a simple way: $\zeta = -H_L$. Note also that $\delta\phi = 0$ in this gauge. If we take the $d$-dimensional definitions of $u$, $v$, $\theta$ and $z$ to be:

$$u \equiv \frac{1}{2}\left(\frac{d-2}{d-3}\right)\frac{a^{(d-2)/2}}{\phi_0'}\Phi \qquad\qquad v \equiv z\zeta \tag{4.44}$$

$$\theta \equiv 1/z \equiv \frac{a'}{a^{d/2}\phi_0'}, \tag{4.45}$$

then sections 4.3 and 4.4 immediately generalize to $d$ dimensions. Indeed, it is straightforward to show that equations (4.12) through (4.29) all remain true in $d$ dimensions, provided we take $\epsilon$, $u$, $v$, $\theta$, and $z$ to be given by their $d$-dimensional definitions (4.42), (4.44) and (4.45).

In particular, this means that the duality extends to $d$-dimensions: $u$, $f_{\rm grow}$ and $f_{\rm decay}$ are all invariant under $\epsilon \to 1/\epsilon$. The duality pairs expanding solutions in the range $\frac{d-4}{d-2} \leq \epsilon < 1$ with contracting solutions in the range $1 < \epsilon < \infty$. This means that in $d = 4$, $\epsilon$ lies in the range $0 \leq \epsilon < \infty$, so that every expanding model has a contracting dual, and *vice versa*; whereas in $d > 4$, the total range of $\epsilon$ is somewhat smaller, so that every expanding model has a contracting dual, but contracting models with $\epsilon > \frac{d-2}{d-4}$ have no expanding dual.

In the $x \to 0$ limit in $d$ dimensions, define the growing and decaying spectral indices

$$x^{d-1}|f_{\rm grow}|^2 \propto x^{n_{\rm grow}-1} \tag{4.46a}$$

$$x^{d-1}|f_{\rm decay}|^2 \propto x^{n_{\rm decay}-1}. \tag{4.46b}$$

Then, using (4.72a), (4.24), (4.27), (4.28) and (4.29) we find

$$n_{\rm grow} - 1 = d - 3 - \left|\frac{\epsilon+1}{\epsilon-1}\right| \tag{4.47a}$$

$$n_{\rm decay} - 1 = d - 1 - \left|\left|\frac{\epsilon+1}{\epsilon-1}\right| - 2\right| \tag{4.47b}$$

with $\epsilon$ given by (4.42). Note that these expressions are invariant under $\epsilon \to 1/\epsilon$.

As shown in [99], we may again introduce tensor perturbations through eq. (4.33). Now define the gauge-invariant variable $f_T \equiv a^{(d-2)/2} h_T$ which obeys

$$f_T'' + [k^2 - (a^{(d-2)/2})''/a^{(d-2)/2}] f_T = 0 \tag{4.48}$$

along with the boundary condition (4.36). Using (4.36), (4.43) and (4.48) we find that solution for $f_T$ is still given by (4.37) in $d$ dimensions. In the $k \to 0$ limit, define the tensor spectral index $k^{d-1}|f|^2 \propto k^{n_T}$. Then, using (4.24) and (4.37) we find

$$n_T = d - 1 - \left|\frac{\epsilon-3}{\epsilon-1}\right| \tag{4.49}$$

with $\epsilon$ given by (4.42). Note that this expression is *not* invariant under $\epsilon \to 1/\epsilon$. Thus, the situation in $d$ spacetime dimensions is the same as it was in 4 spacetime dimensions: the scalar perturbations are invariant under $\epsilon \to 1/\epsilon$ (in the sense we have explained), but the tensor perturbations are not. This completes the generalization to $d$ spacetime dimensions.

Before ending this section, let us very briefly discuss how the duality generalizes in FRW models with arbitrary spatial curvature ($K = -1, 0, +1$). Although this generalization may be purely academic, we hope that — along with the rest of this section — it may provide some clues for solving the more physically-relevant puzzle of how to recast the duality in a more general form that makes its physical origin more transparent and accomodates time-varying $\epsilon$. This subject is the focus of the next section.

To understand the generalization to arbitrary $K$, first note that when $K = 0$ the duality is actually slightly more general than we have emphasized up to this point: the perturbation $u(k, \tau)$ produced by an expanding universe characterized by $\epsilon$ actually matches the perturbation $\hat{u}(k/c, c\tau)$ produced by a contracting universe characterized by $\hat{\epsilon} = 1/\epsilon$. Here $c$ is an arbitrary constant, which ultimately arises from the well-known fact that the normalization of the scale factor $a(\tau)$ is arbitrary in a spatially-flat ($K = 0$) universe. When $K \neq 0$, it turns out [21] that the duality generalizes as follows. The perturbation $u(k, y)$ produced by an expanding universe characterized by $\epsilon$ matches the perturbation $\hat{u}(\hat{k}, \hat{y})$ produced by a contracting universe characterized by $\hat{\epsilon} = 1/\epsilon$, provided $k$ and $\hat{k}$ are related by

$$\frac{1}{(\epsilon-1)^2}\left[\frac{(2k)^2}{(d-2)^2}+K(\epsilon-2)^2+4K\frac{\epsilon-3}{d-2}\right]=\frac{1}{(\hat{\epsilon}-1)^2}\left[\frac{(2\hat{k})^2}{(d-2)^2}+K(\hat{\epsilon}-2)^2+4K\frac{\hat{\epsilon}-3}{d-2}\right]. \quad (4.50)$$

Here $\epsilon$ is given by (4.42), and we have defined the time coordinate $y \equiv |\tau/q|$, where $q = \frac{2}{(d-2)(\epsilon-1)}$. Since $K$ is non-zero, the normalization of the scale factor $a(\tau)$ is no longer arbitrary. This is why there is no longer an arbitrary constant $c$ in the duality, and instead there is a fixed relationship between $\{k, y\}$ and $\{\hat{k}, \hat{y}\}$.

## 4.9 Generalization to time-varying $\epsilon$

In the previous section, we have seen that the duality easily generalizes to FRW models with arbitrary spacetime dimension $d$, as long as $\epsilon$ is constant or slowly-varying. Now we would like to generalize to the situation where $\epsilon$ is time-varying: given an expanding (inflationary) background solution with $\epsilon = \epsilon(\tau)$, what other background solutions produce the same spectrum of $u$ perturbations? Preferably, we would like to find an answer to this question that retains the simplicity and elegance of the $\epsilon \to 1/\epsilon$ prescription. We do not yet have the final answer to this intriguing challenge, but we will describe a few interesting partial answers, with the hope that they will provide clues to uncovering the full story.

We have seen that $u$ satisfies the equation of motion (4.12a) and the boundary condition (4.14a), where

$$\theta^2 = \frac{a^{-d} a'^2}{(d-2)[2(a'/a)^2 - a''/a]}. \tag{4.51}$$

Note that we have used Eqs. (4.3), (4.41) and (4.45) to rewrite $\theta(\tau)$ purely in terms of $a(\tau)$ and its derivatives.

Two models, with scale factors $a(\tau)$ and $\hat{a}(\tau)$, will share the same $u(k,\tau)$ provided

$$\widehat{\theta}''/\widehat{\theta} = \theta''/\theta. \tag{4.52}$$

It is useful to note that the differential condition (4.52) is equivalent to the integral condition

$$1 = \left(A_1 + A_2 \int^\tau \frac{\mathrm{d}\tau}{\theta^2}\right)\left(\widehat{A}_2 + \widehat{A}_2 \int^\tau \frac{\mathrm{d}\tau}{\widehat{\theta}^2}\right) \tag{4.53}$$

for arbitary constants $A_1$, $A_2$, $\widehat{A}_1$ and $\widehat{A}_2$.

If we are given a fiducial background solution with scale factor $a(\tau)$, then we can use (4.51) to compute $\theta(\tau)$. Next we can solve (4.52) to find a 2-parameter family of solutions:

$$\widehat{\theta} = \theta(\tau)\left[B_1 + B_2 \int^\tau \frac{d\tau'}{\theta^2(\tau')}\right]. \tag{4.54}$$

And finally, for each function $\widehat{\theta}(\tau)$ in this 2-parameter family, we can solve (4.51) to find a 2-parameter family of solutions for $\hat{a}(\tau)$. By this counting, we find that each fiducial background solution $a(\tau)$ should give rise to a 4-parameter family of background solutions

$\hat{a}(\tau)$ that generate the same $u(k,\tau)$. It is easy to see, though, that one of these 4 parameters is unphysical — it corresponds to the freedom to rescale the spatial coordinates, and hence $a(\tau)$, by an arbitary constant. Thus, we expect to find a 3-parameter family of genuinely distinct solutions $\hat{a}(\tau)$.

In some sense, then, these expressions — (4.51) together with (4.52) or (4.53) — provide an answer to the question of how the duality generalizes when $\epsilon = \epsilon(\tau)$ is time-varying. But it would certainly be nicer to recast this generalization in a simpler, more explicit, and hopefully more elegant form — something closer in spirit to $\epsilon \to 1/\epsilon$. One way to achieve this goal would be to solve the equations (4.51) and (4.52) directly. At first glance, this seems like it would be very difficult since, for a given fiducial background solution $a(\tau)$, these equations together constitute a fourth-order nonlinear differential equation for $\hat{a}(\tau)$. Surprisingly, though, it turns out to be possible to completely solve the problem in full generality in $2+1$ dimensions, and at least partially solve it in $3+1$ dimensions. And, what's more, the solutions turn out to again be surprisingly simple and elegant, just as we had hoped! This hints that in $3+1$ dimensions, an elegant and fully-general solution — a proper generalization of the $\epsilon \to 1/\epsilon$ duality — may still be lurking.

Let us therefore sketch the general solution in $2+1$ dimensions, and the partial solution in $3+1$ dimensions, in the hopes that these may provide clues to finding the general solution in $3+1$ dimensions.

### 4.9.1 $2+1$ dimensions

In $2+1$ spacetime dimensions ($d=3$), we can evaluate the integrals in (4.53) to obtain

$$1 = \left(A_1 + \frac{A_2}{H(\tau)}\right)\left(\widehat{A}_1 + \frac{\widehat{A}_2}{\widehat{H}(\tau)}\right) \tag{4.55}$$

where $H(\tau) = a'/a^2$ is the Hubble parameter. Rearranging, we find that the condition $\widehat{\theta}''/\widehat{\theta} = \theta''/\theta$ is equivalent, in 2+1 dimensions, to the condition that

$$\widehat{H}(\tau) = \frac{B_1 H(\tau) + B_2}{B_3 H(\tau) + B_4}. \tag{4.56}$$

for arbitrary constants $B_1$, $B_2$, $B_3$ and $B_4$. This is a complete and elegant description of the set fo transformations that leave $u(k,\tau)$ invariant in $2+1$ dimensions. Before moving on, let us use this transformation to infer a few other interesting results.

