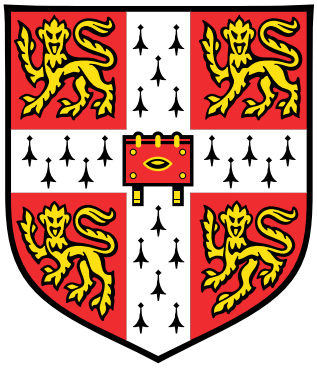
HIGH-ENERGY ASPECTS OF INFLATIONARY COSMOLOGY

HAYDEN LEE

Corpus Christi College
University of Cambridge



Thesis submitted for the degree of
Doctor of Philosophy

May 2017

Abstract

Since the discovery of the cosmic microwave background (CMB), our understanding of the cosmos has been rapidly evolving. Detailed measurements of the CMB temperature fluctuations have led to a standard cosmological model, which traces the origin of the large-scale structure of the universe to quantum fluctuations during inflation. Although the basic framework of inflationary cosmology is now well-established, the microphysical mechanism responsible for the accelerated expansion remains a mystery. In this thesis, we describe how the physics underlying inflation can be probed using two cosmological observables: higher-order correlations of primordial density perturbations (non-Gaussianity) and primordial gravitational waves (tensor modes).

In the first part of the thesis, we explore novel signatures of high-energy physics in higher-order correlation functions of inflationary perturbations. First, we use causality and unitarity to make connections between cosmological observations and the underlying short-distance dynamics of single-field inflation. We obtain a constraint on the size and the sign of the four-point function in terms of the amplitude of the three-point function. We then study the imprints of extra massive particles of arbitrary spin on the three-point function. We classify the couplings of these particles to inflationary scalar and tensor perturbations and derive explicit shape functions for their three-point functions that can serve as templates for future observational searches. Establishing the particle content during inflation would provide important hints for the microscopic theory of inflation.

In the second part, we study ways of testing the nature of inflation using inflationary tensor modes. We consider effects of gravitational corrections to Einstein gravity in models of high-scale inflation. We show that these scenarios can lead to a violation of the tensor consistency condition (i.e. the relation between the amplitude and the scale-dependence of the tensor two-point function) that is satisfied by canonical single-field inflationary models. Finally, we consider the prospects for measuring the inflationary superhorizon signature in future observations. We define an estimator that captures superhorizon correlations and present forecasts for the detectability of the signal with future CMB polarization experiments.

Acknowledgements

I owe my deepest gratitude to my supervisor Daniel Baumann for his exceptional guidance and support throughout my PhD. His infectious passion for physics has been a constant source of inspiration since the day I attended his first lecture in my first semester at Cambridge. It has been a true privilege to have such a dedicated supervisor like Daniel, without whom this thesis would not have been possible.

Beside my supervisor, I am grateful to have had the opportunity to work with my other collaborators, Daniel Green, Guilherme Pimentel, Rafael Porto, and Simon Su. I wish to especially thank Guilherme for his contribution to this thesis and for the numerous enjoyable conversations we had about a range of topics in physics, from which I have learned a lot.

Over the course of my PhD, I had the pleasure of having enlightening discussions with many excellent physicists. For this I wish to thank: Peter Adshead, Mustafa Amin, Matteo Biagetti, Paolo Creminelli, Anne Davis, James Fergusson, Ben Freivogel, Garrett Goon, Eugene Lim, David Marsh, Daan Meerburg, Enrico Pajer, Paul Shellard, Jan Pieter van der Schaar, and Yi Wang.

I want to give a special thanks to my departmental colleagues, Valentin Assassi, Horng Sheng Chia, Helge Gruetjen, Laurence Perreault-Levasseur, Simon Su, and Benjamin Wallisch, who made my PhD journey far more enjoyable.

Above all, I would like to express my heartfelt gratitude to my family for their unconditional love and support. Without their sacrifice and understanding, I would never have been able to pursue my passion of becoming a physicist, and I dedicate this thesis to them.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except specifically indicated in the text. Nevertheless, following the tendency of modern research in theoretical physics, most of the material discussed in this dissertation is the result of collaborative research. In particular, Chapter 4 is based on work done in collaboration with Daniel Baumann, Daniel Green, and Rafael Porto, published in [1], Chapters 5 and 6 are the results of the work done in collaboration with Daniel Baumann and Guilherme Pimentel, published in [2, 3], and Chapter 7 is based on the work done in collaboration with Daniel Baumann and Simon Su, published in [4]. I have made major contributions to the above, both in terms of results and writing.

I hereby declare that my thesis entitled *"High-Energy Aspects of Inflationary Cosmology"* is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

Contents

1. Introduction

Cosmology is the study of the birth, evolution, and fate of the universe. This intriguing subject has been pondered upon by humankind since ancient times and become a proper branch of science thanks to intellectual giants such as Galileo, Kepler, and Newton. Yet, it was not until the last century that we have witnessed paradigm-shifting revolutions in the field of cosmology and formed a realistic picture of the cosmos.

The first of these revolutions was Einstein's general theory of relativity [5]. This extended Newtonian gravity and provided a mathematical framework that can be used to describe the geometry of spacetime and the evolution of the universe. Major contributions to cosmology were made by Friedmann [6] and Lemaître [7] in the subsequent period, who found an exact solution to Einstein's field equations that described an expanding universe. Although the idea that the universe is expanding was considered rather preposterous at the time, Hubble soon confirmed this through the discovery of the relationship between distances to galaxies and their redshifts [8]. This gave birth to the Big Bang theory of the universe.

The second revolution builds on the development of the hot Big Bang theory by Gamow and collaborators, including their pioneering work on the Big Bang nucleosynthesis (BBN)—a theory regarding the production of light elements in the universe. One of their most important predictions was the existence of a relic background cosmic radiation left over from the hot phase of the universe [9]. This was soon confirmed by the landmark discovery of the cosmic microwave background (CMB) by Penzias and Wilson in 1964 [10]. Together with the observed abundances of light elements that matched the prediction from the BBN, the Big Bang theory of the universe was firmly established.

The next revolution was the advent of the era of precision cosmology. Our understanding of the universe has been greatly advanced in the last several decades, both theoretically and observationally, thanks to the series of high-precision satellite missions. In 1992, the Cosmic Background Explorer (COBE) measured the near-perfect blackbody spectrum of the CMB with tiny anisotropies [11]; the Wilkinson Microwave Anisotropy Probe (WMAP) provided a significantly improved measurement of the CMB anisotropies in 2003 [12]; the data from the Planck satellite in 2013 further refined the measurements of the CMB anisotropies [13] (see Fig. 1.1), providing state-of-the-art constraints on cosmological parameters. On another observational front, direct evidence for an accelerated expansion of the universe today came from observations of type Ia

supernovae by the High-Z Supernova Search Team [14] and the Supernova Cosmology Project [15]. All of these observations have led to significant developments of the Big Bang theory, culminating in the establishment of the standard model of cosmology, the *ΛCDM model*.

The ΛCDM model is the simplest parameterization that accounts for all current cosmological observations. This is based on general relativity and contains two extra ingredients beyond the Standard Model (SM) of particle physics: a *cosmological constant* Λ, which is a constant energy density associated with *dark energy* driving the current accelerated expansion of the universe, and non-baryonic *cold dark matter* (CDM), which provides extra mass density that accounts for gravitational lensing and various other astrophysical phenomena. According to this model, the universe today consists of 68% dark energy, 25% dark matter, 5% baryonic matter, and 0.1% radiation, with the observable size of the universe being ten billion light-years.

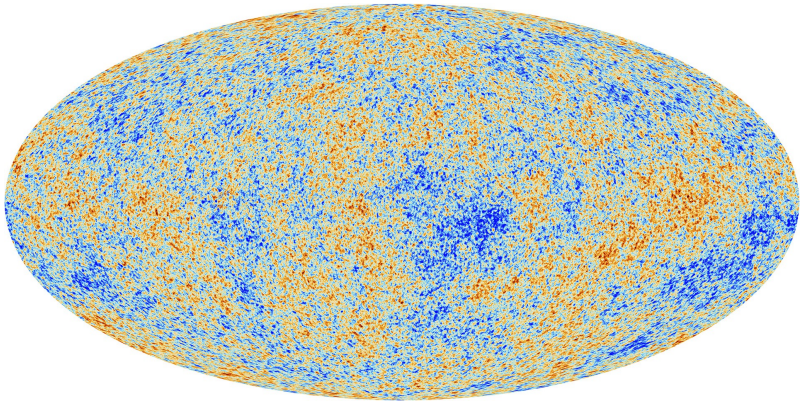


Figure 1.1: The anisotropies of the CMB as observed by the Planck satellite.[1] The variations in the CMB temperature across the sky reflect the density perturbations at recombination.

Our journey to understand the cosmos is not complete, however. Despite the spectacular empirical success of the ΛCDM model, it also comes with a number of shortcomings, including the apparent fine-tuning of the initial conditions of the universe. The extreme isotropy of the observed CMB implies that different regions of the universe should have been in causal contact in the early universe. However, this turns out to be strictly forbidden between regions that are separated by more than 2 degrees on the sky, if the universe was dominated by ordinary matter and radiation before the CMB was generated. This points towards a very special initial condition of the universe that was homogeneous and isotropic to an extreme degree.

The most popular solution to the puzzles associated with the standard Big Bang cosmology is inflation [16]. The inflationary paradigm posits a phase of exponential expansion in the early universe driven by a hypothetical scalar field called the inflaton. During inflation, an initial causal patch quickly leaves the cosmological horizon due to the exponential expansion, which explains how apparently causally disconnected regions that come into view in the matter- and radiation-dominated universe could have once communicated with each other.

[1] http://www.esa.int/spaceinimages/Images/2013/03/Planck_CMB

The status of primordial cosmology today can be compared with the pre-LHC era in elementary particle physics. Although the SM of particle physics was formulated many decades ago, the mechanism for electroweak symmetry breaking remained unproven until the eventual discovery of the Higgs boson in 2013 [17]. Similarly, the basic framework of inflationary cosmology is now well-established, but the precise microscopic mechanism that was responsible for inflation is still elusive. A plentitude of inflationary models exist on the market, yet the current observational constraints are not strong enough to distinguish between many of them. Upcoming cosmological experiments will provide critical observational tests of the inflationary paradigm and help us analyze inflation in a more quantitative manner.

Inflation describes the earliest stage in the history of the universe, governed by physics at ultra-high energy scales. Understanding the inflationary universe thus involves an interplay of the pillars of modern physics—general relativity and quantum physics—as well as ideas from modern high-energy physics. As with other branches of physics, this can be pursued in two complementary ways. In the *top-down* approach, one constructs explicit models of inflation from the underlying physics at high energies. In particular, string theory has led to new insights and approaches to inflationary model building. In the *bottom-up* approach, one instead starts from the most general phenomenological model of inflation valid at low energies. This is useful in characterizing all possible experimental signatures in a systematic manner.

A key challenge in the bottom-up approach to inflation is to extract information about the fundamental physics that governed inflation at high energies with access only to observables in the low-energy theory. In this thesis, we will follow the bottom-up approach and explore novel signatures of new physics during inflation using two important cosmological observables: higher-order correlation functions of density perturbations (*non-Gaussianity*) and primordial gravitational waves (*tensor modes*). We will search in these observables for clues about the physics that drove the inflationary expansion.

1.1 Outline of the thesis

We start by providing an overview of the theoretical basis. In Chapter 2, we review the inflationary paradigm. We first give a brief account of the basic aspects of Big Bang cosmology and present its classic conundrums motivating the need for inflation. We then provide a short review on inflationary cosmology and its observational status. Chapter 3 focuses on effective theories of inflation. We first present the well-known ultraviolet (UV) sensitivity of inflation to Planck-scale physics. We then review two types of the effective theories of inflation, which are constructed from the background and perturbation perspectives.

The remaining chapters consist of the main works of the author. In Chapter 4, we use causality and unitarity to understand the implications of the underlying short-distance dynamics of inflation on cosmological observables. We obtain a consistency condition for primordial non-Gaussianity, which constrains the size and the sign of the four-point function in terms of the amplitude of the

three-point function. In Chapter 5, we study the imprints of massive particles with arbitrary spin on cosmological correlators. Using the framework of the effective theory of inflation, we classify the interactions of these particles. We then derive explicit shape functions for non-Gaussianity, which can serve as templates for future observational searches. In Chapter 6, we explore a novel observational signature of gravitational corrections during slow-roll inflation. We study the coupling of the inflaton field to higher-curvature tensors in models with a minimal breaking of conformal symmetry. Here, we show that these scenarios lead to a correction to the tilt of the tensor power spectrum and hence a violation of the tensor consistency condition. In Chapter 7, we introduce a method for measuring the inflationary superhorizon signature in the CMB polarization. We define an estimator that captures superhorizon correlations and present forecasts for the detectability of the signal with future CMB polarization experiments. We conclude in Chapter 8.

Finally, a number of appendices contain technical details of the results presented in the main text. In Appendix A, we provide further details to the calculations shown in Chapter 4. In Appendix B, we solve the equations of motion for massive fields with arbitrary spin in de Sitter space. The derivation of the shape functions introduced in Chapter 5 is given in Appendix C. In Appendix D, we make a few supplementary remarks to the results of Chapter 6. Appendix E contains additional reference material related to Chapter 7.

1.2 Notation and conventions

We will use natural units in which the speed of light, c, and Planck's constant, $\hbar$, are set to unity, $c = \hbar = 1$, with reduced Planck mass $M_{\rm pl}^2 = 1/8\pi G = 2.4 \times 10^{18}\,\text{GeV}$, where G denotes Newton's constant. Our metric signature is $(-+++)$. We will use Greek letters for spacetime indices, $\mu, \nu, \ldots = 0, 1, 2, 3$, and Latin letters for spatial indices, $i, j, \ldots = 1, 2, 3$. Repeated indices are summed, unless otherwise stated. Spatial three-dimensional vectors are written in boldface, $\mathbf{k}$, and unit vectors are hatted, $\hat{\mathbf{k}}$. The complex conjugate (Hermitian adjoint) of a function (operator) f is denoted by f^* ($f^\dagger$). A shorthand for the symmetrization of tensor indices is $a_{(\mu}b_{\nu)} \equiv \frac{1}{2}(a_\mu b_\nu + a_\nu b_\mu)$. Overdots and primes will denote derivatives with respect to physical time $x^0 \equiv t$ and conformal time $x^\eta \equiv \eta$, respectively. The symbols ∇_μ and ∂_μ denote the covariant and partial derivatives with respect to the coordinate x^μ, respectively. The notation $(\partial f)^2$ means $g^{\mu\nu}\partial_\mu f \partial_\nu f$. Our Fourier convention is

$$f_{\mathbf{k}} = \int \mathrm{d}^3x\, f(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}}\,, \qquad f(\mathbf{x}) = \frac{1}{(2\pi)^3}\int \mathrm{d}^3k\, f_{\mathbf{k}}\, e^{-i\mathbf{k}\cdot\mathbf{x}}\,. \tag{1.1}$$

The dimensionless power spectrum of a Fourier mode $f_{\mathbf{k}}$ is defined as

$$\Delta_f^2(k) \equiv \frac{k^3}{2\pi^2}\langle f_{\mathbf{k}} f_{\mathbf{k}'}\rangle'\,, \tag{1.2}$$

where the prime on the expectation value indicates that the overall momentum-conserving delta function $(2\pi)^3\delta(\mathbf{k}+\mathbf{k}')$ has been dropped.

2. Inflationary paradigm

Inflation is a hypothesized period in the very early universe, during which the universe expanded at an accelerating rate. This seemingly simple idea has deep implications and is one of the cornerstones of modern cosmology.

In the early 1980s, the first generation of inflationary scenarios was born. The original model by Guth [16] called *old inflation* was based on the first-order phase transition from a metastable false vacuum to the true lower-energy vacuum, forming bubbles that led to the present universe. This scenario, however, generated too much inhomogeneities by the bubble collision after the end of inflation, leading to what is known as the graceful exit problem. A *new inflation* scenario based on second-order phase transition was subsequently proposed by Linde [18] and independently by Albrecht and Steinhardt [19]. Although this scenario addressed some problems of the original inflation model, it also faced a fine-tuning issue of the initial condition.

An improved model of new inflation was then proposed by Linde [20], which did not share the problems of old and new inflation. This inflationary scenario, called *chaotic inflation*, does not rely on phase transitions from a special thermal state. Instead, the universe inflates as a scalar field slowly rolls down a sufficiently flat potential (in Planck units) towards the global minimum under fairly natural initial conditions. The success of this scenario opened up a new window of inflationary model-building based on the existence of the slow-roll regime, which has become the standard picture of inflation today.

Nearly three decades have passed since the birth of inflationary cosmology, during which theorists' inventiveness has led to the construction of a slew of inflationary models. In principle, the energy scale during inflation can be as high as $10^{16}\,\mathrm{GeV}$, making the inflationary universe the ultimate testing ground for theories of high-energy physics. A complete story of inflation will involve a full-fledged model constructed from Planck-scale physics, which must also successfully lead to the production of all particles of the SM after inflation. We are currently far from achieving such a grand goal, and inflation in its current state exists as a paradigm, which allows various constructions within a flexible framework. Nevertheless, despite the absence of the precise knowledge of its microscopic mechanism, the basic consequences of the inflationary paradigm are rather robust.

Although inflation was originally designed to address the classic problems of the Big Bang cosmology, it has even further-reaching implications. Most importantly, it provides a dynami-

cal mechanism for generating primordial density perturbations that gave rise to the temperature fluctuations in the CMB and provided seeds for the large-scale structure of the universe. The subtle features of the observed CMB anisotropies—small deviations from scale invariance, adiabaticity, and Gaussianity—are in excellent agreement with the basic predictions of the inflationary paradigm. Thanks to experimentalists' efforts, we have recently entered a theoretically interesting regime where we started ruling out some of the canonical models of inflation.

In this chapter, we will give an overview of the general properties and observational consequences of inflation. We start in §2.1 by reviewing the basic elements of Big Bang cosmology. We also discuss its classic problems and describe how inflation successfully addresses them. In §2.2, we present the main idea of inflation in the homogeneous limit and then review the basics of quantum fluctuations during inflation in §2.3. Finally, we show the current observational status of the inflationary paradigm in §2.4.

2.1 Big Bang cosmology

A well-established edifice of modern cosmology upon which inflation is based is the hot Big Bang model. This picture is supported by a wealth of observational evidence, including the CMB, the abundances of light elements, and the accelerated expansion of the universe. In this section, we will briefly review the successful story of the standard Big Bang cosmology.

2.1.1 A brief cosmic history

A summary of main events in the history of the universe is presented in Table 2.1. Due to the incessant expansion of spacetime, the energy density of the universe has been diluted ever since the Big Bang, and different physics and matter content were responsible for driving the expansion at different times. In essence, the history of the universe can be divided into three main eras, as we describe in the following.

Speculative epoch. Physics of energy scales up to about 1 TeV has been probed by the LHC. Unfortunately, energy scales much higher than 1 TeV cannot be probed by terrestrial experiments. Due to the lack of experimental data in this high-energy regime, the physics responsible for the very first moment of the Big Bang is rather uncertain. The standard picture is that there was a brief period of inflation, during which the universe underwent an accelerated expansion and primordial fluctuations were generated. The non-detection of primordial gravitational waves puts an upper bound of 10^{16} GeV on the inflationary energy scale, as we shall see below. Inflation is expected to be followed by a period of reheating, during which the familiar particles of the SM were produced and a smooth transition to a radiated-dominated universe happened.

Thermal epoch. This thermal epoch is a cornerstone of Big Bang cosmology. At high temperatures, all relativistic species were in thermal equilibrium. As the universe expanded and cooled, some particles started to become decoupled from the equilibrium.

Event	Time t	Redshift z	Density $\rho^{1/4}$	Temperature T
Quantum gravity?	10^{-43} s	-	10^{19} GeV	-
Grand unification?	10^{-36} s	-	10^{16} GeV	-
Inflation?	$> 10^{-36}$ s	-	$< 10^{16}$ GeV	-
EW symmetry breaking	10^{-12} s	10^{15}	100 GeV	10^{15} K
QCD phase transition	10^{-6} s	10^{12}	100 MeV	10^{12} K
ν decoupling	1 s	10^{10}	1 MeV	10^{10} K
$e^{\pm}$ annihilation	3 s	$6 \cdot 10^{9}$	0.5 MeV	$5 \cdot 10^{9}$ K
Big Bang nucleosynthesis	300 s	$4 \cdot 10^{8}$	0.1 MeV	10^{9} K
Ω_m domination	70 kyr	3400	1 eV	10^{4} K
Recombination	300 kyr	1300	0.3 eV	3600 K
γ decoupling	380 kyr	1100	0.26 eV	3000 K
Reionization	10 Myr	9	2.3 meV	27 K
Ω_Λ domination	900 Myr	0.4	0.32 meV	3.7 K
t_0	14 Gyr	0	0.23 meV	2.7 K

Table 2.1: Chronology of the universe. The last two columns show the energy density of the universe and the photon temperature, respectively.

At 100 GeV, the electroweak symmetry became spontaneously broken and particles acquired mass through the Higgs mechanism. At the QCD scale of 100 MeV, there was a quark-gluon phase transition and composite hadronic particles were formed. When the age of the universe was of about 1 second, the weak interaction became sufficiently weak so that neutrinos decoupled from the rest of the plasma. The temperature subsequently fell below the electron mass, after which electron-positron annihilation occurred. Around a few minutes after the Big Bang, protons and neutrons were bound together to form the first light elements—primarily hydrogen and helium-4—through the BBN.

Cold epoch. We are currently living in the cold epoch, in which the energy density from radiation is negligible. At 1 eV, pressureless matter became the dominant source to the energy density of the universe. An important event called *recombination* then happened at around 0.3 eV, which was when electrons and protons combined to form hydrogen atoms. Subsequently, the number of free electrons dropped rapidly, and the photon mean free path became longer than the horizon distance. In other words, photons became completely decoupled from matter at this point, and the universe became transparent. We see these relics of photons as the CMB today.

Small irregularities in the matter density became more pronounced due to gravitational at-

traction. As time went by, this gravitational instability led to the formation of structures such as galaxies. Beyond the galactic scale, we observe the uneven distribution of matter in the form of walls, filaments, and voids. The universe recently entered a dark-energy-dominated era and started expanding at an accelerating rate. The current estimated age of the universe is 13.8 billion years [21].

2.1.2 FLRW background

Since the time of Copernicus, we have learned that Earth does not occupy a privileged location in the universe, neither do our Solar System and the Milky Way. In modern cosmology, we assume that the matter distribution in our universe is spatially homogeneous and isotropic on sufficiently large scales—a notion referred to as the *cosmological principle*. Stated in simple terms, the universe looks essentially the same at any location and in all directions. This is not merely a philosophical assertion, but is heavily evidenced by the near uniformity of the observed CMB temperature and the large-scale structure in the universe. Mathematically, the cosmological principle imposes a stringent restriction on the possible types of spatial geometry. These are labelled by the curvature κ, categorized into flat ($\kappa = 0$), closed ($\kappa > 0$) and open ($\kappa < 0$) spaces. The Friedmann–Lemaître–Robertson–Walker (FLRW) line element that describes the homogeneous and isotropic spatial geometry in spherical polar coordinates is

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-\kappa r^2} + r^2\mathrm{d}\Omega_2^2\right], \tag{2.1}$$

where $\mathrm{d}\Omega_2^2 \equiv \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2$ is the line element of a 2-sphere. The symmetries of spatial translations and rotations are enough to fully specify the metric up to a scale factor $a(t)$ and the parameter κ. Introducing *conformal time* η, which is related to physical time t by

$$\mathrm{d}\eta = \frac{\mathrm{d}t}{a(t)}, \tag{2.2}$$

the FLRW line element also can be written as

$$\mathrm{d}s^2 = a^2(\eta)\left[\mathrm{d}\eta^2 + \frac{\mathrm{d}r^2}{1-\kappa r^2} + r^2\mathrm{d}\Omega_2^2\right]. \tag{2.3}$$

This is conformal to the line element in flat spacetime and hence particularly convenient for describing the propagation of light and the causal structure of spacetime.

Perfect fluids provide an idealized yet good description of the matter distribution in the universe. The energy-momentum tensor of a relativistic perfect fluid takes the form

$$T^{\mu\nu} = (\rho + P)u^\mu u^\nu + Pg^{\mu\nu}, \tag{2.4}$$

where ρ is the energy density and P is the pressure of the fluid in its local rest frame, and u^μ is the 4-velocity of the fluid with respect to a comoving observer. The time-time component of

Einstein's field equations gives the Friedmann equation:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu} \quad \Rightarrow \quad H^2 = \frac{8\pi G}{3}\rho - \frac{\kappa}{a^2}\,, \tag{2.5}$$

where $R_{\mu\nu}$ is the Ricci tensor, $R \equiv R^{\mu}{}_{\mu}$ is the Ricci scalar, and the Hubble parameter is defined by $H = \dot{a}/a$, with an overdot denoting the derivative with respect to physical time t. The conservation of the energy-momentum tensor in the local frame leads to the continuity equation:

$$\nabla^{\mu} T_{\mu\nu} = 0 \quad \Rightarrow \quad \dot{\rho} = -3H(\rho + P)\,. \tag{2.6}$$

The continuity equation together with the equation of state for the fluid determines the cosmological evolution of the fluid:

$$P = w\rho \quad \Rightarrow \quad \rho \propto a^{-3(1+w)}\,, \tag{2.7}$$

where w is a dimensionless parameter. The three main matter components of our universe are radiation, pressureless matter, and dark energy. Observations indicate that dark energy can be described by a cosmological constant Λ, which contributes to the energy-momentum tensor as

$$T_{\mu\nu} = \frac{\Lambda}{8\pi G} g_{\mu\nu}\,. \tag{2.8}$$

A summary of the equations of state and expansion rates for different matter components is given in Table 2.2.

	w	$\rho(a)$	$a(t)$	$a(\eta)$
radiation	$\frac{1}{3}$	a^{-4}	$t^{1/2}$	η
matter	0	a^{-3}	$t^{2/3}$	η^2
dark energy	-1	a^0	e^{Ht}	$-\eta^{-1}$

Table 2.2: Summary of equations of state and expansion rates in single-component universes.

It is convenient to define the critical density as

$$\rho_c \equiv \frac{3H^2}{8\pi G}\,, \tag{2.9}$$

and the dimensionless density parameters by

$$\Omega_r \equiv \frac{\rho_r}{\rho_c}\,, \quad \Omega_m = \frac{\rho_m}{\rho_c}\,, \quad \Omega_\kappa \equiv -\frac{\kappa}{H^2 a^2}\,, \quad \Omega_\Lambda \equiv \frac{\Lambda}{3H^2}\,, \tag{2.10}$$

for radiation, matter, curvature, and dark energy, respectively. In terms of these quantities, we

can recast the Friedmann equation (2.5) in the form

$$H^2 = H_0^2 \left[\Omega_{r,0} \left(\frac{a_0}{a} \right)^4 + \Omega_{m,0} \left(\frac{a_0}{a} \right)^3 + \Omega_{\kappa,0} \left(\frac{a_0}{a} \right)^2 + \Omega_{\Lambda,0} \right], \tag{2.11}$$

where the subscript '0' indicates that the quantity is evaluated today.

2.1.3 Big Bang puzzles

Despite the empirical success of the standard hot Big Bang model, it has major shortcomings in its failure to provide an explanation for the apparent fine-tuning of the initial conditions. We describe below the two most notable problems and show how inflation provides a natural solution to these problems.

Flatness problem. In the previous section, we saw that there are three unique geometries describing the spatial slices of the universe, characterized by the sign of κ. Observations are consistent with a negligible contribution to the effective energy density from curvature today, $|\Omega_{\kappa,0}| \lesssim 0.005$ [21], i.e. the current universe looks highly flat. This puts a tremendous constraint on its initial value: for example, we require $|\Omega_\kappa| \lesssim 10^{-16}$ at the time of nucleosynthesis, and even smaller at earlier times. Otherwise, the universe would have either recollapsed ($\kappa > 0$) or expanded too quickly ($\kappa < 0$). This apparent fine-tuning of the initial geometry of the universe is called the *flatness problem*. In Newtonian gravity, the curvature can be interpreted as the total energy of particles in a uniform expanding medium. The flatness problem can then be rephrased as a problem associated with their initial velocities, because it requires an exquisite cancellation between the kinetic and potential energies of particles in the early universe.