We can integrate (4.56) to find

$$\widehat{a}(\tau) = -\left[\int^{\tau} \frac{B_1 H(\tau') + B_2}{B_3 H(\tau') + B_4} \mathrm{d}\tau'\right]^{-1}. \tag{4.57}$$

Alternatively, by differentiating (4.56) we obtain

$$\widehat{H}' = \frac{(B_1 B_4 - B_2 B_3) H'}{(B_3 H + B_4)^2}. \tag{4.58}$$

It is easy to show that our single scalar field system must satisfy the condition $H'(\tau) \leq 0$ (because it satisfies the weak energy condition). But, assuming $a(\tau)$ satisfies the condition $H'(\tau) \leq 0$, then $\widehat{a}(\tau)$ will also satisfy the corresponding condition $\widehat{H}'(\tau) \leq 0$ if and only if $B_1 B_4 - B_2 B_3 > 0$. Furthermore, note that the transformation law (4.56) is invariant if we rescale all the $B_i$'s by an arbitrary constant: $B_i \to C B_i$. Thus, we can always take the $B_i$'s to satisfy

$$B_1 B_4 - B_2 B_3 = 1. \tag{4.59}$$

Now, instead of thinking of the right-hand-side of (4.56) as a fraction, think of the numerator and denominator as the components of a 2-component vector. Then we can think of the transformation (4.56) as a *linear* transformation:

$$\widehat{H} = \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix} \begin{pmatrix} H \\ 1 \end{pmatrix}. \tag{4.60}$$

This is a good way to think about it since, as one can check, successive transformations of the form (4.56) simply compose under ordinary matrix multiplication. In combination with (4.59), this shows that the group of transformations of $H(\tau)$ leaving $u$ invariant is just $SL(2,\mathbf{R})$.

Next, let us characterize the set of transformations that leave both $u(k,\tau)$ and $v(k,\tau)$ simultaneously invariant. First recall that $v$ may be obtained from $u$ via the equation (4.13a)

$$v = 2[u' - (\theta'/\theta)u]. \tag{4.61}$$

After the transformation (4.56), we have

$$\widehat{v} = 2[\widehat{u}' - (\widehat{\theta}'/\widehat{\theta})\widehat{u}]. \tag{4.62}$$

Now, using $\widehat{u} = u$ and $\theta = H(-H')^{-1/2}$, which implies

$$\frac{\widehat{\theta}'}{\widehat{\theta}} = \frac{\theta'}{\theta} - \frac{B_2 H'}{H(B_1 H + B_2)}, \tag{4.63}$$

we find

$$\widehat{v} = v + \frac{2B_2 H'}{H(B_1 H + B_2)} u. \tag{4.64}$$

So there are two cases in which $\widehat{v} = v$: a general case and a special case. The general case is $B_2 = 0$. That is, the transformations (4.56) that leave *both* $u$ *and* $v$ invariant form the lower-triangular subgroup of $SL(2, \mathbf{R})$. The special case is $H'(\tau) = 0$, corresponding to pure deSitter. The transformation (4.56) is degenerate in this case, and just maps deSitter solutions to other deSitter solutions.

Finally, let us consider "dominant" and "subdominant" scalar perturbations (see section 4.4). When $\epsilon$ was nearly constant, we showed that the dominant and subdominant modes were both invariant under the transformation $\epsilon \to 1/\epsilon$. Let us now show that, in $2+1$ dimensions, the dominant and subdominant modes are both invariant under the full set of transformations (4.56), even for arbitrary time-dependent $\epsilon(\tau)$. First construct the subdominant mode from $u$ and $v$:

$$f_{\rm sub}(x) = v - \left[\lim_{k\to 0} \frac{v}{u}\right] u \tag{4.65}$$

where the limit is taken holding $\tau$ fixed. Thus, under the transformation (4.56), this becomes

$$\widehat{f}_{\rm sub}(x) = \widehat{v} - \left[\lim_{k\to 0} \frac{\widehat{v}}{\widehat{u}}\right] \widehat{u}. \tag{4.66}$$

Then, using $\widehat{u} = u$ and equation (4.64) for $\widehat{v}$, we immediately find

$$\widehat{f}_{\rm sub}(x) = f_{\rm sub}(x). \tag{4.67}$$

Now, assuming $u$ and $f_{\rm sub}$ are linearly independent, we can choose $f_{\rm dom} = u$. So both $f_{\rm sub}(x)$ (automatically) and $f_{\rm dom}(x)$ (by choice) are both invariant under the full $SL(2, \mathbf{R})$ symmetry represented by the transformations (4.56).

### 4.9.2 $3+1$ dimensions

Note that in $3+1$ spacetime dimensions ($d=4$), we can rewrite $\theta''/\theta$ as

$$\frac{\theta''}{\theta} = g + \frac{(g^{-1/2})''}{(g^{-1/2})}, \tag{4.68}$$

where

$$g \equiv \frac{(1/a)''}{(1/a)}. \tag{4.69}$$

Thus, if two background solutions $a(\tau)$ and $\hat{a}(\tau)$ generate functions $g(\tau)$ and $\hat{g}(\tau)$ such that $g(\tau) = \hat{g}(\tau)$, then it follows from (4.68) that $\theta''/\theta = \widehat{\theta}''/\widehat{\theta}$. But, given a fiducial background solution $a(\tau)$, the family of background solutions $\hat{a}(\tau)$ that generate the same $g(\tau)$ is simply given by

$$\hat{a}(\tau) = a(\tau)\left[C_1 + C_2\int^{\tau} a^2(\tau')\mathrm{d}\tau'\right]^{-1}. \tag{4.70}$$

Thus, we have found a partial solution to the full problem: we expect a 4-parameter family of solutions $\hat{a}(\tau)$ that generate the same $u(k,\tau)$, and we have found a 2-parameter subfamily. One of the parameters (in the full 4-parameter family, and also in our 2-parameter subfamily) corresponds to an unphysical rescaling of the spatial coordinates and the scale factor, as discussed above. So it is more correct to say that we have found a 1-parameter subfamily of the full 3-parameter family of physically-distinct background solutions $a(\tau)$ that generate the same $u(k,\tau)$ in $3+1$ dimensions.

It is interesting to explore this 1-parameter subfamily in a particular example. Let the fiducial background solution be $a(\tau) = |\tau|^p$ (often called power-law inflation). Until now, power-law inflation has been the only example in the literature of a model where one can solve exactly for both the background evolution, as well as the evolution of the scalar and tensor perturbations. It has therefore provided an important cross-check for numerical calculations and analytic approximations (notably the slow-roll approximation). We will now show that, by applying the techniques discussed above, we obtain a new family of background solutions — a generalization of power-law inflation — in which it is possible to solve exactly for both the background evolution as well as the evolution of the scalar and tensor perturbations. Considering their generality, these solutions are actually rather

simple, and we hope that they may be useful in some of the contexts where the exact solutions provided by power law inflation have been useful in the past.

To solve for the background evolution in this example, we first use Eq. (4.70) to obtain a family of scale factors $a(\tau)$. We then use the Friedmann equations (4.2) to solve for the corresponding scalar field evolution $\phi(\tau)$, and the scalar field potential $V(\phi)$.

$$a(\tau) = a_0 \left( r_+ T^{L^+/2} + r_- T^{L^-/2} \right)^{-1} \tag{4.71a}$$

$$\phi(\tau) = -L \ln T \tag{4.71b}$$

$$V(\phi) = V_0 \left[ r_+^2 L^+ (L^+ + 1) \mathrm{e}^{\phi L^-/L} + 4 r_+ r_- L^+ L^- \mathrm{e}^{\phi/L} + r_-^2 L^- (L^- + 1) \mathrm{e}^{\phi L^+/L} \right]. \tag{4.71c}$$

Here $a_0$, $\tau_0$, $r_+$ and $r_-$ denote arbitrary constants and for brevity we have used the notation $T \equiv |\tau/\tau_0|$ and defined the constants

$$\alpha \equiv \frac{1}{2} \left| \frac{1+\epsilon}{1-\epsilon} \right| \tag{4.72a}$$

$$L^\pm \equiv 1 \pm 2\alpha \tag{4.72b}$$

$$L \equiv \sqrt{-L^+ L^-/2} \tag{4.72c}$$

$$V_0 \equiv \frac{1}{2 a_0^2 \tau_0^2}. \tag{4.72d}$$

For reference, here are several useful background quantities that are directly obtained from

the background solutions (4.71)

$$\theta = \frac{T^{1/2}}{2a_0 L}\left[g_+(\tau)+g_-(\tau)\right] \tag{4.73a}$$

$$z = \frac{2a_0 L}{T^{1/2}}\left[g_+(\tau)+g_-(\tau)\right]^{-1} \tag{4.73b}$$

$$\frac{a'}{a} = \frac{-1}{2\tau}\left[\frac{r_+ L^+ T^\alpha + r_- L^- T^{-\alpha}}{r_+ T^\alpha + r_- T^{-\alpha}}\right] \tag{4.73c}$$

$$\frac{\theta'}{\theta} = \frac{+1}{2\tau}\left[\frac{L^+ g_+(\tau) + L^- g_-(\tau)}{g_+(\tau)+g_-(\tau)}\right] \tag{4.73d}$$

$$\frac{z'}{z} = \frac{-1}{2\tau}\left[\frac{L^+ g_+(\tau) + L^- g_-(\tau)}{g_+(\tau)+g_-(\tau)}\right] \tag{4.73e}$$

$$\frac{a''}{a} = \frac{r_+^2 L^+(L^++2)T^{2\alpha} + 6 r_+ r_- L^+ L^- + r_-^2 L^-(L^-+2)T^{-2\alpha}}{4\tau^2(r_- T^\alpha + r_- T^{-\alpha})^2} \tag{4.73f}$$

$$\frac{\theta''}{\theta} = -\frac{L^+ L^-}{4\tau^2} \tag{4.73g}$$

$$\frac{z''}{z} = \frac{L^+(L^++2)g_+^2(\tau) + 6L^+L^- g_+(\tau)g_-(\tau) + L^-(L^-+2)g_-^2(\tau)}{4\tau^2[g_+(\tau)+g_-(\tau)]^2} \tag{4.73h}$$

where we have defined

$$g_\pm(\tau) \equiv r_\pm L^\pm T^{\pm\alpha}. \tag{4.74}$$

Next, we can substitute these background quantities into the equation of motions of motion — (4.12a) for the scalar perturbation variable $u$, (4.12b) for the scalar perturbation variable $v$, and (5.9) for the tensor perturbation variable $f_T$ — and thus find the general solutions

$$u = \sum_{n=1,2} A_n(k) x^{1/2} H_\alpha^{(n)}(x) \tag{4.75a}$$

$$v = \sum_{n=1,2} B_n(k) x^{1/2}\left[\frac{g_+(\tau)H_{\alpha+1}^{(n)}(x) - g_-(\tau)H_{\alpha-1}^{(n)}(x)}{g_+(\tau)+g_-(\tau)}\right] \tag{4.75b}$$

$$f_T = \sum_{n=1,2} C_n(k) x^{1/2}\left[\frac{r_+ T^\alpha H_{\alpha+1}^{(n)}(x) - r_- T^{-\alpha} H_{\alpha-1}^{(n)}(x)}{r_+ T^\alpha + r_- T^{-\alpha}}\right] \tag{4.75c}$$

where $A_1(k)$, $A_2(k)$, $B_1(k)$, $B_2(k)$, $C_1(k)$, and $C_2(k)$ are (possibly $k$-dependent) constants. To determine these constants, we can impose the boundary conditions — (4.14a) for $u$,