In order to understand the flatness problem in more mathematical terms, let us use the Friedmann and continuity equations to express the evolution for Ω_κ as

$$\frac{\mathrm{d}\Omega_\kappa}{\mathrm{d}\ln a} = (1+3w)\Omega_\kappa(1-\Omega_\kappa)\,. \tag{2.12}$$

We have set $\Omega_\Lambda = 0$, which is a good approximation for most of the history of the universe, and assumed that the universe is dominated by a single fluid with an equation of state $P = w\rho$. By examining the sign of the right-hand side of the above equation, we see that $\Omega_\kappa = 0$ is an unstable fixed point for $w \geq -1/3$, while $\Omega_\kappa = 0$ becomes an attractor for $w < -1/3$. Moreover, the rate of change is exponential in Hubble time. This implies that the curvature component grows exponentially in a universe dominated by ordinary fluids such as radiation or pressureless matter. On the other hand, an expansion driven by a fluid component that violates the strong energy condition (SEC), $w < -1/3$, will drive Ω_κ effectively to zero. The acceleration equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(1+3w)\rho\,, \tag{2.13}$$

tells us that this corresponds to having an accelerated expansion. The Hubble rate is almost

constant during inflation, which implies that $w \approx -1$. This in turn implies that $\Omega_\kappa \propto e^{-2N}$, where $N \propto \ln a$ is the number of e-folds during inflation. If we suppose that $|\Omega_\kappa| \sim 1$ at the beginning of inflation, then we require inflation to last at least $N = 60$ e-folds in order to ensure that $|\Omega_\kappa| < 10^{-50}$ at the end of inflation and thereby solve the flatness problem.

Horizon problem. Causality plays a crucial role in all branches of physics. For example, black holes have an event horizon, the surface beyond which causal contact with outside observers is lost. In cosmology, there exists a notion of a *particle horizon*, which is the maximum distance that particles can have travelled within the age of the universe.

The observed CMB is extremely isotropic (with the level of anisotropy being one part in 10^4), which implies that photons had been in causal contact before the time of photon decoupling. Let us define

$$\chi(t_1, t_2) \equiv \int_{a(t_1)}^{a(t_2)} \frac{\mathrm{d}a}{a} \frac{1}{aH} \,. \tag{2.14}$$

The particle horizon at the time of last scattering, t_L, is then given by $d_H \equiv \chi(0, t_L)$. Let us assume that the universe has evolved according to the standard history of the hot Big Bang without inflation. Before last scattering, the universe was dominated either by radiation or matter, during which the Hubble rate evolved as $H \propto a^{-2}$ and $H \propto a^{-3/2}$, respectively. For these normal matter contents, the integral is dominated by late times, since $(aH)^{-1}$ is monotonically increasing. In this case, d_H will be finite and small. Note that comoving distances that are separated by more than d_H cannot have been in causal contact. The angle subtended by the comoving particle horizon at recombination is $\theta_H \equiv d_H/d_A$, where $d_A \equiv \chi(t_L, t_0)$ is the angular diameter distance to the surface of last scattering in a flat universe. Numerically, we obtain $\theta_H \approx 1^\circ$, which implies that the near-uniform surface of last scattering was made out of 10^4 causally disconnected regions. This apparent fine-tuning of the initial condition of the universe is called the *horizon problem*.

Now, suppose that there was an era of inflation, during which the universe expanded as $a(t) \propto e^{Ht}$, before radiation came to dominate the universe. In that case, the integral for d_H becomes dominated by the period of inflation at early times, in contrast to what the naive extrapolation above suggests. The condition that all regions of the surface of last scattering have been in causal contact then translates to

$$d_H > d_A \quad \Rightarrow \quad e^N \gtrsim \frac{a_* H_*}{a_0 H_0} \,, \tag{2.15}$$

where the subscript '$*$' indicates that the quantity is evaluated at the end of inflation, and we used the fact that the integral for d_A is dominated by the matter-dominated era. In other words, we require inflation to have lasted sufficiently long, so that the comoving Hubble radius $(aH)^{-1}$ today is smaller than its size at the end of inflation. This is satisfied if inflation lasted at least $N \gtrsim 60$ e-folds.[1]

[1] This is a conservative bound assuming $\rho_*^{1/4} \sim 10^{16}$ GeV, and the exact number of required N depends on the

2.2 Inflation I: homogeneous limit

As we have seen above, positing a period of inflation can resolve the classic shortcomings of Big Bang cosmology. In this section, we describe the basic elements of the homogeneous limit of inflation.

Definitions. We begin by describing several equivalent definitions of inflation. In the simplest form, inflation is just a period of accelerated expansion, $\ddot{a} > 0$. As we have seen, this implies that the fluid that drives the expansion violates the SEC, $w < -1/3$. If $w \approx -1$, then the space will expand quasi-exponentially. More exotic theories allow a violation of the null energy condition (NEC), $w < -1$, but we will assume that the NEC always holds.[2]

Inflation can also be characterized in terms of the evolution of the comoving Hubble radius $(aH)^{-1}$. In particular, since

$$\frac{\mathrm{d}}{\mathrm{d}t}(aH)^{-1} = -\frac{\ddot{a}}{\dot{a}^2} < 0\,, \tag{2.16}$$

the comoving Hubble radius decreases during inflation. From (2.14), we see that the comoving Hubble radius has the notion of an instantaneous cosmological horizon, in the sense that the distances that are separated by greater than the Hubble radius cannot be in causal contact at a particular moment. The observable universe therefore becomes smaller during inflation.

The condition of having an accelerated expansion furthermore implies

$$\varepsilon \equiv -\frac{\dot{H}}{H^2} = 1 - \frac{\ddot{a}a}{\dot{a}^2} < 1\,. \tag{2.17}$$

Notice than an exponential expansion is achieved by taking the limit

$$\varepsilon \to 0 \qquad \Rightarrow \qquad H \to \text{const.}\,, \quad a(t) \to e^{Ht}\,. \tag{2.18}$$

The exact $\varepsilon = 0$ case corresponds to having a de Sitter background, which is a vacuum solution of the Einstein equations with a positive cosmological constant. A summary of the four widely-used equivalent definitions of inflation is shown in Table 2.3.

accelerated expansion	violation of SEC	shrinking Hubble radius	slowly varying H
$\ddot{a} > 0$	$P < -\frac{1}{3}\rho$	$\frac{\mathrm{d}}{\mathrm{d}t}(aH)^{-1} < 0$	$\varepsilon = -\frac{\dot{H}}{H^2} < 1$

Table 2.3: Equivalent definitions of inflation

energy scale at the end of inflation and the details of the subsequent cosmological evolution.

[2]More precisely, the NEC requires that the energy-momentum tensor satisfies $T_{\mu\nu}n^\mu n^\nu \geq 0$ for any null vector n^μ.

Slow-roll inflation. The exponential expansion of inflation would never end, if the universe were in a perfect de Sitter phase. For realistic inflationary scenarios, we instead consider a quasi-de Sitter phase, for which the expansion rate is almost—but not exactly—exponential. This is no longer a vacuum solution of Einstein's field equations, which means that we must add a matter content to drive inflation—in particular, one that violates the SEC. This can be achieved by a scalar field ϕ, called the inflaton, whose action with minimal coupling to gravity is

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}M_{\rm pl}^2 R - \frac{1}{2}(\partial\phi)^2 - V(\phi)\right], \tag{2.19}$$

where $V(\phi)$ is the inflaton potential. In a pure FLRW background, ϕ is a function only of time. The variation of this action with respect to ϕ yields the equation of motion

$$\ddot{\phi} + 3H\dot{\phi} = -V'(\phi)\,, \tag{2.20}$$

with a prime on $V(\phi)$ denoting the derivative with respect to ϕ. Moreover, the Friedmann equation implies[3]

$$H^2 = \frac{1}{3M_{\rm pl}^2}\left[\frac{1}{2}\dot{\phi}^2 + V(\phi)\right]. \tag{2.21}$$

Differentiating this with respect to time and using (2.20), we get $2M_{\rm pl}^2\dot{H} = -\dot{\phi}^2$. In order to have a nearly exponential expansion for a sufficient period, we require that the following *slow-roll conditions* to be satisfied:

$$\varepsilon = \frac{\dot{\phi}^2}{2H^2M_{\rm pl}^2} \ll 1\,, \quad \tilde{\eta} \equiv \frac{\dot{\varepsilon}}{H\varepsilon} \ll 1\,. \tag{2.22}$$

These conditions imply that the potential energy dominates the kinetic energy, $\frac{1}{2}\dot{\phi}^2 \ll V(\phi)$, for a sufficient period, and, as a result, that the scalar field slowly rolls down the potential. In this regime, the equivalent slow-roll conditions can also be defined in terms of the potential as

$$\varepsilon_V \equiv \frac{M_{\rm pl}^2}{2}\left(\frac{V'}{V}\right)^2 \ll 1\,, \quad \eta_V \equiv M_{\rm pl}^2\frac{V''}{V} \ll 1\,. \tag{2.23}$$

These conditions state that the first and second derivatives of the potential are small in Planck units, and we will refer to these as the *flatness conditions*. For example, one of the simplest potentials that satisfy the flatness conditions is $V(\phi) = \frac{1}{2}m^2\phi^2$ with $\phi \gg M_{\rm pl}$.

[3] We will henceforth switch from using G to $M_{\rm pl} = 1/\sqrt{8\pi G}$, which is more convenient to use in the high-energy regime.

2.3 Inflation II: quantum fluctuations

We know that the universe is not perfectly homogeneous and isotropic because we see structures around us. One of the beautiful aspects of inflation is that it provides a natural mechanism for generating primordial density perturbations.

Observations of the CMB anisotropies provide information about several important properties of the primordial density perturbations: that they are nearly (i) *adiabatic*, meaning that no fluctuations were in composition, (ii) *scale invariant*, implying that the variance is constant across logarithmic k-intervals, and (iii) *Gaussian*, so that their statistical distribution follows entirely from their two-point function. All of these are predicted features in the simplest picture of slow-roll inflation, where these perturbations were generated from quantum fluctuations of the scalar field driving the accelerated expansion.

The inflaton can be expressed as $\phi(t,\mathbf{x}) = \bar{\phi}(t) + \delta\phi(t,\mathbf{x})$, where $\bar{\phi}$ denotes the homogeneous background solution (the one studied in the previous section) and $\delta\phi$ is a spacetime-dependent perturbation. The latter arises from vacuum fluctuations of the inflaton, which "freeze" on superhorizon scales and become classical perturbations. These then lead to inhomogeneities in the energy density after inflation and ultimately to the observed anisotropies of the CMB. In this section, we will give a lightening review of quantum fluctuations during inflation.

Inflaton perturbations. Due to one of the flatness conditions, $\eta_V \ll 1$, we require the inflaton to be sufficiently light. For simplicity, let us take the inflaton to be massless. The linearized equation of motion for $\delta\phi$ in an *unperturbed* de Sitter background is

$$\ddot{\delta\phi} + 3H\dot{\delta\phi} - a^{-2}\nabla^2\delta\phi = 0\,. \tag{2.24}$$

In Fourier space, this equation expressed in terms of conformal time is

$$\delta\phi_{\mathbf{k}}'' - \frac{2}{\eta}\delta\phi_{\mathbf{k}}' + \frac{k^2}{H^2\eta^2}\delta\phi_{\mathbf{k}} = 0\,. \tag{2.25}$$

We now follow the standard procedure of quantization by promoting $\delta\phi$ to an operator and expanding it in terms of time-dependent mode functions

$$\delta\phi_{\mathbf{k}} = \delta\phi_k(\eta)a_{\mathbf{k}} + \delta\phi_k^*(\eta)a_{-\mathbf{k}}^\dagger\,. \tag{2.26}$$

The momentum-dependent coefficients $a_{\mathbf{k}}$ and $a_{\mathbf{k}}^\dagger$ play the role of the annihilation and creation operators, respectively, which satisfy the canonical commutation relations

$$[a_{\mathbf{k}}, a_{\mathbf{k}'}^\dagger] = (2\pi)^3\delta(\mathbf{k}+\mathbf{k}')\,, \quad [a_{\mathbf{k}}, a_{\mathbf{k}'}^\dagger] = [a_{\mathbf{k}}^\dagger, a_{\mathbf{k}'}^\dagger] = 0\,. \tag{2.27}$$

The vacuum is a state satisfying $a_{\mathbf{k}}|0\rangle = 0$. The canonical commutation relations (2.27) impose the normalization condition $a^2(\delta\phi_k'\delta\phi_k^* - \delta\phi_k\delta\phi_k^{*\prime}) = i$ on the mode functions, known as the

Wronskian condition. The normalized solution of the equation (2.25) is

$$\delta\phi_k(\eta) = \frac{iH}{\sqrt{2k^3}}(1+ik\eta)e^{-ik\eta}\,, \tag{2.28}$$

where we have imposed the Bunch-Davies initial condition, so that it reduces to the correct vacuum solution in the far past.[4] At late times when the modes have exited the horizon, $|k\eta| \ll 1$, the power spectrum of the inflaton perturbations becomes

$$P_{\delta\phi}(k) = \langle \delta\phi_{\mathbf{k}}\delta\phi_{-\mathbf{k}}\rangle' = |\delta\phi_k|^2 = \frac{H^2}{2k^3}\,. \tag{2.29}$$

Note that this is time independent and exactly scale invariant.

A few comments are in order. The above analysis was carried out in a fixed de Sitter background and ignored the coupling to metric fluctuations. However, for true inflationary solutions, the perturbations of the spacetime metric need to be taken into account as well. The full gravitational dynamics is most conveniently described by using the Arnowitt-Deser-Misner (ADM) formalism [24], where the metric takes the form

$$\mathrm{d}s^2 = -N^2\mathrm{d}t^2 + h_{ij}(N^i\mathrm{d}t + \mathrm{d}x^i)(N^j\mathrm{d}t + \mathrm{d}x^j)\,. \tag{2.30}$$

Neglecting the tensor fluctuations inside h_{ij} for the moment (which we will describe below), we can fix the gauge by setting $\phi = \bar{\phi} + \delta\phi$ and $h_{ij} = a^2\delta_{ij}$. This is called *spatially flat gauge*, in which the scalar degree of freedom is solely described by $\delta\phi$. The ADM formalism is designed so that when the above metric is substituted into the action (2.19), N and N^i act as Lagrange multipliers. Solving for the constraints, we will find that $N = 1$ and $N^i = 0$ at leading order with corrections that are suppressed by slow-roll parameters [25]. Therefore, the result (2.29) is correct up to slow-roll-suppressed gravitational corrections. On the other hand, the slow-roll approximation breaks down at sufficiently late times, and, as a result, the inflaton fluctuations evolve on superhorizon scales. This means that the above result should only be trusted up to a few e-folds after the horizon exit. On superhorizon scales, it is more convenient to describe the scalar degree of freedom in a different gauge, which we will describe below.

Metric perturbations. Another convenient choice of gauge is *comoving gauge*, where (after reintroducing tensor fluctuations) we set

$$\phi = \bar{\phi}\,, \quad h_{ij} = a^2e^{2\zeta}[e^\gamma]_{ij}\,, \quad \partial_i\gamma_{ij} = 0\,, \quad \gamma_{ii} = 0\,. \tag{2.31}$$

This choice completely fixes the gauge freedom at non-zero momentum.[5] In this gauge, the scalar

[4] There exists a one-parameter family of de Sitter-invariant states called the α-vacua, which makes the choice of a vacuum state in de Sitter space ambiguous. However, since the mode does not feel any curvature in the limit of short wavelength (the equivalence principle), there is a particular de Sitter-invariant state called the Bunch-Davies vacuum that reduces to the Minkowski vacuum in the asymptotic past [22, 23].

[5] At zero momentum, there are residual symmetries involving gauge transformations that do not fall off at

fluctuations are fully described by the gauge-invariant variable ζ [27, 28] called the (comoving) *curvature perturbation*, while the tensor fluctuations are captured by the *tensor perturbation* γ_{ij}. Expanding the action (2.19) up to second order in ζ, we obtain [25]

$$S = M_{\rm pl}^2 \int \mathrm{d}t\, \mathrm{d}^3x\, a^3 \varepsilon \left[\dot{\zeta}^2 + a^{-2}(\partial_i \zeta)^2\right]. \tag{2.32}$$

No slow-roll approximation is required for deriving this action, so it remains valid even at the end of inflation. Note that, due to the appearance of ε in the overall normalization, the action vanishes in the exact de Sitter limit $\varepsilon = 0$. This means that ζ becomes a pure gauge mode in an exact de Sitter background, as it should for a vacuum solution of Einstein's field equations. Variation of the above action with respect to ζ yields the equation of motion

$$\ddot{\zeta} + 3H(1+\tilde{\eta})\dot{\zeta} - a^{-2}\partial_i^2 \zeta = 0\,. \tag{2.33}$$

We can solve this equation by going to Fourier space and follow the standard quantization procedure for ζ as before.[6] At leading order in the slow-roll approximation, we can neglect $\tilde{\eta}$ in the friction term, after which the equation takes the same form as (2.24). The resulting *scalar power spectrum* is then

$$P_\zeta(k) = \frac{H^2}{4\varepsilon M_{\rm pl}^2}\frac{1}{k^3}\,. \tag{2.34}$$

The extra factors in the denominator appears relative to (2.29) because ζ is not canonically normalized. Although the slow-roll approximation is made in deriving the time-independent result (2.34), we can show that the exact equation (2.33) always has a constant ζ solution on superhorizon scales.[7] To compare the results between comoving and spatially flat gauges, we note that a time reparametrization of the form

$$t \;\to\; t + \xi(t, \mathbf{x})\,, \quad \text{with} \quad \xi = -\frac{\delta\phi}{\dot{\phi}} + \mathcal{O}(\delta\phi^2)\,, \tag{2.36}$$

takes us from spatially flat gauge to comoving gauge. This gives the following relation between ζ in comoving gauge and $\delta\phi$ in spatially flat gauge:

$$\zeta = -\frac{H\delta\phi}{\dot{\phi}} + \mathcal{O}(\delta\phi^2)\,. \tag{2.37}$$

infinity [26]. As we will describe in §3.3.2, these are conformal transformations that are non-linearly realized by ζ and γ_{ij}.

[6] This needs to take into account the fact that ζ is not canonically normalized.

[7] Let us consider the behavior of ζ outside the horizon. On superhorizon scales, we can neglect the gradient term in (2.33), and the general solution is

$$\zeta(t) = c_1 + c_2 \int^t \mathrm{d}t_1 \exp\left[-\int^{t_1} \mathrm{d}t_2\,(3+\tilde{\eta})H\right]. \tag{2.35}$$

We see that there is a constant solution with $c_1 \neq 0$. The other solution with $c_2 \neq 0$ decays as long as $\tilde{\eta} > -3$.

Using this relation and $\varepsilon = \dot{\bar{\phi}}^2/2H^2M_{\rm pl}^2$, it is straightforward to see that (2.29) and (2.34) are equivalent at leading order. Since (2.29) is only valid in the slow-roll regime, the conversion from spatially flat to comoving gauge should be made shortly after the mode exits the horizon.

In an inflationary background, the power spectrum (2.34) is not exactly scale invariant because the Hubble parameter slowly evolves with time. The small deviation from scale invariance is characterized by the scalar *spectral index* n_s, defined by

$$n_s - 1 \equiv \frac{\mathrm{d}\ln k^3 P_\zeta(k)}{\mathrm{d}\ln k}\,. \tag{2.38}$$

Evaluating the power spectrum at horizon crossing, $k = aH$, and using the relation $\mathrm{d}\ln k \approx H\mathrm{d}t$, we obtain $n_s - 1 = -2\varepsilon - \tilde{\eta}$ to first order in the slow-roll approximation.[8]

Apart from the scalar fluctuations, inflation also predicts a stochastic background of tensor perturbations or primordial gravitational waves. Expanding the action (2.19) up to second order in γ, we obtain

$$S = \frac{1}{4}M_{\rm pl}^2\int \mathrm{d}t\,\mathrm{d}^3x\,a^3\left[\dot{\gamma}_{ij}^2 - a^{-2}(\partial_k\gamma_{ij})^2\right]. \tag{2.39}$$

The transverse and traceless conditions (2.31) imply that γ contains two physical degrees of freedom, which in Fourier space can be decomposed into definite polarization modes as

$$\gamma_{ij,\mathbf{k}} = \sum_{\lambda=\pm 2}\varepsilon_{ij}^{\lambda}\gamma_{\mathbf{k}}^{\lambda}\,, \tag{2.40}$$

where $\varepsilon_{ij}^{\lambda}$ is the spin-2 transverse polarization tensor that satisfies $\varepsilon_{ii}^{\lambda} = 0$, $k_i\varepsilon_{ij}^{\lambda} = 0$, and $\varepsilon_{ij}^{\lambda}\varepsilon_{ij}^{\lambda'} = 4\delta_{\lambda\lambda'}$.[9] Each polarization mode then obeys the same equation of motion as a massless scalar field studied in the previous section. Following the same quantization procedure as before, it can be shown that the *tensor power spectrum* is [32]

$$P_\gamma(k) = \langle\gamma_{ij,\mathbf{k}}\gamma_{ij,-\mathbf{k}}\rangle' = \frac{H^2}{M_{\rm pl}^2}\frac{4}{k^3}\,. \tag{2.41}$$

This is exactly scale invariant in the de Sitter limit, but a small deviation from scale invariance in inflation is quantified by the *tensor tilt* n_t, defined by

$$n_t \equiv \frac{\mathrm{d}\ln k^3 P_\gamma(k)}{\mathrm{d}\ln k}\,. \tag{2.42}$$

Primordial gravitational waves from inflation in Einstein gravity give $n_t = -2\varepsilon$.

[8] The exact scale-invariant case, called the Harrison-Zeldovich spectrum, was proposed well before inflation [29–31]. Inflation provides a physical mechanism that naturally realizes the scale-invariant spectrum at leading order.

[9] Notice that we adopt a slightly different normalization of the polarization tensor from the more common convention $\varepsilon_{ij}^{s}\varepsilon_{ij}^{s'} = 2\delta_{ss'}$.

2.4 Observational status of the paradigm

We have seen that there exist two fundamental fluctuations created during inflation: the curvature perturbation ζ and the tensor perturbation γ_{ij}. In this section, we show how these fluctuations lead to the CMB anisotropies and review the current observational status of the inflationary paradigm.

2.4.1 CMB anisotropies

Although the temperature of the CMB is extremely isotropic with the mean blackbody temperature of $\overline{T} = 2.7255 \pm 0.0006$ K [33], there are small anisotropies in one part in 10^4. The temperature variation δT is a function of the direction on the sky, $\mathbf{n}$. Since we observe the CMB on the celestial sphere, we can expand the CMB anisotropy $\Theta \equiv \delta T(\mathbf{n})/\overline{T}$ in terms of spherical harmonics as

$$\Theta(\mathbf{n}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\mathbf{n})\,. \tag{2.43}$$

Assuming that the initial perturbations are statistically isotropic, the two-point correlation of the expansion coefficients $a_{\ell m}$ becomes diagonal:

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = C_\ell \delta_{\ell\ell'} \delta_{mm'}\,. \tag{2.44}$$

The CMB angular power spectrum is related to the primordial power spectrum by

$$C_\ell = 4\pi \int_0^\infty \frac{\mathrm{d}k}{k}\, T_\ell^2(k) P_\zeta(k)\,, \tag{2.45}$$

where $T_\ell(k)$ is the transfer function that encodes the cosmological evolution and projection effects. Since the transfer function can be computed using known physics, an accurate measurement of the angular power spectrum allows us to reconstruct the shape of the primordial power spectrum.

Figure 2.1 shows the latest measurement of the CMB temperature power spectrum by Planck and the best fit to the ΛCDM model. On large scales, the modes are still outside the horizon, and the angular power spectrum stays approximately constant, reflecting the scale invariance of the initial spectrum. On small scales, we see acoustic oscillations due to the gravitational compression and radiative pressure felt by the photon-baryon fluid. Crucially, the very fact that we see these acoustic peaks implies that the modes had coherent initial phases, which would be the case for perturbations generated during inflation [34]. In contrast, generic post-inflationary mechanisms produce modes with random phases, resulting in a featureless power spectrum. These properties of the observed CMB provide compelling qualitative evidence for inflation.

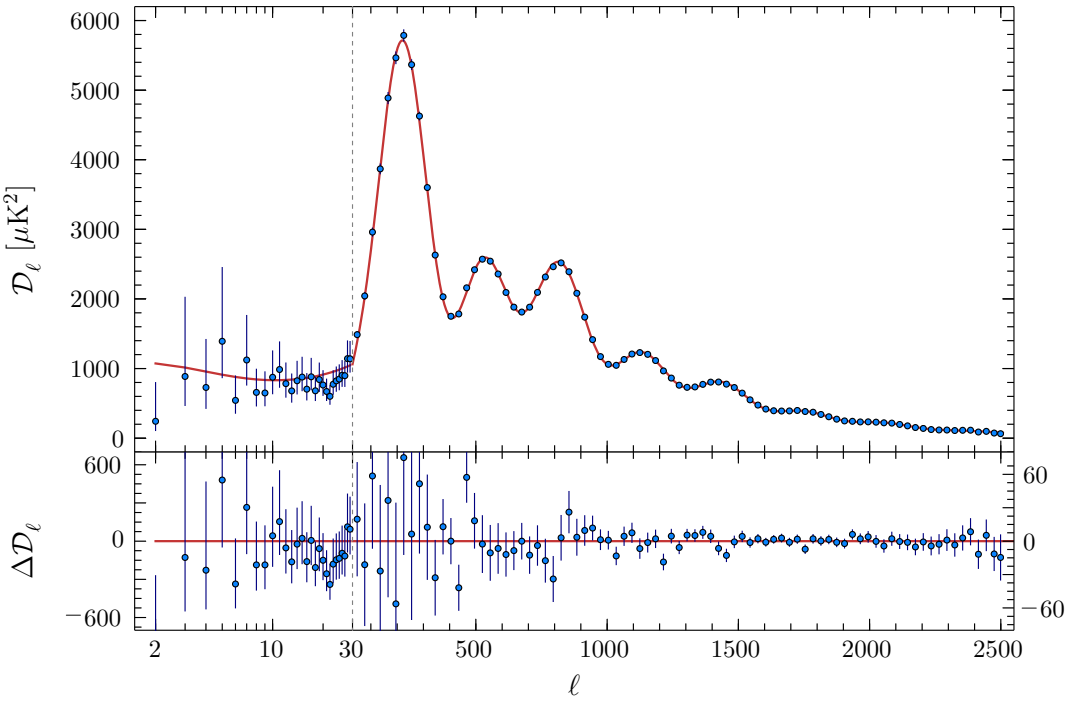


Figure 2.1: The Planck 2015 temperature power spectrum $\mathcal{D}_\ell \equiv \ell(\ell+1)C_\ell/2\pi$. The upper and lower panels show the best fit to the ΛCDM model and the residuals with respect to this fit, respectively. The error bars show 1σ uncertainties. (Figure adapted from [21].)

2.4.2 Constraints on inflation

The latest constraints on the six parameters of the ΛCDM model are quoted in Table 2.4. Four of these parameters describe the composition of the late-time universe, and two parameters set the initial conditions for the density perturbations: the amplitude of the scalar power spectrum, A_s, and the spectral index, n_s. We consider a phenomenological parameterization of the power spectrum by defining a dimensionless power spectrum

$$\Delta^2_\zeta(k) \equiv \frac{k^3}{2\pi^2} P_\zeta(k) = A_s \left(\frac{k}{k_*}\right)^{n_s-1} , \tag{2.46}$$

where k_* is a pivot scale (e.g. $k_* = 0.05\,\mathrm{Mpc}^{-1}$ for Planck). This parameterization takes into account the fact that the scalar power spectrum in (2.34) is not exactly scale invariant during inflation, and the $n_s = 1$ case corresponds to the scale-invariant spectrum in de Sitter space.