(4.14b) for $v$, and (4.36) for $f_T$ — to obtain

$$u = \frac{\mathcal{P}}{2k}\left(\frac{\pi x}{4k}\right)^{1/2} H_\alpha^{(1)}(x), \tag{4.76a}$$

$$v = \mathcal{P}\left(\frac{\pi x}{4k}\right)^{1/2}\left[\frac{g_+(\tau)H_{\alpha+1}^{(1)}(x) - g_-(\tau)H_{\alpha-1}^{(1)}(x)}{g_+(\tau)+g_-(\tau)}\right], \tag{4.76b}$$

$$f_T = \mathcal{P}\left(\frac{\pi x}{4k}\right)^{1/2}\left[\frac{r_+T^\alpha H_{\alpha+1}^{(1)}(x) - r_-T^{-\alpha}H_{\alpha-1}^{(1)}(x)}{r_+T^\alpha + r_-T^{-\alpha}}\right], \tag{4.76c}$$

where we have defined the complex phase

$$\mathcal{P} \equiv \exp[i(2\alpha+3)\pi/4]. \tag{4.77}$$

In particular, notice that the $u$ solution is completely independent of the parameters $r_+$ and $r_-$, and is further invariant under the transformation $\epsilon \to 1/\epsilon$. In other words, regardless of the choice of $r_+$ and $r_-$, the solution (4.19a) for $u(k,\tau)$ is precisely the one that would be produced by a simple power-law inflation model characterized by equation of state $\epsilon$. By contrast, the $v$ and $f_T$ solutions depend explicitly on $r_+$ and $r_-$ and are not invariant under the transformation $\epsilon \to 1/\epsilon$.

## 4.10 Discussion

Beyond its inherent theoretical interest, our duality may be observationally relevant if (as discussed in section 4.1) long-wavelength correlations produced during a contracting phase successfully propagate into a subsequent expanding phase. This suggests a fundamental degeneracy: an ideal measurement of the "primordial" scalar perturbation spectrum may be unable to determine whether the perturbations were generated by an expanding phase or by its contracting dual. Luckily, as shown in section 4.5, tensor perturbations break this degeneracy: a contracting model produces a much bluer tensor spectrum than its expanding dual. In particular, a detection of tensors in the cosmic microwave background would indicate that these perturbations were generated in an expanding phase, since the dual contracting phase would produce an undetectably small tensor spectrum on these cosmological length scales [93, 23].

We have explained how the duality works in a broad range of contexts, but a number of interesting questions remain. What, if anything, is one to make of the geometrical aspects of the duality, discussed in section 4.7? What is the best way to formulate and understand the duality physically, and what is its general form? In particular, is there a simple and general rule relating models, even when $\epsilon$ is time-dependent, that retains the elegance of the $\epsilon \to 1/\epsilon$ prescription? There are hints that the answer may be "yes," and the previous section discusses several ideas along this line. We have studied the duality in the context of a simple model — an FRW spacetime containing a single scalar field $\phi$ with potential $V(\phi)$ and a canonical kinetic term; but what happens in more complicated models? And we have used linear perturbation theory; but what happens in the nonlinear regime?

# Chapter 5

# The gravitational wave spectrum from the cyclic model

This last chapter is based on work done in collaboration with Paul Steinhardt and Neil Turok. Apart from fairly minor updates and modifications, it originally appeared as [23]. I thank Andrew Tolley and Justin Khoury for helpful conversations.

## 5.1 Introduction

The recently-proposed cyclic model[159, 158] differs radically from standard inflationary cosmology [74, 107, 4], while retaining the inflationary predictions of homogeneity, flatness, and nearly scale-invariant density perturbations. It has been suggested that the cosmic gravitational wave background provides the best experimental means for distinguishing the two models. Inflation predicts a nearly scale-invariant (slightly red) spectrum of primordial tensor perturbations, whereas the cyclic model predicts a blue spectrum [93]. The difference arises because inflation involves an early phase of hyper-rapid cosmic acceleration, whereas the cyclic model does not.

In this chapter, we compute the gravitational wave spectrum for cyclic models to obtain both the normalization and spectral shape as a function of model parameters, improving

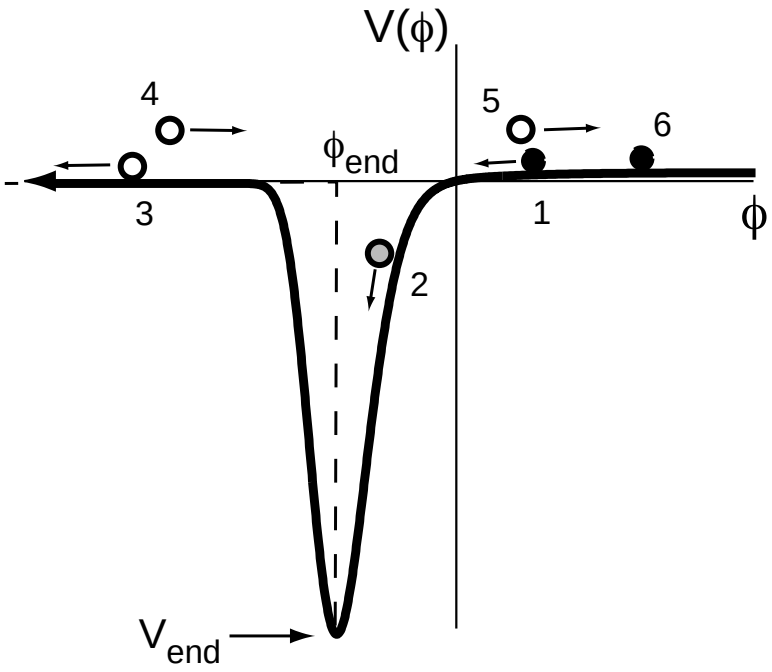



Figure 5.1: Schematic of cyclic potential with numbers representing the stages described in the text. To the left of $\phi_{end}$, where the scalar kinetic energy dominates, we approximate $V$ with a Heaviside function, jumping to zero as shown by the dashed line.

upon earlier heuristic estimates. We make the assumption that perturbations pass smoothly through the bounce, as discussed in the previous chapter. Under this assumption, we find that the spectrum is strongly blue. The amplitude is too small to be observed by currently proposed detectors on all scales. Hence, the discovery of a stochastic background of gravitational waves would be evidence in favor of inflation, and would rule out the cyclic model.

Readers unfamiliar with the cyclic model may consult [160] for an informal tour, and [96] for a recent analysis of phenomenological constraints. Cyclic cosmology draws strongly on earlier ideas associated with the "ekpyrotic universe" scenario [93, 91, 94]. Briefly, the scenario can be described in terms of the periodic collision of orbifold planes moving in an extra spatial dimension, or, equivalently, in terms of a four-dimensional theory with an evolving (modulus) field $\phi$ rolling back and forth in an effective potential $V(\phi)$. The

potential (Fig. 1) is small and positive for large $\phi$, falling steeply negative at intermediate $\phi$, and increasing again for negative $\phi$.

Each cycle consists of the following stages: (1) $\phi$ large and decreasing: the universe expands at an accelerated rate as $V(\phi) > 0$ acts as dark energy; (2) $\phi$ intermediate and decreasing: the universe is dominated by a combination of scalar kinetic and potential energy, leading to slow contraction and to the generation of fluctuations; (3) $\phi$ negative and decreasing (beginning at conformal time $\tau_{end} < 0$): the generation of fluctuations ends, $\phi$ rolls past $\phi_{end}$ so that the string coupling vanishes as $\phi$ runs off to $-\infty$ and, in the four-dimensional description, the universe contracts rapidly, dominated by scalar field kinetic energy, to the bounce ($\tau = 0$) at which matter and radiation are generated; (4) $\phi$ increasing from minus infinity: the universe remains dominated by scalar field kinetic energy, which decreases rapidly compared to the radiation energy; (5) $\phi$ large and increasing (beginning at $\tau_r > 0$): the scalar field kinetic energy red-shifts to a negligible value and the universe begins the radiation dominated expanding phase; (6) $\phi$ large and nearly stationary: the universe undergoes the transitions to matter and dark energy domination, and the cycle begins anew.

## 5.2 Background evolution

During stage 2 ($\phi > \phi_{end}$), the potential $V(\phi)$ must be exponentially steep to produce acceptable scalar perturbations. But when $V(\phi)$ ceases to be exponentially steep (stage 3, $\phi < \phi_{end}$), the potential energy becomes negligible compared to the kinetic energy, which blue shifts and dominates the universe. Since $\phi$ acts like a free scalar field in this regime, we may simplify calculations by setting $V(\phi)$ to zero without any loss of generality, as discussed in [159, 96]. We therefore model the potential as:

$$V(\phi) = V_0(1 - e^{-c\phi/M_{pl}})\Theta(\phi - \phi_{end}) \tag{5.1}$$

where $M_{pl}$ is the reduced Planck mass and the Heaviside step function $\Theta(\phi - \phi_{end})$ sets $V(\phi)$ to zero when $\phi < \phi_{end}$. Choosing $c = 10$ for example results in a scalar spectral index

$n_s = .96$ which is compatible with current constraints.

Our calculation begins in the "ekpyrotic phase," stage (2), with the Einstein-frame scale factor contracting:

$$a(\tau) = a_{end}\left(\frac{\tau - \tau_{ek}}{\tau_{end} - \tau_{ek}}\right)^{\alpha}, \qquad \tau < \tau_{end}\ , \tag{5.2}$$

where $\alpha \equiv 2/(c^2 - 2) \ll 1$ and $\tau_{ek} \equiv (1 - 2\alpha)\tau_{end}$, being the conformal time the potential would have diverged to minus infinity had the exponential form continued. At $\tau = \tau_{end}$, the ekpyrotic phase ends and the "contracting kinetic phase," stage (3), begins:

$$a(\tau) = \left(\frac{-\tau}{(1+\chi)\tau_r}\right)^{1/2}, \qquad \tau_{end} < \tau < 0\ . \tag{5.3}$$

At $\tau = 0$, the universe bounces and the "expanding kinetic phase," stage (4), begins:

$$a(\tau) = \left(\frac{\tau}{\tau_r}\right)^{1/2}, \qquad 0 < \tau < \tau_r\ . \tag{5.4}$$

Radiation is produced at the bounce, but is less than the scalar kinetic energy until, at $\tau = \tau_r$, the expanding kinetic phase ends, and standard radiation-dominated, matter-dominated, and dark-energy-dominated epochs ensue. The transition times, $\tau_r$ and $\tau_{end}$, are given by

$$\tau_r = (\sqrt{2}H_r)^{-1}, \qquad \tau_{end} = -\tau_r/\Gamma, \tag{5.5}$$

and

$$\Gamma \equiv \left|\frac{\tau_r}{\tau_{end}}\right| = \left[\frac{1}{1+\chi}\left(\frac{2\alpha}{1-2\alpha}\right)\left(\frac{V_{end}}{H_r^2 M_{pl}^2}\right)\right]^{1/3}, \tag{5.6}$$

where $H_r \equiv H(\tau_r)$ is the Hubble constant at $\tau_r$, $V_{end} = -V(\phi_{end})$ is the depth of the potential at its minimum, and $\chi \ll 1$ is a small positive constant that measures the amount of radiation created at the bounce. Note that $a(\tau)$ and $a'(\tau)$ are both continuous at the transition time $\tau = \tau_{end}$, and we have chosen to normalize $a(\tau)$ to unity at the start of radiation domination ($a(\tau_r) = 1$).

## 5.3 Primordial strain spectrum, $\Delta h(k, \tau_r)$.

A quasi-stationary stochastic background of gravitational waves is characterized by the quantity $\Delta h(k, \tau)$, the rms dimensionless strain per unit logarithmic wavenumber at time

$\tau$. Accounting for both polarizations, it is given by $\Delta h(k,\tau) = k^{3/2}|h_k(\tau)|/\pi$, where the Fourier amplitude $h_k(\tau)$ satisfies

$$h_k'' + 2\frac{a'}{a}h_k' + k^2 h_k = 0\ . \tag{5.7}$$

In the cyclic model, all modes inside the horizon today exited the horizon during the contracting phase, and re-entered during the expanding phase. Early in the ekpyrotic phase (that is, in stage (2), with $\tau \to -\infty$), these modes were far inside the Hubble volume, which had been smoothed, flattened and cleaned of debris by the dark energy epoch, stage (1). All modes of interest are therefore expected to be in their usual Minkowski vacuum at the start of stage (2), implying the boundary condition

$$h_k(\tau) \to \frac{e^{-ik\tau}}{a(\tau)M_{pl}\sqrt{2k}} \quad \text{as} \quad \tau \to -\infty. \tag{5.8}$$

To solve equation (5.7), it is useful to define $f_k(\tau) \equiv a(\tau)\, h_k(\tau)$ and rewrite (5.7) as

$$(f_k)'' + (k^2 - \frac{a''}{a})f_k = 0\ . \tag{5.9}$$

During the ekpyrotic phase, $a(\tau)$ is given by (5.2), and the general solution of (5.9) is

$$f_k(\tau) = \sqrt{y}\left(A_1(k)H_n^{(1)}(y) + A_2(k)H_n^{(2)}(y)\right), \tag{5.10}$$

where $A_{1,2}(k)$ are arbitrary constants, $n \equiv \frac{1}{2} - \alpha$, $y \equiv -k(\tau - \tau_{ek})$, and $H_n^{(1,2)}$ are the Hankel functions. Asymptotically, $H_n^{(1,2)}(y) \to \sqrt{\frac{2}{\pi y}}\mathrm{e}^{\pm iy}$, so (5.8,5.10) imply

$$A_1(k) = \frac{1}{2}\sqrt{\frac{\pi}{k}}, \qquad A_2(k) = 0, \tag{5.11}$$

where we have dropped a physically irrelevant phase. In the contracting kinetic phase, stage (4), $a(\tau)$ is given by (5.3), and the general solution of (5.9) is

$$f_k(\tau) = \sqrt{-k\tau}\Big(B_1(k)H_0^{(1)}(-k\tau) + B_2(k)H_0^{(2)}(-k\tau)\Big) \tag{5.12}$$

where $B_{1,2}(k)$ are arbitrary constants. Then, continuity of $h_k$ and $h_k'$ at $\tau = \tau_{end}$ implies

$$\begin{aligned} B_{1,2}(k) \;&=\; \mp\frac{i\pi}{4}\sqrt{\frac{\pi\alpha}{2k}}x_e\Big[H_1^{(2,1)}(x_e)H_n^{(1)}(2\alpha x_e) + \\ &\quad + H_0^{(2,1)}(x_e)H_{n-1}^{(1)}(2\alpha x_e)\Big] \end{aligned} \tag{5.13}$$

where $x_e \equiv k|\tau_{end}|$. Finally, in the expanding kinetic phase, $a(\tau)$ is given by (5.4), and the general solution of (5.9) is

$$f_k(\tau) = \sqrt{k\tau}\left(C_1(k)H_0^{(1)}(k\tau) + C_2(k)H_0^{(2)}(k\tau)\right). \tag{5.14}$$

From (5.7), each gravity wave polarization acts just like a massless scalar, and QFT for a massless scalar in the contracting/expanding Milne geometry near the cyclic model's bounce has already been treated by Tolley *et al* in [163, 164]. (Our problem is even simpler: tensor perturbations are gauge-invariant.) They find a unique sensible matching condition. One can think of continuing the fields from the contracting Milne wedge to the expanding one through the Minkowski space in which they are both naturally embedded. Equivalently, one may analytically continue the positive (negative) frequency parts of $h_k \equiv f_k/a$ around the origin in the lower (upper) half of the complex $\tau$-plane, so $H_0^{(1,2)}(-k\tau) \to -H_0^{(2,1)}(k\tau)$. [163, 164] This yields

$$C_{1,2}(k) = -\sqrt{1+\chi}\, B_{2,1}(k)\,. \tag{5.15}$$

The pre-factor arises because $a(\tau)$ differs by a factor of $\sqrt{1+\chi}$ between the kinetic contracting and expanding phases; see Eqs. (5.3) and (5.4). Combining our results, we arrive at the "primordial" dimensionless strain spectrum at the beginning of the radiation dominated epoch:

$$\begin{aligned}\Delta h(k,\tau_r) = \quad & (k^2/\pi M_{pl})\,\sqrt{2(1+\chi)\tau_r} \\ & \left|B_2(k)H_0^{(1)}(x_r) + B_1(k)H_0^{(2)}(x_r)\right|\end{aligned}$$

where $x_r \equiv k\tau_r$ and $k < k_{end}$. For $k > k_{end}$, the spectrum is cut off because these modes are not amplified and, instead, Eq. (5.16) converges to the result for a static Minkowski background.

## 5.4 Present-day strain spectrum, $\Delta h(k, \tau_0)$.

To convert from the primordial spectrum to the present-day spectrum $\Delta h(k,\tau_0) \equiv T_h(k)\Delta h(k,\tau_r)$ we need to know the transfer function, $T_h(k)$. To approximate $T_h(k)$, note that $\Delta h(k,\tau)$ is roughly time-independent outside the horizon, and decays as $a^{-1}$ once a mode re-enters

the horizon. Therefore, the transfer function is $\sim 1/(1+z_r)$ for modes already inside the horizon at the onset of radiation domination $\tau_r$, and $\sim 1/(1+z_k)$ for modes that entered at red shift $z_k$ between $\tau_r$ and $\tau_0$. Using the fact that $H \propto a^{-2}$ during radiation domination and $H \propto a^{-3/2}$ during matter domination (neglecting the change in $g_*$) we find

$$T(k) \approx \left(\frac{k_0}{k}\right)^2 \left[1 + \frac{k}{k_{eq}} + \frac{k^2}{k_{eq}k_r}\right] \tag{5.16}$$

where $k_0 \equiv a_0 H_0$, $k_{eq} \equiv a_{eq} H_{eq}$, $k_r \equiv a_r H_r$, and $k_{end} \equiv a_{end}|H_{end}|$ denote the modes on the horizon today ($\tau_0$), at matter-radiation equality ($\tau_{eq}$), at the start of radiation domination ($\tau_r$), and at the end of the ekpyrotic phase ($\tau_{end}$), respectively.