The observed broken scale-invariant spectrum provides a strong quantitative evidence for inflation, with the exact scale invariance now being disfavored at slightly more than 5σ [35]. Unfortunately, the value of n_s alone contains limited information about the dynamics of inflation. In order to further elucidate the physics of inflation, it is imperative to measure other cosmological observables which encode richer information about the primordial universe. Below we will

highlight two important cosmological observables that have the potential to probe the detailed physics of inflation.

H_0	$\Omega_{c,0}h^2$	$\Omega_{b,0}h^2$	τ	$\ln(10^{10}A_s)$	n_s
67.74 ± 0.46	0.1188 ± 0.0010	0.02230 ± 0.00014	0.066 ± 0.012	3.064 ± 0.023	0.9667 ± 0.0040

Table 2.4: Six independent parameters of the ΛCDM model at the 1σ level from Planck in combination with external data [21]. The symbols $\Omega_{c,0}$ and $\Omega_{b,0}$ denote the cold dark matter and baryonic matter densities, respectively, τ is the optical depth, and $h \equiv H_0/100\,\mathrm{km\ s^{-1}Mpc^{-1}}$.

Non-Gaussianity. If the curvature perturbations were perfectly Gaussian, then their statistical properties would be completely dictated by their two-point correlation function or power spectrum. Although the power spectrum is very tightly constrained due to its near scale invariance, higher-order correlation functions are more model specific and thus can serve as powerful discriminants for inflationary models. Due to the highly Gaussian nature of the observed density perturbations, the three-point function or its Fourier counterpart, the *bispectrum*, is the leading-order statistics that can distinguish between Gaussian and non-Gaussian perturbations. The primordial bispectrum is defined by

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_\zeta(k_1, k_2, k_3)\,. \tag{2.47}$$

The delta function enforces momentum conservation, and the function $B_\zeta(k_1, k_2, k_3)$ depends on the shape of the triangle formed by the three momentum vectors. The size of the bispectrum is customarily characterized by the non-linearity parameter

$$f_{\mathrm{NL}} \equiv \frac{5}{18} \frac{B_\zeta(k,k,k)}{P_\zeta(k)^2}\,, \tag{2.48}$$

which is its normalized amplitude in the equilateral configuration $k_1 = k_2 = k_3 = k$.

In standard single-field slow-roll inflation, gravitational non-linearities are the main source of non-Gaussianity. The level of non-Gaussianity in these scenarios is very small, due to the sheer weakness of gravitational interactions. On the other hand, there exist a large number of well-motivated extensions to single-field slow-roll inflation that can accommodate a detectable amplitude of non-Gaussianity, each of which predict a distinct shape for the bispectrum. We will briefly review three of the most-studied shapes below.

In single-field inflation, higher-derivative interactions of the inflaton can induce a non-trivial sound speed c_s for the curvature perturbation. At leading order in derivatives, the dominant contribution to the bispectrum is produced by the two cubic self-interactions $\dot\zeta^3$ and $\dot\zeta(\partial_i\zeta)^2$, with the size of the latter being tied to the deviation of c_s from unity [36]. Although the former is an independent operator, radiative corrections naturally give it a size of the same order [37]. These interactions give rise to two independent non-Gaussian shapes which are both peaked in

the equilateral configuration. For actual data analysis, it is convenient to work with templates that are computationally more efficient, rather than dealing with the exact shapes. The original shapes from the cubic interactions are well-captured by a linear combination of the two templates called *equilateral* and *orthogonal* shapes, defined by [38, 39]

$$B_\zeta^{\rm equil} = \frac{3}{5} f_{\rm NL}^{\rm equil} \left[\left(P_1 P_2^2 P_3^3\right)^{1/3} - 2\left(P_1 P_2 P_3\right)^{2/3} - P_1 P_2 + \text{perms.} \right], \tag{2.49}$$

$$B_\zeta^{\rm ortho} = \frac{3}{5} f_{\rm NL}^{\rm ortho} \left[3(P_1 P_2^2 P_3^3)^{1/3} - 3P_1 P_2 - 8\left(P_1 P_2 P_3\right)^{2/3} + \text{perms.} \right], \tag{2.50}$$

where $P_i \equiv P_\zeta(k_i)$. The amplitude of the equilateral non-Gaussianity typically scales as $f_{\rm NL}^{\rm equil} \sim c_s^{-2}$ for small c_s.

The class of inflationary models which involves additional light scalar fields is called *multi-field inflation*. In this scenario, extra scalar perturbations can transfer their non-Gaussianity to the curvature perturbation. This transfer happens locally on superhorizon scales, and thus the type of non-Gaussianity generated in this scenario is called *local* non-Gaussianity. The bispectrum in this scenario takes the form [12]

$$B_\zeta^{\rm local} = \frac{6}{5} f_{\rm NL}^{\rm local} \left(P_1 P_2 + \text{perms.}\right). \tag{2.51}$$

As opposed to the equilateral shape, the local shape is peaked in the squeezed configuration, where one side of the triangle is taken to be much smaller than the other two.

Current observational constraints on the size of the local, equilateral, and orthogonal non-Gaussian shapes are [40]

$$f_{\rm NL}^{\rm local} = 0.8 \pm 5.0\,, \quad f_{\rm NL}^{\rm equil} = -4 \pm 43\,, \quad f_{\rm NL}^{\rm ortho} = -26 \pm 21 \quad (68\%\ \text{CL})\,. \tag{2.52}$$

These constraints can be compared with the minimal amount of non-Gaussianity expected in single-field slow-roll inflation, $f_{\rm NL} = \mathcal{O}(\varepsilon, \tilde{\eta}) = \mathcal{O}(10^{-2})$ [25, 41], and a natural theoretical threshold, $f_{\rm NL} = \mathcal{O}(1)$ [1, 42] (see Chapter 4). There exist many different shapes of non-Gaussianity that arise from other physical inflationary scenarios. We will explore primordial non-Gaussianity more in Chapters 4 and 5.

Tensor modes. Primordial gravitational waves are arguably the most quintessential prediction of inflation. Similar to the scalar power spectrum, we consider the following phenomenological parametrization of the tensor power spectrum:

$$\Delta_\gamma^2(k) \equiv \frac{k^3}{2\pi^2} P_\gamma(k) = A_t \left(\frac{k}{k_*}\right)^{n_t}. \tag{2.53}$$

Again, due to the expected near scale invariance, this is characterized by two numbers: the tensor amplitude, A_t, and the tensor tilt, n_t. We conventionally characterize the size of the tensor power spectrum by the *tensor-to-scalar ratio*, r, defined as the ratio of the amplitudes of the tensor and

scalar power spectra

$$r \equiv \frac{\Delta_\gamma^2}{\Delta_\zeta^2} = 16\varepsilon\,. \tag{2.54}$$

Constraints on r can be placed both indirectly from the CMB temperature power spectrum and directly from the CMB B-mode polarization (see Chapter 7). Figure 2.2 shows the constraints on r and n_s from the joint analysis of BICEP2/Keck Array and Planck [43]. We see that the ϕ^2 model of inflation is now ruled out at the 2σ level.

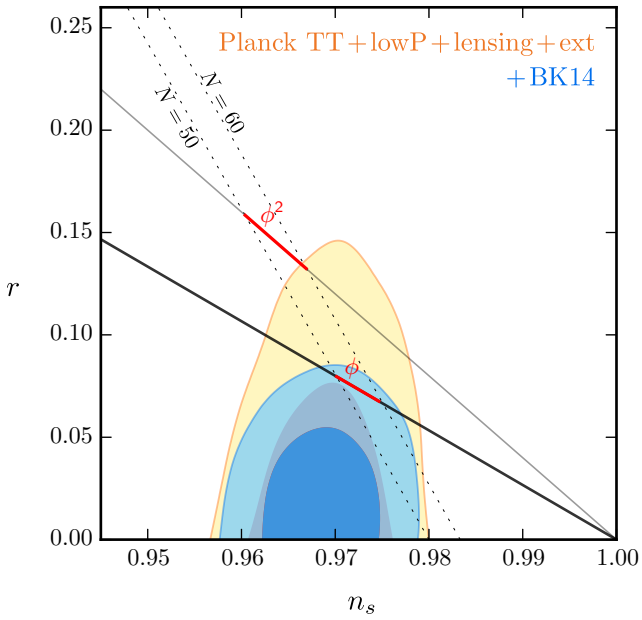


Figure 2.2: Current constraints in the r vs. n_s plane at the pivot scale $k_* = 0.002\,\mathrm{Mpc}^{-1}$. The yellow contour represents the combined constraint from Planck plus external data, while the blue contour also includes the BICEP2/Keck Array data. The dotted lines show lines of constant number of e-folds, N, for power-law potentials, and the solid lines show the r-n_s relation for linear and quadratic potentials. (Figure adapted from [43].)

From (2.41), we see that the tensor amplitude is a direct measure of the Hubble scale during inflation. In the slow-roll regime, this is related to the energy density $\rho \approx V$ during inflation by the Friedmann equation $3H^2M_{\mathrm{pl}}^2 \approx V$. Using the known value for M_{pl}, we can thus express ρ in terms of r as

$$\rho^{1/4} \approx \left(\frac{r}{0.01}\right)^{1/4} \times 10^{16}\ \mathrm{GeV}\,. \tag{2.55}$$

With the next generation of ground-based, balloon, and satellite CMB experiments, we expect to be able to reach $r \sim 10^{-3}$ [44, 45]. Thus, a detection of primordial gravitational waves in the near future would be the tell-tale signature that inflation occurred near the grand unification scale.

In single-field slow-roll inflation, the tensor tilt is given by $n_t = -2\varepsilon = 2\dot{H}/H^2$. Note that n_t must be negative, if the NEC is satisfied, $\dot{H} \leq 0$. Moreover, n_t is not an independent parameter, but is related to r in the following way:

$$\frac{n_t}{r} = -\frac{1}{8}\,. \tag{2.56}$$

This is known as the *tensor consistency condition* and is obeyed by all conventional single-field slow-roll inflationary models. In single-field models with non-trivial sound speed or multi-field inflation, we instead have $n_t/r > -1/8$. We will study inflationary tensor modes in more detail in Chapters 6 and 7.

3. Effective theories of inflation

Nature consists of many different length scales: from the very small—the quantum realm of particles—to the very large—the size of the cosmic horizon. Fortunately, describing a particular physical phenomenon does not require us to know physics at all scales at once. For instance, one does not speak of quantum chromodynamics when doing atomic spectroscopy or quantum mechanics when describing the motions of planets. Essentially all the relevant macroscopic dynamics at some length scale can be described without knowing the detailed microscopic description at a shorter scale, as long as the separation of these scales is sufficiently large. This is the essence of *effective field theory* (EFT).

In physical systems with a separation of scales, the effects of the short-distance physics on the long-wavelength dynamics are said to decouple, up to corrections that are suppressed by the hierarchy of scales.[1] Moreover, the higher-order corrections in the EFT can be classified by the symmetries of the problem. This allows us to parameterize our ignorance about the microscopic theory in a systematic way. Furthermore, in order to compute observables at a finite level of accuracy, only a finite number of corrections need to be included, making the EFT a predictive framework. Precision measurements of these higher-order corrections can provide important theoretical hints for the structure of the underlying theory.

The methods of EFT allow us to study the physics of inflation in a model-insensitive way. In the simplest setting, the low-energy theory of inflation is described by a single scalar degree of freedom, consisting of a homogeneous time-dependent background solution and spacetime-dependent perturbations. In this case, one can consider effective descriptions for the background and perturbations separately. The former provides a useful framework for inflationary model-building, whereas the latter allows us to describe the phenomenology of single-field inflation in the most general way.

This chapter is organized as follows. We begin by describing the general philosophy of EFT in §3.1. Next, we give details of EFT approaches to inflation from the perspectives of the background field and that of the perturbations. In §3.2, we provide an overview of Weinberg's work on the EFT

[1] The multipole expansion in gravity or electrodynamics provides a good illustration for the idea. Sufficiently far from a a distribution of masses or charges, the potential can be approximated by a monopole (point source), while higher-order terms in the expansion can be systematically incorporated. The amplitudes of the higher-order moments (dipole, quadrupole, etc.) are suppressed by a ratio of scales: the size of the extended source over the distance to the source.

of slow-roll inflation [46] and comment on the sensitivity of this EFT to Planck-scale corrections. In §3.3, we review the EFT of inflationary perturbations developed by Cheung et al. [36].

3.1 EFT philosophy

In this section, we describe the basic philosophy of EFT from the two perspectives: the top-down and bottom-up approaches.

Top down. Imagine performing experiments at an energy scale E. Let the full theory consist of a heavy and a light field labelled by Ψ and ϕ, with masses $M > E$ and $m < E$, respectively. The Lagrangian of the theory can be expressed as

$$\mathcal{L}[\Psi, \phi] = \mathcal{L}_\Psi[\Psi] + \mathcal{L}_\phi[\phi] + \mathcal{L}_{\rm mix}[\Psi, \phi]\,, \tag{3.1}$$

where $\mathcal{L}_\Psi$ and $\mathcal{L}_\phi$ denote renormalizable terms for the individual fields, and $\mathcal{L}_{\rm mix}$ includes interactions between the two fields. We would like to describe the dynamics of this theory using only the low-energy degrees of freedom that are active at $E < M$. The *Wilsonian effective action*, $S_{\rm eff}$, is obtained by performing the following functional integral over the heavy field (and the high-frequency modes of the light field):

$$e^{iS_{\rm eff}[\phi]} = \int \mathcal{D}\Psi\, e^{iS[\Psi,\phi]}\,. \tag{3.2}$$

In this case, the heavy field is said to have been *integrated out*. This effective action describes the low-energy dynamics of the theory, which incorporates the effects of the heavy modes above E. The functional integral is often evaluated using a saddle-point approximation, which is equivalent to replacing Ψ by its classical solution. The resulting effective action will formally be non-local, but becomes (approximately) local, once expanded in inverse powers of the heavy scale.

To better illustrate these abstract concepts, consider a toy example with the following two-field Lagrangian for the full theory:

$$\mathcal{L}[\Psi, \phi] = -\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{1}{2}(\partial\Psi)^2 - \frac{1}{2}M^2\Psi^2 + g\phi^2\Psi\,, \tag{3.3}$$

where g is a dimensionful coupling constant. We expect the heavy field to be approximately non-dynamical at low energies, $E \ll M$, meaning that to leading order we can use the classical equations of motion for Ψ, namely $(-\Box + M^2)\Psi = g\phi^2$, as a constraint. When the classical solution of Ψ is substituted back to the above Lagrangian, it will yield an effective Lagrangian for ϕ only

$$\mathcal{L}_{\rm eff}[\phi] = -\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 + \frac{1}{2}g^2\phi^2\frac{1}{-\Box + M^2}\phi^2\,. \tag{3.4}$$

Note that formally this Lagrangian contains a non-local term. However, at low energies, we can

take an expansion in powers of $1/M$, which gives a series of local operators

$$\mathcal{L}_{\text{eff}}[\phi] = -\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 + \frac{1}{2}\frac{g^2\phi^2}{M^2}\sum_{n=0}^{\infty}\left(\frac{\Box}{M^2}\right)^n \phi^2 \,. \tag{3.5}$$

As long as the derivative expansion is valid, the infinite series of higher-derivative terms can be safely truncated, so that the effective Lagrangian becomes local. At which order one truncates the series depends on the required precision.

Bottom up. It is often impossible to derive an effective Lagrangian in the above manner from the top down, either due to the intricacy of the calculation or because the UV theory is unknown. In those cases, we can parameterize our ignorance about the UV dynamics and construct an EFT from the bottom up. After identifying the relevant degrees of freedom (in this case ϕ) and the symmetries in question (e.g. Lorentz invariance, internal symmetries), we write down the most general effective Lagrangian consistent with these symmetries

$$\mathcal{L}_{\text{eff}}[\phi] = \mathcal{L}_{\delta\leq 4}[\phi] \; + \sum_i c_i \, \frac{\mathcal{O}_i[\phi]}{\Lambda^{\delta_i - 4}} \,, \tag{3.6}$$

where $\mathcal{O}_i$ is a set of local operators of mass dimension $\delta_i > 4$. The part of the Lagrangian denoted $\mathcal{L}_{\delta\leq 4}$ contains a finite number of renormalizable interactions, including the kinetic and mass terms. The rest of the Lagrangian consists of an infinite series of higher-dimensional, or non-renormalizable, interactions. These are weighted by inverse powers of a cutoff scale, Λ, which is assumed to be higher than the typical energy scales of interest. In this bottom-up construction of an EFT, the *Wilson coefficients* c_i are free parameters that need to be fixed by experiments, but we typically expect these coefficients to be of order unity.[2] Naively, the presence of an infinite number of terms seems to suggest that theoretical predictability is lost. However, because only a *finite* number of non-renormalizable interactions have the same dimension, we can truncate the series and make predictions with a finite precision.

The algorithm for constructing an EFT from the bottom up can be summarized as follows: (i) identify the physical degrees of freedom and symmetries relevant at a chosen energy scale, (ii) write down the most general effective Lagrangian consistent with these symmetries up to some chosen order, and (iii) fix the parameters of the EFT by comparing with experimental data. We will later apply these steps to construct effective theories of inflation.

Naturalness. Having light scalar fields in an EFT seems problematic, once quantum corrections are included.[3] By dimensional analysis, the prefactor of an operator $\mathcal{O}_i$ receives radiative corrections proportional to $\Lambda^{4-\delta_i}$. This implies that the mass parameter will receive a correction,

[2] We use the term "order unity" in the loose sense, referring to some small range around unity, say, from 10^{-2} to 10^2. For order one Wilson coefficients, we also assume that dimensionless parameters in the UV theory are also of order one.

[3] We will focus on the naturalness issue of light scalar fields, but the following discussion applies more generally to relevant operators whose mass dimensions are less than four.

$\Delta m^2 \sim \Lambda^2$, that is quadratically dependent on the cutoff.[4] In other words, quantum corrections tend to drive the masses of scalar fields towards the scale of the heavy physics that has been integrated out. This is problematic, since it violates the starting assumption for having an EFT, namely the presence of a hierarchy of scales. The apparent sensitivity of the masses of scalar fields to UV physics is a *naturalness problem*. Symmetries can help to protect the masses of scalar fields against large quantum corrections, while the absence of such symmetries invokes a *fine-tuning* of parameters.[5]

3.2 Effective theory of slow-roll inflation

The observed size of the CMB temperature fluctuations indicates that the physical wavenumber of the curvature perturbation during inflation had the value $k/a = H \sim \sqrt{\varepsilon} \times 10^{14}\,\mathrm{GeV}$ at horizon exit. Note that this is much lower than the (reduced) Planck scale $M_{\rm pl} \sim 10^{18}\,\mathrm{GeV}$, which is the natural scale at which quantum gravitational effects become important. However, new physics may appear at a scale much closer to the Hubble scale, in which case higher-dimensional operators may play an important phenomenological role. An EFT can be used to describe these interactions in a systematic manner.

In this section, we introduce an EFT for the inflationary background. We take the relevant low-energy degree of freedom to be a real scalar field ϕ, the inflaton. The effective Lagrangian for the inflaton will have the structure

$$\mathcal{L}_{\rm eff}[\phi] = \mathcal{L}_{\delta\leq 4}[\phi] + \sum_{n=1}^{\infty}\left(c_n\frac{\phi^{4+n}}{\Lambda^n} + d_n\frac{(\partial\phi)^2\phi^n}{\Lambda^n} + \cdots\right), \tag{3.7}$$

where c_n and d_n are Wilson coefficients and the ellipses denote higher-derivative terms. We first describe the UV sensitivity of the inflationary observables in the context of this EFT. We then generalize the EFT by allowing for couplings to curvature tensors and consider the leading derivative interactions, following the work of Weinberg [46].

3.2.1 Ultraviolet sensitivity

Inflation is distinct from many other models in particle physics, in that it intrinsically involves gravitational dynamics. An unusual feature of low-energy inflationary observables is that they can be extremely sensitive to high-scale physics through higher-dimensional operators. Some of these operators are relevant even when they are suppressed by the Planck scale—a phenomenon known as the *UV sensitivity* of inflation. We illustrate this issue with two examples below.

[4] In fact, the quadratic divergence in Λ carries no physical meaning as it depends on the regularization scheme. For example, power law dependences on the cutoff do not appear in dimensional regularization. However, a physical scale M (e.g. the mass of a heavy particle in the UV theory) in general will still lead to a contribution $\Delta m^2 \sim M^2$ to the renormalized mass parameter. We therefore use the unphysical Λ^2 divergence as a proxy for the expected physical dependence on M^2.

[5] Fields with non-zero spin do not receive quadratically divergent corrections, as their masses are protected by symmetries, namely gauge invariance or chiral symmetry.

Eta problem. Consider a dimension-6 operator in the effective Lagrangian of the form

$$\mathcal{L}_{\text{eff}} \supset cV_0(\phi)\frac{\phi^2}{\Lambda^2}\,, \tag{3.8}$$

where c is a dimensionless coefficient, $V_0(\phi)$ denotes the renormalizable part of the inflaton potential, and $\Lambda \gg H$ is the cutoff. This operator does not violate any obvious symmetry and thus should be present in a generic EFT.[6] The eta parameter, η_V, defined in (2.23) then receives the following correction:

$$\Delta\eta_V \approx 2c\frac{M_{\text{pl}}^2}{\Lambda^2}\,. \tag{3.9}$$

Since $\Lambda \leq M_{\text{pl}}$ and c is typically of order one, we find $\Delta\eta_V \gtrsim 1$. This is problematic as it violates one of the inflationary flatness conditions, $\eta_V \ll 1$. In particular, we see that the problem remains even when we take $\Lambda \to M_{\text{pl}}$. This striking sensitivity of the eta parameter to Planck-scale physics is called the *eta problem*.[7]

As with other fine-tuning problems, the eta problem is not a fundamental problem of inflation itself. Rather, it highlights the lack of our current understanding of the microscopic theory underlying inflation, such as its symmetry structure. For example, imposing a shift symmetry of the inflaton, $\phi \to \phi + \text{const.}$, will ensure that dangerous terms which can give rise to large corrections to the eta parameter are absent in the effective Lagrangian. Having said that, it is also believed that continuous global symmetries cannot exist in quantum gravity [48]. This implies that such a shift symmetry cannot be an exact symmetry and must be broken by non-perturbative gravitational effects. It is currently unclear what the typical size of these symmetry-breaking effects are. Addressing the eta problem therefore requires an advancement in our understanding of quantum gravity and a proper quantification of these non-perturbative effects (although see [49]).

Large-field inflation. Another prominent problem arises in models of *large-field* inflation, for which the inflaton field undergoes a super-Planckian excursion during inflation, $\Delta\phi > M_{\text{pl}}$. A

[6] For example, such an operator appear in generic supergravity models of inflation [47].

[7] The radiative stability of the inflaton mass also leads to an eta problem. As we have seen, radiative corrections will tend to drive the inflaton mass towards the cutoff, $\Delta m^2 \sim \Lambda^2$. This in turn lead to the following large correction to the eta parameter:

$$\Delta\eta_V = \frac{\Lambda^2}{3H^2} > 1\,. \tag{3.10}$$

Unless protected by some symmetry, the mass of the inflaton need be fine-tuned to ensure that inflation lasts sufficiently long. For example, supersymmetry can partially address the issue. Although necessarily broken during inflation due to the positive vacuum energy, supersymmetry still leads to the cancellation between the quantum contributions from bosons and fermions at high energies above the Hubble scale. As a consequence, the inflaton gains a mass of order H or $\Delta\eta_V \sim 1$, meaning that a relatively minor fine-tuning of a part in 100 is required to solve the eta problem.

lower bound on the inflaton field excursion is given by the so-called *Lyth bound* [50]

$$\frac{\Delta\phi}{M_{\rm pl}} \gtrsim N\sqrt{\frac{r}{8}} \approx \left(\frac{r}{10^{-3}}\right)^{1/2}\left(\frac{N}{10}\right), \tag{3.11}$$

where $60 > N > 40$ is the number of e-folds between the time observable scales exit the horizon and the end of inflation. The tensor-to-scalar ratio, $r = 16\varepsilon$, was assumed to be nearly constant during inflation. Observable tensor modes, $r \gtrsim 10^{-3}$, therefore implies a super-Planckian field excursion, $\Delta\phi \gtrsim M_{\rm pl}$. To see the ramification of this, consider a generic form of the inflaton potential

$$V(\phi) = V_0(\phi) + \sum_{n=1}^{\infty} c_n \frac{\phi^{4+n}}{\Lambda^n}, \tag{3.12}$$

which is a subset of the effective Lagrangian (3.7) involving non-derivative terms. We see that if the inflaton has a super-Planckian field value, then the perturbative expansion becomes no longer valid, even when $\Lambda \to M_{\rm pl}$. Fine-tuning is not an option this time, since it would require tuning of an infinite number of parameters. Imposing shift symmetry is therefore the only possibility to ensure that these dangerous interactions are not present. As with the eta problem, building a consistent UV-completion of large-field inflation is a challenge in inflationary model-building. Attempts to address this in string theory are described in [51].

3.2.2 Higher-curvature action

Besides self-interactions of the inflaton, there can also be contributions to the effective theory of inflation from gravitational corrections. Any gravitational theory should reduce to Einstein gravity at low energies, but new gravitational physics may enter at a scale much below the Planck scale (e.g. the string scale in string theory). How low, then, could the scale of new physics be, while an effective description of inflation remains valid? For a valid perturbative expansion, the cutoff scale Λ must be larger than any physical scale that is intrinsic to the low-energy system. During inflation, the inflaton acquires a time-dependent background value, with $\dot{\bar{\phi}} = \sqrt{2\varepsilon}HM_{\rm pl}$, which suggests that the cutoff should be larger than this scale, $\Lambda^2 \gg \dot{\bar{\phi}}$.[8] This will ensure that, for instance, the dimension-8 operator $(\partial\phi)^4/\Lambda^4$ leads to a perturbative correction to the kinetic term for the inflaton perturbation, $\dot{\bar{\phi}}^2(\partial\delta\phi)^2/\Lambda^4 \ll (\partial\delta\phi)^2$.

Assuming that the cutoff scale is above $\dot{\bar{\phi}}$, so that we can use an effective treatment of the background field, let us write down the leading corrections to the slow-roll action (2.19). This

[8] As we will see in the next section, in the EFT of inflationary perturbations we consider a derivative expansion, not of the background inflaton field in spatially flat gauge, but of the metric fluctuations in comoving gauge. The power counting then changes dramatically, and the cutoff can be made lower than $\dot{\bar{\phi}}$.

was systematically studied by Weinberg in [46] and is given by

$$\Delta S = \int \mathrm{d}^4x\sqrt{-g}\left[f_1(\varphi)(\partial\varphi)^4 + f_2(\varphi)(\partial\varphi)^2\Box\varphi + f_3(\varphi)(\Box\varphi)^2 + f_4(\varphi)R^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi \right.$$
$$\left. + f_5(\varphi)R(\partial\varphi)^2 + f_6(\varphi)R\Box\varphi + f_7(\varphi)R^2 + f_8(\varphi)R^{\mu\nu}R_{\mu\nu} + f_9(\varphi)W^2 + f_{10}(\varphi)W\widetilde{W}\right], \quad (3.13)$$

where we have introduced the dimensionless field $\varphi \equiv \phi/\Lambda$. We have decided to express the last two terms in the action (3.13) in terms of the Weyl tensor

$$W_{\mu\nu\rho\sigma} \equiv R_{\mu\nu\rho\sigma} - \frac{1}{2}\left(g_{\mu\rho}R_{\nu\sigma} - g_{\mu\sigma}R_{\nu\rho} - g_{\nu\rho}R_{\mu\sigma} + g_{\nu\sigma}R_{\mu\rho}\right) + \frac{R}{6}(g_{\mu\rho}g_{\nu\sigma} - g_{\nu\rho}g_{\mu\sigma}) \ . \quad (3.14)$$

This can always be done by redefining other coefficients, since

$$W^2 \equiv W^{\mu\nu\,\rho\sigma}W_{\mu\nu\,\rho\sigma} = R^{\mu\nu\,\rho\sigma}R_{\mu\nu\,\rho\sigma} - 2R^{\mu\nu}R_{\mu\nu} + \frac{1}{3}R^2 \ , \quad (3.15)$$

and similarly for the parity-violating term $W\widetilde{W} \equiv (\sqrt{-g})^{-1/2}\epsilon^{\mu\nu\,\rho\sigma}W_{\mu\nu}{}^{\kappa\lambda}W_{\rho\sigma\,\kappa\lambda}$.