The gravitational wave spectrum can be divided into three regimes. There is a low frequency (LF) regime corresponding to long wavelength modes that re-enter after matter-radiation equality ($k < k_{eq}$), and a medium frequency (MF) regime consisting of modes which re-enter between equality and the onset of radiation domination ($k_{eq} < k < k_r$). (We ignore the recent dark energy dominated phase, which has negligible effect.) The spectrum for these two regimes is:

$$\Delta h \approx \frac{\Gamma^{\frac{1}{2}} k_0^2}{\pi M_{pl} H_r^\alpha} \begin{cases} k^{-1+\alpha} & (LF) \\ k^\alpha / k_{eq} & (MF) \end{cases} \tag{5.17}$$

Finally, modes which exit the horizon during the ekpyrotic phase (before $\tau_{end}$), and re-enter during the expanding kinetic phase (after the bound but before $\tau_r$) result in a high frequency (HF) band ($k_r < k < k_{end}$):

$$\Delta h \approx \left(\frac{\sqrt{2}}{\pi}\right)^{\frac{3}{2}} \frac{(\Gamma H_r)^{\frac{1}{2}-\alpha} k_0^2}{M_{pl} k_{eq} k_r} \left|\cos\left(k\tau_r - \frac{\pi}{4}\right)\right| k^{\frac{1}{2}+\alpha} \quad (HF) \tag{5.18}$$

In this expression, we can replace the oscillatory factor cos(...) factor by its averaged value, because it is such a rapidly oscillating function of $k$ that these oscillations cannot be observed by any known experiment [6]. The HF band runs over a range $k_{end}/k_r = \Gamma$, and this quantity is strongly constrained by the requirement that the scalar field cross the negative region of the potential before radiation domination begins, which requires that [96]

$$H_r \lesssim \frac{V_{end}^{\frac{1}{2}}}{M_{Pl}} \left(\frac{V_0}{V_{end}}\right)^{\sqrt{\frac{3}{2}}/c}, \tag{5.19}$$

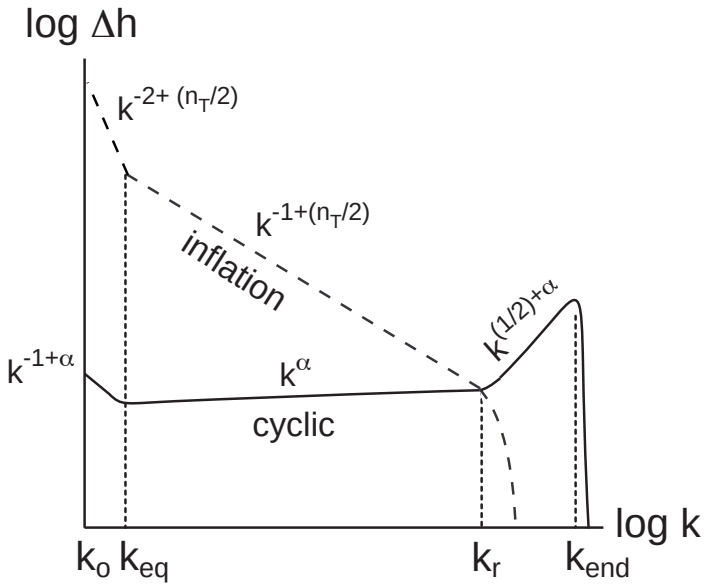



Figure 5.2: A schematic comparison of the dimensionless strain observed today $\Delta h(k, \tau_0)$, as predicted by inflation and the cyclic model. Here $n_T$ is the inflationary tensor spectral index (a small negative number), and $\alpha \ll 1$ in the cyclic model is a small positive number. $k_r$ denotes the mode on the horizon at the start of radiation domination.

where $V_0$ is today's value of the dark energy density. This equation, combined with (5.6), gives a lower bound on $\Gamma$, $\Gamma \gtrsim (V_{end}/V_0)^{\sqrt{2/3c^2}}$. For example, for $V_{end}$ around the GUT scale and $c = 10$, we find $\Gamma \geq 10^8$. Fig. 5.2 schematically depicts $\Delta h(k, \tau_0)$ in the cyclic scenario and compares it to the inflationary spectrum. [168]

Another useful quantity is $\Omega_{gw}(k, \tau_0)$, the gravitational wave energy per unit logarithmic wavenumber, in units of the critical density [161, 168]

$$\Omega_{gw}(k, \tau_0) \equiv \frac{k}{\rho_{cr}} \frac{d\rho_{gw}}{dk} = \frac{1}{6} \left( \frac{k}{k_0} \right)^2 \Delta h(k, \tau_0)^2 \ . \tag{5.20}$$

In the cyclic model, $\Omega_{gw}(k, \tau_0)$ is very blue, with nearly all the gravitational wave energy concentrated at the high-frequency end of the distribution.

## 5.5 Observational constraints and detectability.

The strongest observational constraint on the gravitational spectrum in the cyclic model comes from the requirement that the successful predictions of big bang nucleosynthesis (BBN) not be affected, which requires

$$\int_{k_{BBN}}^{k_{end}} \Omega_{gw}(k, \tau_0) \frac{dk}{k} \lesssim \frac{0.1}{1 + z_{eq}} \ . \tag{5.21}$$

From the above equations, (5.18) and (5.20), and using $1 + z_{eq} \approx k_{eq}^2/k_0^2$, and $T_r \sim H_r^{\frac{1}{2}} M_{Pl}^{\frac{1}{2}}$ for the temperature at radiation domination, we obtain a total $\Omega$ in gravitational waves of $\sim (2\alpha V_{end}/T_r M_{Pl}^3)^{\frac{4}{3}} [36\pi^3 (1 + z_{eq})]^{-1}$. From (5.21), this implies

$$T_r \gtrsim \frac{\alpha}{20} V_{end} M_{Pl}^{-3}, \tag{5.22}$$

where, for simplicity, we have ignored the factors which depends on the number of thermal degrees of freedom, and further weaken this bound.

The other observational constraints are much weaker. [5, 162] From the CMB anisotropy, one infers $\Delta h(f \sim 10^{-18}\text{Hz}) \lesssim 10^{-6}$; from precision pulsar timing, $\Delta h(f \sim 10^{-8}\text{Hz}) \lesssim 10^{-14}$. Optimistic goals for the future laser-interferometer gravitational wave experiments LISA, advanced LIGO, and BBO are strain sensitivities of $\Delta h(f \sim 10^{-4}\text{Hz}) \sim 10^{-20.5}$,

$\Delta h(f \sim 10^2\text{Hz}) \sim 10^{-24}$, and $\Delta h(f \sim 1\text{ Hz}) \sim 10^{-26.5}$ [129], respectively. Fig. 3 shows results for values of $T_r$ and $V_{end}$ consistent with all constraints on the cyclic model [96].

Even if the parameters are chosen to saturate the BBN constraint, the spectrum is still orders of magnitude below the sensitivity of anticipated instruments. Hence, the detection of a scale-invariant, stochastic gravitational wave imprint in the CMB polarization would be consistent with inflation and rule out the cyclic model.

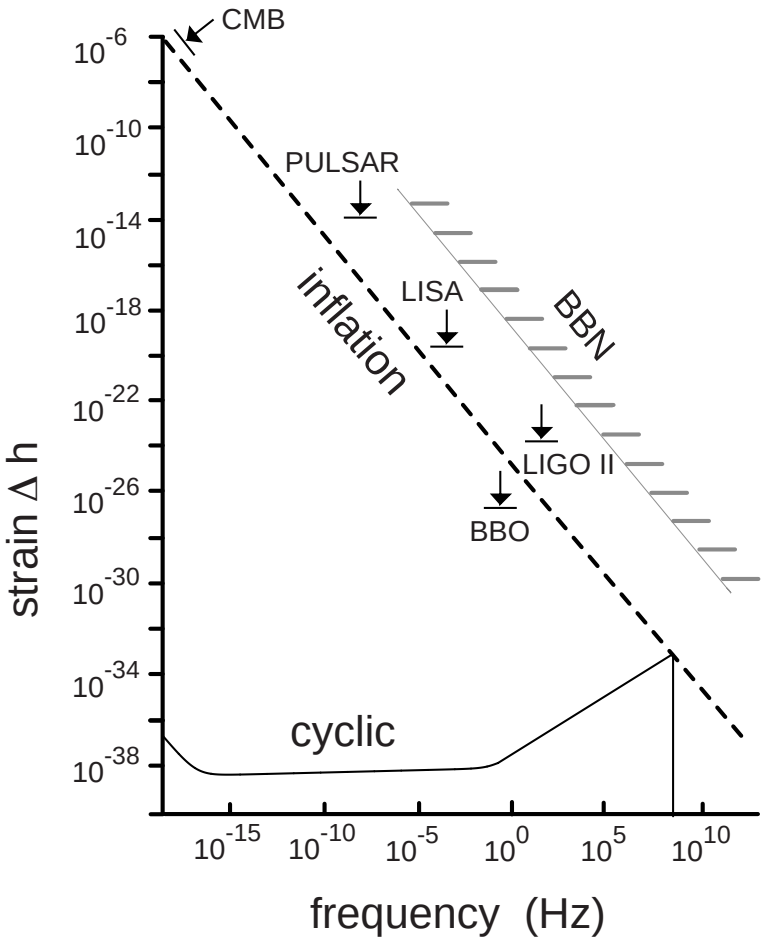



Figure 5.3: The present-day dimensionless strain, $\Delta h(k, \tau_0)$, predicted by the cyclic model with $T_r = 10^7$ GeV and $V_{end}^{1/4} = 10^{14}$ GeV. These parameters yield a gravity wave density four orders of magnitude below the BBN bound. Some observational bounds and (optimistic) future strain sensitivities are indicated.

# References


[1] L. F. Abbott and Mark B. Wise. Constraints on generalized inflationary cosmologies. *Nucl. Phys.*, B244:541–548, 1984.

[2] Fred C. Adams, J. Richard Bond, Katherine Freese, Joshua A. Frieman, and Angela V. Olinto. Natural inflation: Particle physics models, power law spectra for large scale structure, and constraints from cobe. *Phys. Rev.*, D47:426–455, 1993, hep-ph/9207245.

[3] Andreas Albrecht, Pedro Ferreira, Michael Joyce, and Tomislav Prokopec. Inflation and squeezed quantum states. *Phys. Rev.*, D50:4807–4820, 1994, astro-ph/9303001.

[4] Andreas Albrecht and Paul J. Steinhardt. Cosmology for grand unified theories with radiatively induced symmetry breaking. *Phys. Rev. Lett.*, 48:1220–1223, 1982.

[5] B. Allen. The stochastic gravity-wave background: Sources and detection. 1996, gr-qc/9604033.

[6] Bruce Allen, Eanna E. Flanagan, and Maria Alessandra Papa. Is the squeezing of relic gravitational waves produced by inflation detectable? *Phys. Rev.*, D61:024024, 2000, gr-qc/9906054.

[7] Nima Arkani-Hamed, Hsin-Chia Cheng, Paolo Creminelli, and Lisa Randall. Extranatural inflation. *Phys. Rev. Lett.*, 90:221302, 2003, hep-th/0301218.

[8] Nima Arkani-Hamed, Lawrence J. Hall, Christopher F. Kolda, and Hitoshi Murayama. A new perspective on cosmic coincidence problems. *Phys. Rev. Lett.*, 85:4434–4437, 2000, astro-ph/0005111.

[9] James M. Bardeen. Gauge invariant cosmological perturbations. *Phys. Rev.*, D22:1882–1905, 1980.

[10] James M. Bardeen, Paul J. Steinhardt, and Michael S. Turner. Spontaneous creation of almost scale - free density perturbations in an inflationary universe. *Phys. Rev.*, D28:679, 1983.