Due to the presence of higher-derivative terms, the full action will inevitably contain ghost degrees of freedom with negative kinetic terms, corresponding to poles with wrong signs. However, we emphasize that this form of the effective action is valid for inflationary models with $\Lambda > \dot{\bar{\varphi}}$, i.e. when ΔS is treated as a perturbative correction to the leading action given in (2.19). This means that we can use the field equations from the leading action, and the wrong-sign poles will arise for energy scales greater than Λ, beyond the validity of the effective action. The leading-order field equations are

$$\Lambda^2\Box\varphi = U'(\varphi)\,, \quad M_{\rm pl}^2 R_{\mu\nu} = -\Lambda^2\partial_\mu\varphi\partial_\nu\varphi - U(\varphi)g_{\mu\nu}\,, \quad (3.16)$$

with $U(\varphi) \equiv V(\Lambda\varphi)$. We can further use field redefinitions to simplify the action. Physical observables such as scattering amplitudes or correlation functions are not affected by such changes of the field variables. This is because local non-linear field redefinitions—in the form of $\varphi \to \varphi + F(\varphi)$, where the function F involves at least two powers of the φ field—do not affect the quadratic action and thus leave the asymptotic states intact. After using the field equations and suitable field redefinitions, one arrives at the simple action [46]

$$\Delta S = \int \mathrm{d}^4x\sqrt{-g}\left[f_1(\varphi)(\partial\varphi)^4 + f_9(\varphi)W^2 + f_{10}(\varphi)W\widetilde{W}\right], \quad (3.17)$$

where the effects of other higher-curvature terms have been absorbed into the inflaton potential. The first term in the above action is a higher-derivative self-interaction of the inflaton that appears e.g. in models of k-inflation [52]. The Weyl tensors vanish in FLRW backgrounds, so the corrections from the last two terms start at quadratic order in fluctuations. We will analyze the perturbations of this action in Chapter 6.

3.3 Effective theory of cosmological perturbations

We now turn our attention to a general framework of describing cosmological perturbations. We focus on the light degree of freedom that arises because the time dependence of the inflationary background breaks some of the symmetries of the idealized de Sitter limit. The effective theory for the Goldstone boson corresponding to this symmetry breaking provides a systematic way to characterize the dynamics of inflationary fluctuations. The key advantage of this approach is that it allows us to remain agnostic about the mechanism that creates the inflationary background.

The outline of the section is as follows. We first describe the basic concepts of spontaneous symmetry breaking and their applications to inflation. We then review the EFT of the Goldstone boson of broken time translations—often just referred to as the *EFT of inflation*—which was originally developed in [36, 53] (see [51, 54] for recent reviews).

3.3.1 Spontaneous symmetry breaking

If the symmetries of the action do not leave the ground state of the theory invariant, then we say that these symmetries are spontaneously broken. For systems with spontaneously broken symmetries, there is an immensely useful way of describing the low-energy dynamics of the system. This makes use of an EFT in the broken symmetry phase, whose construction relies solely on the symmetry breaking pattern without the need to know the mechanism responsible for the breaking.

Internal symmetries. Consider a theory consisting of a set of scalar fields collectively denoted by Φ, whose action is invariant under an internal symmetry group G. Group elements act on the fields by

$$\Phi \;\to\; e^{i\theta^a T^a}\Phi\,, \tag{3.18}$$

where θ^a are parameters and T^a are the generators of G, with $a = 1, ..., \dim G$. Suppose that the fields acquire vacuum expectation values, $\langle\Phi\rangle \equiv \bar{\Phi}$, which are only invariant under the subgroup $K \subset G$, but transform non-trivially under the coset G/K. In this case, we say that the symmetry group G is spontaneously broken to K, and the low-energy action is determined purely by the symmetry breaking pattern. Goldstone's theorem states that spontaneous symmetry breaking leads to the excitation of massless scalar particles called *Goldstone bosons* [55, 56]. There are in total $\dim G/K$ Goldstone bosons, one for each broken symmetry generator.[9]

Under the action of the coset, a vacuum state gets transformed into another degenerate vacuum state. This implies the presence of flat directions in the field space, along which the energy of the state stays constant. A coset element is represented by

$$U(x) \equiv e^{i\pi^a(x)\tau^a}\,, \tag{3.19}$$

[9]This is true for relativistic systems. The counting of Goldstone bosons is more subtle for non-relativistic systems (see e.g. [57]).

where π^a are the Goldstone bosons that parameterize the space of degenerate vacuum states and τ^a are the broken symmetry generators. The unbroken symmetries acts linearly on the Goldstone bosons, while the broken symmetries act non-linearly.[10] In particular, the non-linear transformation involves a constant shift at leading order. This implies that the Goldstone bosons are massless and that they only involve derivative interactions. Since the excitations along the non-flat directions are massive, the dynamics of the Goldstone bosons are decoupled at sufficiently low energies.

To describe the low-energy degrees of freedom in the broken symmetry phase, we write down the most general Lagrangian that is invariant under G and built out of the coset element U defined in (3.19):

$$\begin{aligned}\mathcal{L}_{\text{eff}} \;=\; & -\frac{f_\pi^2}{4}\mathrm{Tr}\big(\partial_\mu U\partial^\mu U^\dagger\big) + c_1\mathrm{Tr}\big(\partial_\mu U\partial^\mu U^\dagger\partial_\nu U\partial^\nu U^\dagger\big) \\ & + c_2\mathrm{Tr}\big(\partial_\mu U\partial_\nu U^\dagger\big)\mathrm{Tr}\big(\partial^\mu U\partial^\nu U^\dagger\big) + c_3\Big[\mathrm{Tr}\big(\partial_\mu U\partial^\mu U^\dagger\big)\Big]^2 + \cdots\,,\end{aligned} \tag{3.20}$$

where c_i are dimensionless constants and $\mathrm{Tr}(\cdots)$ denotes a trace. We have omitted terms with more than four derivatives and with more than single derivative acting on U. The parameter f_π with dimension of mass is included on dimensional grounds. The Goldstone Lagrangian can be obtained by expanding in π. For example, the two-derivative term gives

$$-\frac{f_\pi^2}{4}\mathrm{Tr}\big(\partial_\mu U\partial^\mu U^\dagger\big) = -\frac{1}{2}\partial_\mu\pi_c^a\partial^\mu\pi_c^a - \frac{1}{6f_\pi^2}\big[(\pi_c^a\partial_\mu\pi_c^a)^2 - \pi_c^2(\partial_\mu\pi_c^a\partial^\mu\pi_c^a)\big] + \cdots\,, \tag{3.21}$$

where $\pi_c^a \equiv f_\pi\pi^a$ are the canonically normalized fields.[11] We see that two-derivative interactions of π^a are completely fixed by the condition that π^a must transform non-linearly under the broken symmetries. After canonical normalization, these interactions consist of a series of non-renormalizable terms suppressed by powers of f_π. This scale is fixed by the vacuum expectation value of the original theory and is called the *symmetry breaking scale*.[12] At energies above f_π, the perturbative expansion in terms of the Goldstone boson breaks down, and the symmetries get restored. By dimensional analysis, the perturbative expansion for the four-derivative terms in (3.20) will break down at $E \sim \Lambda$, where $\Lambda^4 \equiv f_\pi^4/c_i$, which is below f_π for $c_i > 1$. This means that the inclusion of higher-derivative terms can lower the scale at which the perturbative expansion breaks down.

Spacetime symmetries. For cosmological applications, we will be interested in spontaneously broken spacetime symmetries. While most of the concepts discussed above for internal symmetries also apply for spacetime symmetries, there is also a crucial difference. For spontaneous spacetime symmetry breaking, the spectrum of Goldstone modes is more subtle, and the number of broken

[10] The Goldstone bosons are said to transform linearly if $\pi^a \to \pi'^a = A^a{}_b\pi^b$ for some constant matrix $A^a{}_b$. Otherwise, they are said to transform non-linearly.

[11] Our normalization convention is $\mathrm{Tr}(\tau^a\tau^b) = \delta^{ab}/2$.

[12] For instance, the scale f_π plays the role of the pion decay constant in the Chiral Lagrangian [58].

symmetry generators in general does not match the number of Goldstone bosons [59]. This is because the generators of local spacetime symmetries can be linearly dependent on each other, so that some of them may not give rise to independent Goldstone bosons. A classic example is phonons in solids. In that case, there are three Goldstone bosons although six isometries (three translations and three rotations) are spontaneously broken.

Let us explicitly demonstrate the relationship between broken spacetime symmetries. For example, consider a scalar field ϕ with a spacetime-dependent vacuum expectation value

$$\langle\phi\rangle = c_\mu x^\mu \,, \tag{3.22}$$

where c_μ is a constant vector. Due to the explicit appearance of the spacetime coordinate x^μ, this configuration spontaneously breaks diffeomorphism invariance. Note that we can also express the above as $\langle\partial_\mu\phi\rangle = c_\mu$. Let us introduce a vierbein e^a_μ, whose indices, $a = 0, \ldots, 3$, are raised and lowered by a local Minkowski metric η_{ab}. Using the vierbein, the components of $\partial_\mu\phi$ in a coordinate basis can then be related to the components of $\partial_a\phi$ in a local Lorentz frame

$$\langle\partial_a\phi\rangle = e^\mu_a\langle\partial_\mu\phi\rangle = c_\mu e^\mu_a \,. \tag{3.23}$$

This shows that spontaneously broken diffeomorphism invariance implies spontaneously broken local Lorentz invariance (and vice versa). The same argument applies to arbitrary tensors $T_{\mu\nu\ldots}$ with constant vacuum expectation values.

3.3.2 Symmetries in inflation

Inflation is described by a quasi-de Sitter phase, so it does not retain all of the isometries of de Sitter space. Before developing an effective theory describing the symmetry breaking, we review the symmetries of de Sitter space and show how they are broken during inflation.

de Sitter isometries. De Sitter space is a maximally symmetric vacuum solution to Einstein gravity with a positive cosmological constant. The corresponding line element is

$$\mathrm{d}s^2 = \frac{1}{H^2\eta^2}\left(-\mathrm{d}\eta^2 + \mathrm{d}\mathbf{x}^2\right), \tag{3.24}$$

where $a(\eta) = -1/H\eta$ and the range of conformal time is $\eta \in (-\infty, 0)$. Inspection of (3.24) reveals the isometries of the de Sitter spacetime: in addition to spatial rotations and translations, these include a dilatation (D) and three special conformal transformations (SCTs)

$$\begin{aligned} \text{D}: &\quad \eta \to (1+\lambda)\eta\,, &&\quad \mathbf{x} \to (1+\lambda)\mathbf{x}\,, \\ \text{SCTs}: &\quad \eta \to (1-2\mathbf{b}\cdot\mathbf{x})\eta\,, &&\quad \mathbf{x} \to \mathbf{x} - 2(\mathbf{b}\cdot\mathbf{x})\mathbf{x} + (\mathbf{x}^2 - \eta^2)\mathbf{b}\,, \end{aligned} \tag{3.25}$$

where λ and $\mathbf{b}$ are infinitesimal parameters. In total, de Sitter space therefore has 10 isometries constituting the isometry group SO(1,4). Correspondingly, de Sitter space possesses 10 Killing

vectors or generators of the isometry group.[13] Of particular interest is the Killing vector for the dilatation, $\xi^\mu = (-\eta, \mathbf{x})$, which is locally timelike:

$$\xi^\mu \xi_\mu = \frac{1}{H^2\eta^2}(-\eta^2 + \mathbf{x}^2) < 0 \quad \text{for} \quad a|\mathbf{x}| < H^{-1}\,. \tag{3.26}$$

We can therefore think of de Sitter space as being time-translation symmetric within a finite causal patch of radius H^{-1}.

Symmetry breaking in inflation. Because inflation needs to end, it cannot be described by a pure de Sitter space. There must be a physical "clock" that keeps track of the remaining period of inflation, which can be associated e.g. with a set of time-dependent homogeneous background matter fields $\bar{\phi}_m$ (e.g. $\bar{\phi}_m = \bar{\phi}$ for single-field slow-roll inflation) or the energy density of the background.[14] This time dependence induces a preferred time slicing of the spacetime in terms of constant-$\bar{\phi}_m$ hypersurfaces and breaks time diffeomorphism invariance (see Fig. 3.1). We therefore expect the existence of a corresponding Goldstone boson, π, that encodes fluctuations along the direction of the broken symmetry. This is introduced as a spacetime-dependent shift of the time coordinate [53]

$$U(t, \mathbf{x}) \equiv t + \pi(t, \mathbf{x})\,. \tag{3.27}$$

The Goldstone boson is defined so that it transforms non-linearly $\pi \to \pi - \xi(t, \mathbf{x})$ under a time diffeomorphism $t \to t + \xi(t, \mathbf{x})$. The action expressed in terms of U will thus be invariant under full spacetime diffeomorphisms, thus "restoring" the broken symmetry.[15] An important property of the field π is that it can be identified with so-called *adiabatic perturbations*, i.e. with perturbations induced by a synchronized, spacetime-dependent local time shift of the background matter fields:

$$\delta\phi_m(t, \mathbf{x}) \equiv \bar{\phi}_m\big(t + \pi(t, \mathbf{x})\big) - \bar{\phi}_m(t)\,. \tag{3.28}$$

In spatially flat gauge, the Goldstone π captures all the information contained in the cosmological perturbations.

Alternatively, we can think of the symmetry breaking in terms of the isometry group of the background spacetime. The symmetry breaking pattern that leads to inflation is

$$\mathrm{SO}(1,4) \quad \to \quad \mathrm{ISO}(3)\,, \tag{3.29}$$

[13] The vector ξ^μ is a Killing vector if the Lie derivative of the metric with respect to ξ^μ vanishes, which amounts to the Killing equation $\nabla_{(\mu}\xi_{\nu)} = 0$.

[14] One does not need to postulate a fundamental scalar field for this purpose. For example, it could be some strongly coupled dynamics that give rise to the time dependence during inflation.

[15] Like gauge invariance, the diffeomorphism group of general relativity consists of local transformations and hence is not a true symmetry—instead, it reflects a redundancy in our description. We can always reintroduce local symmetries by explicitly introducing new degrees of freedom.