[11] Sergei Bashinsky. Coupled evolution of primordial gravity waves and relic neutrinos. 2005, astro-ph/0505502.

[12] T. J. Battefeld, S. P. Patil, and R. Brandenberger. Perturbations in a bouncing brane model. *Phys. Rev.*, D70:066006, 2004, hep-th/0401010.

[13] D. Baumann, A. Dymarsky, I. Klebanov, J. Maldacena, L. McAllister, and P. Steinhardt. Corrections to d3-brane potentials: Implications for brane inflation. (in preparation) 2006.

[14] D. Baumann, P. J. Steinhardt, and N. Turok. In preparation (2006).

[15] Daniel Baumann et al. On d3-brane potentials in compactifications with fluxes and wrapped d-branes. 2006, hep-th/0607050.

[16] C. L. Bennett et al. First year wilkinson microwave anisotropy probe (wmap) observations: Preliminary maps and basic results. *Astrophys. J. Suppl.*, 148:1, 2003, astro-ph/0302207.

[17] N. D. Birrell and P. C. W. Davies. Quantum fields in curved space. Cambridge, Uk: Univ. Pr. ( 1982) 340p.

[18] Tirthabir Biswas and Alessio Notari. Can inflation solve the hierarchy problem? 2005, hep-ph/0511207.

[19] James Bock et al. Task force on cosmic microwave background research. 2006, astro-ph/0604101.

[20] D. Boyanovsky, Hector J. de Vega, and Norma G. Sanchez. Clarifying inflation models: Slow-roll as an expansion in 1/n(efolds) and no fine tuning. *Phys. Rev.*, D73:023008, 2006, astro-ph/0507595.

[21] Latham A. Boyle. Expansion/contraction duality in curved frw models. 2004 (unpublished).

[22] Latham A. Boyle and Paul J. Steinhardt. Probing the early universe with inflationary gravitational waves. 2005, astro-ph/0512014.

[23] Latham A. Boyle, Paul J. Steinhardt, and Neil Turok. The cosmic gravitational wave background in a cyclic universe. *Phys. Rev.*, D69:127302, 2004, hep-th/0307170.

[24] Latham A. Boyle, Paul J. Steinhardt, and Neil Turok. A new duality relating density perturbations in expanding and contracting friedmann cosmologies. *Phys. Rev.*, D70:023504, 2004, hep-th/0403026.

[25] Latham A. Boyle, Paul J. Steinhardt, and Neil Turok. Inflationary predictions reconsidered. *Phys. Rev. Lett.*, 96:111301, 2006, astro-ph/0507455.

[26] Robert Brandenberger and Fabio Finelli. On the spectrum of fluctuations in an effective field theory of the ekpyrotic universe. *JHEP*, 11:056, 2001, hep-th/0109004.

[27] Helen A. Bridgman and David Wands. Cosmological perturbation spectra from sl(4,r)-invariant effective actions. *Phys. Rev.*, D61:123514, 2000, hep-th/0002215.

[28] R. Brustein, M. Gasperini, and G. Veneziano. Duality in cosmological perturbation theory. *Phys. Lett.*, B431:277–285, 1998, hep-th/9803018.

[29] Matthew R. Buckley and Hitoshi Murayama. How can we test seesaw experimentally? 2006, hep-ph/0606088.

[30] A. Buonanno, K. A. Meissner, C. Ungarelli, and G. Veneziano. Classical inhomogeneities in string cosmology. *Phys. Rev.*, D57:2543–2556, 1998, hep-th/9706221.

[31] A. Buonanno, K. A. Meissner, C. Ungarelli, and G. Veneziano. Quantum inhomogeneities in string cosmology. *JHEP*, 01:004, 1998, hep-th/9710188.

[32] Carlo Burigana et al. Cosmological reionization after wmap: perspectives from planck and future cmb missions. 2004, astro-ph/0411415.

[33] Cyril Cartier. Scalar perturbations in an alpha'-regularised cosmological bounce. 2004, hep-th/0401036.

[34] Cyril Cartier, Ruth Durrer, and Edmund J. Copeland. Cosmological perturbations and the transition from contraction to expansion. *Phys. Rev.*, D67:103517, 2003, hep-th/0301198.

[35] Sirichai Chongchitnan and George Efstathiou. Prospects for direct detection of primordial gravitational waves. *Phys. Rev.*, D73:083511, 2006, astro-ph/0602594.

[36] D. Cirigliano, Hector J. de Vega, and Norma G. Sanchez. Clarifying inflation models: The precise inflationary potential from effective field theory and the wmap data. *Phys. Rev.*, D71:103518, 2005, astro-ph/0412634.

[37] Edmund J. Copeland, Richard Easther, and David Wands. Vacuum fluctuations in axion-dilaton cosmologies. *Phys. Rev.*, D56:874–888, 1997, hep-th/9701082.

[38] Edmund J. Copeland, James E. Lidsey, and David Wands. S-duality-invariant perturbations in string cosmology. *Nucl. Phys.*, B506:407–420, 1997, hep-th/9705050.

[39] Vincent Corbin and Neil J. Cornish. Detecting the cosmic gravitational wave background with the big bang observer. *Class. Quant. Grav.*, 23:2435–2446, 2006, gr-qc/0512039.

[40] Marina Cortes and Andrew R Liddle. The consistency equation hierarchy in single-field inflation models. *Phys. Rev.*, D73:083523, 2006, astro-ph/0603016.

[41] Ben Craps and Burt A. Ovrut. Global fluctuation spectra in big crunch / big bang string vacua. *Phys. Rev.*, D69:066001, 2004, hep-th/0308057.

[42] Jeff Crowder and Neil J. Cornish. Beyond lisa: Exploring future gravitational wave missions. *Phys. Rev.*, D72:083005, 2005, gr-qc/0506015.

[43] Hooman Davoudiasl, Ryuichiro Kitano, Graham D. Kribs, Hitoshi Murayama, and Paul J. Steinhardt. Gravitational baryogenesis. *Phys. Rev. Lett.*, 93:201301, 2004, hep-ph/0403019.

[44] W. de Sitter. Einstein's theory of gravitation and its astronomical consequences. Third paper. *Mon. Not. Roy. Astron. Soc.*, 78:3–28, November 1917.

[45] Fabrizio Di Marco and Alessio Notari. 'graceful' old inflation. *Phys. Rev.*, D73:063514, 2006, astro-ph/0511396.

[46] Duane A. Dicus and Wayne W. Repko. Comment on 'damping of tensor modes in cosmology'. *Phys. Rev.*, D72:088302, 2005, astro-ph/0509096.

[47] S. Dimopoulos, S. Kachru, J. McGreevy, and Jay G. Wacker. N-flation. 2005, hep-th/0507205.

[48] S. Dimopoulos, S. Raby, and Frank Wilczek. Supersymmetry and the scale of unification. *Phys. Rev.*, D24:1681–1683, 1981.

[49] A. G. Doroshkevich, I. D. Novikov, and A. G. Polnarev. Temperature fluctuations in the primordial background radiation due to gravitational waves. *Soviet Astronomy*, 21:529–535, October 1977.

[50] Ruth Durrer and Filippo Vernizzi. Adiabatic perturbations in pre big bang models: Matching conditions and scale invariance. *Phys. Rev.*, D66:083503, 2002, hep-ph/0203275.

[51] Anatoly Dymarsky, Igor R. Klebanov, and Nathan Seiberg. On the moduli space of the cascading su(m+p) x su(p) gauge theory. *JHEP*, 01:155, 2006, hep-th/0511254.

[52] Richard Easther and Eugene A. Lim. Stochastic gravitational wave production after inflation. *JCAP*, 0604:010, 2006, astro-ph/0601617.

[53] Richard Easther and Liam McAllister. Random matrices and the spectrum of n-flation. *JCAP*, 0605:018, 2006, hep-th/0512102.

[54] George Efstathiou and Sirichai Chongchitnan. The search for primordial tensor modes. 2006, astro-ph/0603118.

[55] George Efstathiou and Katherine J. Mack. The lyth bound revisited. *JCAP*, 0505:008, 2005, astro-ph/0503360.

[56] Albert Einstein. The field equations of gravitation. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys. )*, 1915:844–847, 1915.

[57] Albert Einstein. Cosmological considerations in the general theory of relativity. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys. )*, 1917:142–152, 1917.

[58] Joel K. Erickson, Steven Gratton, Paul J. Steinhardt, and Neil Turok. Cosmic perturbations through the cyclic ages. 2006, hep-th/0607164.

[59] Joel K. Erickson, Daniel H. Wesley, Paul J. Steinhardt, and Neil Turok. Kasner and mixmaster behavior in universes with equation of state w ¿= 1. *Phys. Rev.*, D69:063514, 2004, hep-th/0312009.

[60] R. Fabbri and M. d. Pollock. The effect of primordially produced gravitons upon the anisotropy of the cosmological microwave background radiation. *Phys. Lett.*, B125:445–448, 1983.

[61] Katherine Freese, Joshua A. Frieman, and Angela V. Olinto. Natural inflation with pseudo - nambu-goldstone bosons. *Phys. Rev. Lett.*, 65:3233–3236, 1990.

[62] A. Friedman. On the curvature of space. *Z. Phys.*, 10:377–386, 1922.

[63] A. Friedmann. On the possibility of a world with constant negative curvature of space. *Z. Phys.*, 21:326–332, 1924.

[64] M. Fukugita and T. Yanagida. Baryogenesis without grand unification. *Phys. Lett.*, B174:45, 1986.

[65] M. Gasperini, Massimo Giovannini, and G. Veneziano. Cosmological perturbations across a curvature bounce. *Nucl. Phys.*, B694:206–238, 2004, hep-th/0401112.

[66] Murray Gell-Mann, Pierre Ramond, and Richard Slansky. Complex spinors and unified theories. Print-80-0576 (CERN).

[67] H. Georgi, Helen R. Quinn, and Steven Weinberg. Hierarchy of interactions in unified gauge theories. *Phys. Rev. Lett.*, 33:451–454, 1974.

[68] S. L. Glashow. The future of elementary particle physics. *NATO Adv. Study Inst. Ser. B Phys.*, 59:687, 1979.

[69] Steven Gratton, Justin Khoury, Paul J. Steinhardt, and Neil Turok. Conditions for generating scale-invariant density perturbations. *Phys. Rev.*, D69:103505, 2004, astro-ph/0301395.

[70] L. P. Grishchuk. Amplification of gravitational waves in an istropic universe. *Sov. Phys. JETP*, 40:409–415, 1975.

[71] L. P. Grishchuk and Yu. V. Sidorov. On the quantum state of relic gravitons. *Class. Quant. Grav.*, 6:L161–L165, 1989.

[72] L. P. Grishchuk and Yu. V. Sidorov. Squeezed quantum states of relic gravitons and primordial density fluctuations. *Phys. Rev.*, D42:3413–3421, 1990.

[73] A. H. Guth and S. Y. Pi. Fluctuations in the new inflationary universe. *Phys. Rev. Lett.*, 49:1110–1113, 1982.

[74] Alan H. Guth. The inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev.*, D23:347–356, 1981.

[75] S. W. Hawking. The development of irregularities in a single bubble inflationary universe. *Phys. Lett.*, B115:295, 1982.

[76] Mark B. Hoffman and Michael S. Turner. Kinematic constraints to the key inflationary observables. *Phys. Rev.*, D64:023506, 2001, astro-ph/0006321.

[77] Gary T. Horowitz and Joseph Polchinski. Instability of spacelike and null orbifold singularities. *Phys. Rev.*, D66:103512, 2002, hep-th/0206228.

[78] E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173, March 1929.

[79] Kevin M. Huffenberger, H. K. Eriksen, and F. K. Hansen. Point source power in three-year wilkinson microwave anisotropy probe data. 2006, astro-ph/0606538.

[80] Shamit Kachru et al. Towards inflation in string theory. *JCAP*, 0310:013, 2003, hep-th/0308055.

[81] K. Kajantie, M. Laine, K. Rummukainen, and Y. Schroder. The pressure of hot qcd up to g**6 ln(1/g). *Phys. Rev.*, D67:105008, 2003, hep-ph/0211321.

[82] Renata Kallosh, Lev Kofman, and Andrei D. Linde. Pyrotechnic universe. *Phys. Rev.*, D64:123523, 2001, hep-th/0104073.

[83] Renata Kallosh, Lev Kofman, Andrei D. Linde, and Arkady A. Tseytlin. Bps branes in cosmology. *Phys. Rev.*, D64:123524, 2001, hep-th/0106241.

[84] Marc Kamionkowski, Arthur Kosowsky, and Albert Stebbins. A probe of primordial gravity waves and vorticity. *Phys. Rev. Lett.*, 78:2058–2061, 1997, astro-ph/9609132.

[85] Marc Kamionkowski, Arthur Kosowsky, and Albert Stebbins. Statistics of cosmic microwave background polarization. *Phys. Rev.*, D55:7368–7388, 1997, astro-ph/9611125.

[86] Manoj Kaplinghat, Lloyd Knox, and Yong-Seon Song. Determining neutrino mass from the cmb alone. *Phys. Rev. Lett.*, 91:241301, 2003, astro-ph/0303344.

[87] S. Kawamura et al. The japanese space gravitational wave antenna decigo. *Class. Quant. Grav.*, 23:S125–S132, 2006.

[88] M. Kawasaki, Masahide Yamaguchi, and T. Yanagida. Natural chaotic inflation in supergravity. *Phys. Rev. Lett.*, 85:3572–3575, 2000, hep-ph/0004243.

[89] M. Kawasaki, Masahide Yamaguchi, and T. Yanagida. Natural chaotic inflation in supergravity and leptogenesis. *Phys. Rev.*, D63:103514, 2001, hep-ph/0011104.

[90] S. Y. Khlebnikov and I. I. Tkachev. Relic gravitational waves produced after preheating. *Phys. Rev.*, D56:653–660, 1997, hep-ph/9701423.

[91] Justin Khoury, Burt A. Ovrut, Nathan Seiberg, Paul J. Steinhardt, and Neil Turok. From big crunch to big bang. *Phys. Rev.*, D65:086007, 2002, hep-th/0108187.

[92] Justin Khoury, Burt A. Ovrut, Paul J. Steinhardt, and Neil Turok. A brief comment on 'the pyrotechnic universe'. 2001, hep-th/0105212.

[93] Justin Khoury, Burt A. Ovrut, Paul J. Steinhardt, and Neil Turok. The ekpyrotic universe: Colliding branes and the origin of the hot big bang. *Phys. Rev.*, D64:123522, 2001, hep-th/0103239.

[94] Justin Khoury, Burt A. Ovrut, Paul J. Steinhardt, and Neil Turok. Density perturbations in the ekpyrotic scenario. *Phys. Rev.*, D66:046005, 2002, hep-th/0109050.

[95] Justin Khoury, Paul J. Steinhardt, and Neil Turok. Great expectations: Inflation versus cyclic predictions for spectral tilt. *Phys. Rev. Lett.*, 91:161301, 2003, astro-ph/0302012.

[96] Justin Khoury, Paul J. Steinhardt, and Neil Turok. Designing cyclic universe models. *Phys. Rev. Lett.*, 92:031302, 2004, hep-th/0307132.

[97] Jihn E. Kim, Hans Peter Nilles, and Marco Peloso. Completing natural inflation. *JCAP*, 0501:005, 2005, hep-ph/0409138.

[98] William H. Kinney. Inflation: Flow, fixed points and observables to arbitrary order in slow roll. *Phys. Rev.*, D66:083508, 2002, astro-ph/0206032.

[99] Hideo Kodama and Misao Sasaki. Cosmological perturbation theory. *Prog. Theor. Phys. Suppl.*, 78:1–166, 1984.

[100] E. W. Kolb and Michael S. Turner. The early universe. *Front. Phys.*, 69:1–547, 1990.

[101] Hideaki Kudoh, Atsushi Taruya, Takashi Hiramatsu, and Yoshiaki Himemoto. Detecting a gravitational-wave background with next- generation space interferometers. *Phys. Rev.*, D73:064006, 2006, gr-qc/0511145.

[102] G. Lemaitre. The universe in expansion. *Annales Soc. Sci. Brux. Ser. I Sci. Math. Astron. Phys.*, A53:51–85, 1933.

[103] G. Lemaitre. The expanding universe. *Gen. Rel. Grav.*, 29:641–680, 1997.

[104] Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of cmb anisotropies in closed frw models. *Astrophys. J.*, 538:473–476, 2000, astro-ph/9911177.

[105] Andrew R Liddle, David Parkinson, Samuel M Leach, and Pia Mukherjee. The wmap normalization of inflationary cosmologies. 2006, astro-ph/0607275.

[106] James E. Lidsey, David Wands, and Edmund J. Copeland. Superstring cosmology. *Phys. Rept.*, 337:343–492, 2000, hep-th/9909061.

[107] Andrei D. Linde. A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Phys. Lett.*, B108:389–393, 1982.

[108] Andrei D. Linde. Chaotic inflation. *Phys. Lett.*, B129:177–181, 1983.

[109] Andrei D. Linde. Hybrid inflation. *Phys. Rev.*, D49:748–754, 1994, astro-ph/9307002.

[110] Hong Liu, Gregory W. Moore, and Nathan Seiberg. Strings in a time-dependent orbifold. *JHEP*, 06:045, 2002, hep-th/0204168.

[111] Hong Liu, Gregory W. Moore, and Nathan Seiberg. Strings in time-dependent orbifolds. *JHEP*, 10:031, 2002, hep-th/0206182.

[112] David H. Lyth. What would we learn by detecting a gravitational wave signal in the cosmic microwave background anisotropy? *Phys. Rev. Lett.*, 78:1861–1863, 1997, hep-ph/9606387.

[113] David H. Lyth. The primordial curvature perturbation in the ekpyrotic universe. *Phys. Lett.*, B524:1–4, 2002, hep-ph/0106153.

[114] Juan M. Maldacena. Non-gaussian features of primordial fluctuations in single field inflationary models. *JHEP*, 05:013, 2003, astro-ph/0210603.

[115] Jerome Martin and Patrick Peter. On the 'causality argument' in bouncing cosmologies. *Phys. Rev. Lett.*, 92:061301, 2004, astro-ph/0312488.

[116] Jerome Martin, Patrick Peter, Nelson Pinto Neto, and Dominik J. Schwarz. Passing through the bounce in the ekpyrotic models. *Phys. Rev.*, D65:123513, 2002, hep-th/0112128.

[117] Jerome Martin, Patrick Peter, Nelson Pinto-Neto, and Dominik J. Schwarz. Comment on 'density perturbations in the ekpyrotic scenario'. *Phys. Rev.*, D67:028301, 2003, hep-th/0204222.

[118] Peter Minkowski. mu $\rightarrow$ e gamma at a rate of one out of 1-billion muon decays? *Phys. Lett.*, B67:421, 1977.

[119] C. W. Misner, K. S. Thorne, and J. A. Wheeler. Gravitation. San Francisco 1973, 1279p.

[120] Viatcheslav F. Mukhanov. Gravitational instability of the universe filled with a scalar field. *JETP Lett.*, 41:493–496, 1985.

[121] Viatcheslav F. Mukhanov. Quantum theory of gauge invariant cosmological perturbations. *Sov. Phys. JETP*, 67:1297–1302, 1988.

[122] Viatcheslav F. Mukhanov and G. V. Chibisov. Quantum fluctuation and 'nonsingular' universe. (in russian). *JETP Lett.*, 33:532–535, 1981.

[123] Viatcheslav F. Mukhanov, H. A. Feldman, and Robert H. Brandenberger. Theory of cosmological perturbations. part 1. classical perturbations. part 2. quantum theory of perturbations. part 3. extensions. *Phys. Rept.*, 215:203–333, 1992.

[124] H. Murayama, Hiroshi Suzuki, T. Yanagida, and Jun'ichi Yokoyama. Chaotic inflation and baryogenesis by right-handed sneutrinos. *Phys. Rev. Lett.*, 70:1912–1915, 1993.

[125] Kin-wang Ng and A. D. Speliotopoulos. Cosmological evolution of scale invariant gravity waves. *Phys. Rev.*, D52:2112–2117, 1995, astro-ph/9405043.

[126] Kin Wang Ng and A. D. Speliotopoulos. Random and correlated phases of primordial gravitational waves. *Phys. Rev.*, D51:5636–5642, 1995, hep-ph/9407302.

[127] David Parkinson, Pia Mukherjee, and Andrew R Liddle. A bayesian model selection analysis of wmap3. *Phys. Rev.*, D73:123523, 2006, astro-ph/0605003.

[128] H. V. Peiris et al. First year wilkinson microwave anisotropy probe (wmap) observations: Implications for inflation. *Astrophys. J. Suppl.*, 148:213, 2003, astro-ph/0302225.

[129] E. S. Phinney et al. The big bang observer, nasa mission concept study. *URL universe.nasa.gov/program/bbo.html*, 2003.

[130] Yun-Song Piao and Yuan-Zhong Zhang. The primordial perturbation spectrum from various expanding and contracting phases. *Phys. Rev.*, D70:043516, 2004, astro-ph/0403671.

[131] Dmitry I. Podolsky, Gary N. Felder, Lev Kofman, and Marco Peloso. Equation of state and beginning of thermalization after preheating. *Phys. Rev.*, D73:023501, 2006, hep-ph/0507096.

[132] David Polarski and Alexei A. Starobinsky. Semiclassicality and decoherence of cosmological perturbations. *Class. Quant. Grav.*, 13:377–392, 1996, gr-qc/9504030.

[133] A. G. Polnarev. Polarization and Anisotropy Induced in the Microwave Background by Cosmological Gravitational Waves. *Soviet Astronomy*, 29:607–+, December 1985.

[134] Jonathan R. Pritchard and Marc Kamionkowski. Cosmic microwave background fluctuations from gravitational waves: An analytic approach. *Annals Phys.*, 318:2–36, 2005, astro-ph/0412581.

[135] H. P. Robertson. On the Foundations of Relativistic Cosmology. *Proceedings of the National Academy of Science*, 15:822–829, November 1929.

[136] H. P. Robertson. Relativistic Cosmology. *Reviews of Modern Physics*, 5:62–90, January 1933.

[137] V. A. Rubakov, M. V. Sazhin, and A. V. Veryaskin. Graviton creation in the inflationary universe and the grand unification scale. *Phys. Lett.*, B115:189–192, 1982.

[138] R. K. Sachs and A. M. Wolfe. Perturbations of a cosmological model and angular variations of the microwave background. *Astrophys. J.*, 147:73–90, 1967.

[139] Varun Sahni. The energy density of relic gravity waves from inflation. *Phys. Rev.*, D42:453–463, 1990.

[140] Uros Seljak et al. Cosmological parameter analysis including sdss ly-alpha forest and galaxy bias: Constraints on the primordial spectrum of fluctuations, neutrino mass, and dark energy. *Phys. Rev.*, D71:103515, 2005, astro-ph/0407372.

[141] Uros Seljak, Anze Slosar, and Patrick McDonald. Cosmological parameters from combining the lyman-alpha forest with cmb, galaxy clustering and sn constraints. 2006, astro-ph/0604335.

[142] Uros Seljak and Matias Zaldarriaga. A line of sight approach to cosmic microwave background anisotropies. *Astrophys. J.*, 469:437–444, 1996, astro-ph/9603033.

[143] Uros Seljak and Matias Zaldarriaga. Signature of gravity waves in polarization of the microwave background. *Phys. Rev. Lett.*, 78:2054–2057, 1997, astro-ph/9609169.

[144] Ashoke Sen. O(d) x o(d) symmetry of the space of cosmological solutions in string theory, scale factor duality and two-dimensional black holes. *Phys. Lett.*, B271:295–300, 1991.

[145] Naoki Seto. Correlation analysis of stochastic gravitational wave background around 0.1-hz - 1-hz. *Phys. Rev.*, D73:063001, 2006, gr-qc/0510067.

[146] Naoki Seto, Seiji Kawamura, and Takashi Nakamura. Possibility of direct measurement of the acceleration of the universe using 0.1-hz band laser interferometer gravitational wave antenna in space. *Phys. Rev. Lett.*, 87:221103, 2001, astro-ph/0108011.

[147] Naoki Seto and Jun'Ichi Yokoyama. Probing the equation of state of the early universe with a space laser interferometer. *J. Phys. Soc. Jap.*, 72:3082–3086, 2003, gr-qc/0305096.

[148] Tristan L. Smith, Marc Kamionkowski, and Asantha Cooray. Direct detection of the inflationary gravitational wave background. *Phys. Rev.*, D73:023504, 2006, astro-ph/0506422.

[149] Tristan L. Smith, Hiranya V. Peiris, and Asantha Cooray. Deciphering inflation with gravitational waves: Cosmic microwave background polarization vs. direct detection with laser interferometers. *Phys. Rev.*, D73:123503, 2006, astro-ph/0602137.

[150] Yong-Seon Song and Lloyd Knox. The detectability of departures from the inflationary consistency equation. *Phys. Rev.*, D68:043518, 2003, astro-ph/0305411.

[151] Yong-Seon Song and Lloyd Knox. Determination of cosmological parameters from cosmic shear data. *Phys. Rev.*, D70:063510, 2004, astro-ph/0312175.

[152] D. N. Spergel et al. First year wilkinson microwave anisotropy probe (wmap) observations: Determination of cosmological parameters. *Astrophys. J. Suppl.*, 148:175, 2003, astro-ph/0302209.

[153] D. N. Spergel et al. Wilkinson microwave anisotropy probe (wmap) three year results: Implications for cosmology. 2006, astro-ph/0603449.

[154] A. A. Starobinsky. Cosmic background anisotropy induced by isotropic flat- spectrum gravitational-wave perturbations. *Sov. Astron. Lett.*, 11:133, 1985.

[155] Alexei A. Starobinsky. Spectrum of relict gravitational radiation and the early state of the universe. *JETP Lett.*, 30:682–685, 1979.

[156] Alexei A. Starobinsky. Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations. *Phys. Lett.*, B117:175–178, 1982.

[157] P. J. Steinhardt. Cosmological perturbations: Myths and facts. *Mod. Phys. Lett.*, A19:967–982, 2004.

[158] P. J. Steinhardt and N. Turok. A cyclic model of the universe. *Science*, 296:1436–1439, 2002.

[159] Paul J. Steinhardt and Neil Turok. Cosmic evolution in a cyclic universe. *Phys. Rev.*, D65:126003, 2002, hep-th/0111098.

[160] Paul J. Steinhardt and Neil Turok. The cyclic universe: An informal introduction. *Nucl. Phys. Proc. Suppl.*, 124:38–49, 2003, astro-ph/0204479.

[161] K. S. Thorne. Gravitational radiation. In *Hawking, S.W. (ed.), Israel, W. (ed.): Three hundred years of gravitation*, 330-458. (see Book Index).

[162] Kip S. Thorne. Gravitational waves. 1995, gr-qc/9506086.

[163] Andrew J. Tolley and Neil Turok. Quantum fields in a big crunch / big bang spacetime. *Phys. Rev.*, D66:106005, 2002, hep-th/0204091.

[164] Andrew J. Tolley, Neil Turok, and Paul J. Steinhardt. Cosmological perturbations in a big crunch / big bang space-time. *Phys. Rev.*, D69:106005, 2004, hep-th/0306109.

[165] R. C. Tolman. *Relativity, Thermodynamics, and Cosmology.* Relativity, Thermodynamics, and Cosmology, Oxford: Clarendon Press, 1934, 1934.

[166] Roberto Trotta. Applications of bayesian model selection to cosmological parameters. 2005, astro-ph/0504022.

[167] A. A. Tseytlin. Duality and dilaton. *Mod. Phys. Lett.*, A6:1721–1732, 1991.

[168] Michael S. Turner. Detectability of inflation-produced gravitational waves. *Phys. Rev.*, D55:435–439, 1997, astro-ph/9607066.

[169] Michael S. Turner, Martin J. White, and James E. Lidsey. Tensor perturbations in inflationary models as a probe of cosmology. *Phys. Rev.*, D48:4613–4622, 1993, astro-ph/9306029.

[170] Neil Turok, Malcolm Perry, and Paul J. Steinhardt. M theory model of a big crunch / big bang transition. *Phys. Rev.*, D70:106004, 2004, hep-th/0408083.

[171] Carlo Ungarelli, Pierstefano Corasaniti, R. A. Mercer, and Alberto Vecchio. Gravitational waves, inflation and the cosmic microwave background: Towards testing the slow-roll paradigm. *Class. Quant. Grav.*, 22:S955–S964, 2005, astro-ph/0504294.

[172] G. Veneziano. Scale factor duality for classical and quantum strings. *Phys. Lett.*, B265:287–294, 1991.

[173] Licia Verde, Hiranya Peiris, and Raul Jimenez. Optimizing cmb polarization experiments to constrain inflationary physics. *JCAP*, 0601:019, 2006, astro-ph/0506036.

[174] A. G. Walker. Distance in an expanding universe. *Mon. Not. Roy. Astron. Soc.*, 94:159–+, December 1933.

[175] David Wands. Duality invariance of cosmological perturbation spectra. *Phys. Rev.*, D60:023507, 1999, gr-qc/9809062.

[176] S. Weinberg. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity.* Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity, by Steven Weinberg, pp. 688. ISBN 0-471-92567-5. Wiley-VCH , July 1972., July 1972.

[177] Steven Weinberg. The first three minutes. a modern view of the origin of the universe. (german transl.). Muenchen 1977, 269p.

[178] Steven Weinberg. Cosmological constraints on the scale of supersymmetry breaking. *Phys. Rev. Lett.*, 48:1303, 1982.

[179] Steven Weinberg. Damping of tensor modes in cosmology. *Phys. Rev.*, D69:023503, 2004, astro-ph/0306304.

[180] Steven Weinberg. Quantum contributions to cosmological correlations. *Phys. Rev.*, D72:043514, 2005, hep-th/0506236.

[181] Steven Weinberg. Cosmology textbook (in preparation). 2006.

[182] Steven Weinberg. Quantum contributions to cosmological correlations. ii: Can these corrections become large? 2006, hep-th/0605244.

[183] Daniel H. Wesley, Paul J. Steinhardt, and Neil Turok. Controlling chaos through compactification in cosmological models with a collapsing phase. *Phys. Rev.*, D72:063513, 2005, hep-th/0502108.

[184] Edward Witten. Quest for unification. 2002, hep-ph/0207124.

[185] Tsutomu Yanagida. Horizontal gauge symmetry and masses of neutrinos. In Proceedings of the Workshop on the Baryon Number of the Universe and Unified Theories, Tsukuba, Japan, 13-14 Feb 1979.

[186] Matias Zaldarriaga and Uros Seljak. An all-sky analysis of polarization in the microwave background. *Phys. Rev.*, D55:1830–1840, 1997, astro-ph/9609170.

[187] Matias Zaldarriaga and Uros Seljak. Cmbfast for spatially closed universes. *Astrophys. J. Suppl.*, 129:431–434, 2000, astro-ph/9911219.