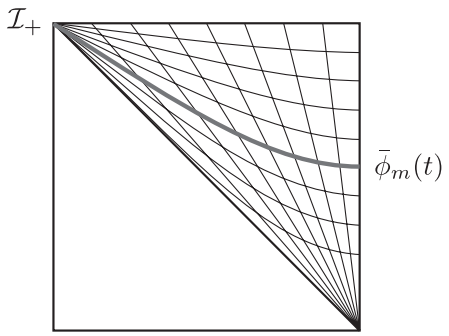


Figure 3.1: Penrose diagram for de Sitter space. The spacelike future boundary at $\eta = 0$ is denoted by $\mathcal{I}_+$. Time-dependent background fields, $\bar{\phi}_m(t)$, induce a preferred time foliation of de Sitter space.

where ISO(3) is the 3-dimensional Euclidean group consisting of spatial rotations and translations. This spontaneously breaks four isometries of de Sitter space: dilatation and three SCTs. To see this, let us perform a time reparameterization $t \to t - \pi(t, \mathbf{x})$ to go to *unitary gauge*, in which the clock is set to be homogeneous ($\delta\phi_m = 0$) and the information is now captured by the curvature perturbation, ζ. (This equivalent to comoving gauge (2.31), where the clock was chosen to be the background inflaton field.) While spatial rotations and translations are linearly realized by ζ, under infinitesimal spatial dilatation and SCTs (3.25), ζ transforms as [26, 60, 61]

$$\begin{aligned} \text{D}: &\quad \zeta \to \lambda(1 + \mathbf{x}\cdot\nabla\zeta)\,, \\ \text{SCTs}: &\quad \zeta \to 2\mathbf{b}\cdot\mathbf{x} + \Big[2(\mathbf{b}\cdot\mathbf{x})\mathbf{x} - \mathbf{x}^2\mathbf{b}\Big]\cdot\nabla\zeta\,. \end{aligned} \tag{3.30}$$

Notice the ζ-independent shifts under these transformations. In this gauge, ζ acts as the Goldstone boson that non-linearly realizes the broken symmetries during inflation.

Both of the above descriptions are useful for different purposes. For example, the parameterization of the broken dilatation and SCTs in terms of the field ζ has been particularly useful in deriving the behavior of the soft limits of inflationary correlation functions [26, 60, 61]. In the following, we will instead mostly use the language of the Goldstone boson π, which is more appropriate for formulating an EFT. The relation between π in spatially flat gauge and ζ in unitary gauge is given by

$$\zeta = -H\pi + \mathcal{O}(\pi^2)\,. \tag{3.31}$$

The non-linear part of the relation contributes to correlation functions at subleading order in the slow-roll approximation and can therefore be neglected [62].[16] Since H is nearly constant during inflation, the two fields are approximately proportional to each other, which allows us to easily translate results from one gauge to the other.

[16] This is true when the Bunch-Davies initial condition is imposed, which we do throughout the thesis. For generic excited initial states, the non-linear relation between π and ζ may produce non-trivial effects on correlation functions [63].

3.3.3 EFT of inflation

We would like to derive the action for the Goldstone boson of time translations introduced above. One way to proceed is by using the method outlined in §3.3.1, that is by writing down the most general action in terms $U = t + \pi$ and then expanding in π. In this subsection, we instead construct the EFT by starting from a more geometric picture, following the work of [53, 62]. We write down the effective actions for both the Goldstone boson and the graviton.

Unitary gauge. In unitary gauge, the time coordinate t itself is chosen to be the clock, and fluctuations in the clock are eaten by the metric. The action is then no longer invariant under full diffeomorphism invariance, but only under time-dependent spatial diffeomorphisms, $x^i \to x^i + \xi^i(t, \mathbf{x})$. The time slicing induced by the clock has a timelike gradient, and the unit vector perpendicular to the surface of constant t is

$$n_\mu \equiv -\frac{\delta^0_\mu}{\sqrt{-g^{00}}}\,. \tag{3.32}$$

The induced spatial metric on the slicing is $h_{\mu\nu} \equiv g_{\mu\nu} + n_\mu n_\nu$. Geometric objects living on the hypersurfaces can be constructed from $h_{\mu\nu}$ and n_μ. Examples are the intrinsic curvature, ${}^{(3)}R_{\mu\nu\rho\sigma}[h]$, and the extrinsic curvature, $K_{\mu\nu} \equiv h_{(\mu}{}^{\rho}\nabla_\rho n_{\nu)}$. Using the Gauss-Codazzi relation, the intrinsic curvature can be written in terms of (the projection of) the four-dimensional Riemann tensor $R_{\mu\nu\rho\sigma}$ and the extrinsic curvature $K_{\mu\nu}$.

Following the EFT philosophy, we construct an effective action by writing down all operators that are compatible with the remaining symmetries. The reduced symmetry of the system now allows many new terms in the action, which can be categorized as follows:

- Terms that are invariant under all diffeomorphisms. These include curvature invariants like the Ricci scalar, R, and contractions of Riemann tensors such as $R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$.
- Operators with uncontracted upper 0 indices, such as g^{00} and R^{00}. These are generated by contracting covariant tensors with n_μ. It is easy to check that these are scalars under spatial diffeomorphisms. In general, products of any four-dimensional covariant tensors with free upper 0 indices are allowed operators, while all spatial indices need to be contracted.
- Operators made out of the three-dimensional quantities describing the geometry of the spatial hypersurfaces. Since these are related to each other by the Gauss-Codazzi relation, we can choose to work with one of them (e.g. $K_{\mu\nu}$).
- Higher-derivative operators constructed by acting the covariant derivative ∇_μ on other terms.
- Arbitrary functions of time t. All operators are therefore allowed to have time-dependent coefficients.

The most general action constructed from these ingredients is [36]

$$S = \int \mathrm{d}^4x\sqrt{-g}\,\mathcal{L}(g^{00}, K_{\mu\nu}, R_{\mu\nu\rho\sigma}, \nabla_\mu, \dots, t)\,, \tag{3.33}$$

where the only free indices entering $\mathcal{L}$ are upper 0's. The spacetime indices are contracted with the four-dimensional metric $g_{\mu\nu}$. Terms involving explicit contractions of the induced metric $h_{\mu\nu}$ do not lead to new operators. Since terms involving curvature tensors are higher order in derivatives, the most general action at leading order can be written in terms of g^{00} alone,

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}M_{\rm pl}^2 R + M_{\rm pl}^2\dot{H}g^{00} - M_{\rm pl}^2(3H^2+\dot{H}) + \sum_{n=2}^{\infty}\frac{M_n^4(t)}{n!}(\delta g^{00})^n\right], \tag{3.34}$$

where $\delta g^{00} \equiv g^{00}+1$ and the time-dependent parameters M_n have mass dimension one. Instead of expanding in powers of g^{00}, we have expressed the action in such a way that only the first three terms contain linear perturbations of the metric, while the rest of the action involves perturbations that are explicitly quadratic order or higher. The coefficients of the operators 1 and g^{00} have been fixed by the requirement that we are expanding around the correct FLRW background with a given expansion rate $H(t)$. This removes all tadpoles, and the action starts quadratic in fluctuations. In the above language, slow-roll inflation corresponds to taking a special limit when $M_n \to 0$. To see this, notice that in unitary gauge, $\phi = \bar{\phi}(t)$, the slow-roll action becomes

$$\int \mathrm{d}^4x\sqrt{-g}\left[-\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right] = \int \mathrm{d}^4x\sqrt{-g}\left[-\frac{1}{2}\dot{\bar{\phi}}^2 g^{00} - V(\bar{\phi})\right]. \tag{3.35}$$

By using the relation $\dot{\bar{\phi}}^2 = -2M_{\rm pl}^2\dot{H}$ and the Friedmann equation $V(\bar{\phi}) = M_{\rm pl}^2(3H^2+\dot{H})$, we see that this indeed corresponds to the leading part of (3.34).

Goldstone action. To make the dynamics of the theory defined by (3.34) more transparent, we introduce the Goldstone boson, π, associated with the spontaneous breaking of time translations. Specifically, we perform a spacetime-dependent time reparameterization, $t \to t = t + \pi(t, \mathbf{x})$, and then promote π to a scalar field that non-linearly realizes time diffeomorphisms and restores full diffeomorphism invariance of the action. The metric transforms in the usual way—e.g.

$$g^{00} \to g^{00} + 2g^{0\mu}\partial_\mu\pi + g^{\mu\nu}\partial_\mu\pi\partial_\nu\pi\,. \tag{3.36}$$

Applying this procedure to (3.34) leads to the following action for the Goldstone boson at leading order in derivatives:

$$\begin{aligned} S = \int \mathrm{d}^4x\sqrt{-g}\Big[&\frac{1}{2}M_{\rm pl}^2 R + M_{\rm pl}^2\dot{H}(t+\pi)(g^{00} + 2g^{0\mu}\partial_\mu\pi + g^{\mu\nu}\partial_\mu\pi\partial_\nu\pi) \\ &- M_{\rm pl}^2\big(3H^2(t+\pi) + \dot{H}(t+\pi)\big) + \sum_{n=2}^{\infty}\frac{M_n^4(t+\pi)}{n!}(\delta g^{00} + 2g^{0\mu}\partial_\mu\pi + g^{\mu\nu}\partial_\mu\pi\partial_\nu\pi)^n\Big]. \end{aligned} \tag{3.37}$$

In general, this action involves a complicated mixing between the Goldstone mode and metric fluctuations. However, for most applications of interest, we can take the so-called decoupling limit ($M_{\rm pl} \to \infty$, $\dot{H} \to 0$, with $M_{\rm pl}^2\dot{H}$ fixed) and evaluate the Goldstone action in the unperturbed background [36], $g_{\mu\nu} \to \bar{g}_{\mu\nu}$.[17] In this case, the transformation (3.36) reduces to $g^{00} \to -1 - 2\dot{\pi} + \bar{g}^{\mu\nu}\partial_\mu\pi\partial_\nu\pi$, and the Goldstone Lagrangian becomes

$$\mathcal{L}_\pi = M_{\rm pl}^2\dot{H}(\partial_\mu\pi)^2 + 2M_2^4\left[\dot{\pi}^2 - a^{-2}\dot{\pi}(\partial_i\pi)^2\right] + \left(2M_2^4 - \frac{4}{3}M_3^4\right)\dot{\pi}^3 + \cdots . \tag{3.38}$$

We see that $M_2 \neq 0$ induces a nontrivial sound speed for the Goldstone boson,

$$c_s^2 \equiv \frac{M_{\rm pl}^2\dot{H}}{M_{\rm pl}^2\dot{H} - 2M_2^4} . \tag{3.39}$$

A small value of c_s (large value of M_2) is correlated with an enhanced cubic interaction $\dot{\pi}(\partial_i\pi)^2$ through a non-linearly realized symmetry. The Planck constraints on primordial non-Gaussianity imply $c_s \geq 0.024$ [35].

While the Goldstone boson is massless in the decoupling limit, it has a mass of order $\sqrt{\varepsilon}H$ when the mixing with the metric perturbations is taken into account. The small mass for π means that it evolves slightly outside the horizon. To describe observable quantities, it is more convenient to use curvature perturbation, ζ. The field ζ is exactly massless and becomes constant outside the horizon, even for large ε. Using the relation (3.31) between ζ and π, the power spectrum of ζ is found to be

$$\Delta_\zeta^2 = \frac{k^3}{2\pi^2}P_\zeta(k) = \frac{1}{4\pi^2}\left(\frac{H}{f_\pi}\right)^4 , \tag{3.40}$$

where $f_\pi^4 \equiv 2M_{\rm pl}^2|\dot{H}|c_s$ is the symmetry breaking scale [64], which is equal to $\dot{\bar{\phi}}^2$ in slow-roll inflation. The observed amplitude of the power spectrum $\Delta_\zeta^2 = (2.14 \pm 0.05) \times 10^{-9}$ [21], implies that $f_\pi \approx 59H$.

Graviton action. We can also systematically include terms that give rise to tensor perturbations. At quadratic order in γ_{ij}, the leading correction to the Einstein-Hilbert action can be written as

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}M_{\rm pl}^2R + \hat{M}_2^2\left(\delta K^{\mu\nu}\delta K_{\mu\nu} - \delta K^2\right)\right] . \tag{3.41}$$

The perturbed part of the extrinsic curvature contains

$$\delta K_j^i \supset \frac{1}{2}\dot{\gamma}_{ij} + \mathcal{O}(\gamma^2) . \tag{3.42}$$

The trace term δK^2 only contains scalar fluctuations, and the combination of extrinsic curvature

[17] It can be shown that the leading quadratic mixing between π and the metric perturbations can be neglected for energies above $E_{\rm mix} \equiv |\dot{H}|^{1/2} = \sqrt{\varepsilon}H$ (see [51] for a derivation). Since we are interested in computing observables at $E = H \gg E_{\rm mix}$, any corrections to taking the decoupling limit will be of order $E_{\rm mix}/H \sim \sqrt{\varepsilon}$ [36].

tensors in the above action was chosen in a way that doesn't modify the dispersion relation of the Goldstone boson π. Expanding the action in γ_{ij}, we find

$$\mathcal{L}_\gamma = \frac{M_{\rm pl}^2}{8}\frac{1}{c_t^2}\left[\dot{\gamma}_{ij}^2 - c_t^2 a^{-2}(\partial_k \gamma_{ij})^2\right] + \cdots , \tag{3.43}$$

where we have defined the tensor sound speed

$$c_t^2 \equiv \frac{M_{\rm pl}^2}{M_{\rm pl}^2 + 2\hat{M}_2^2} . \tag{3.44}$$

For sizable deviation of the tensor sound speed from unity, we require the scale $\hat{M}_2$ to be comparable to the Planck scale. The power spectrum of γ_{ij} is then given by

$$\Delta_\gamma^2 = \frac{k^3}{2\pi^2}P_\gamma(k) = \frac{2}{\pi^2}\frac{1}{c_t}\frac{H^2}{M_{\rm pl}^2} . \tag{3.45}$$

Compared to (2.41), this has an extra factor of $1/c_t$. We will describe the observational consequences of having a non-trivial tensor sound speed in Chapter 6.

4. Causality constraints

In the previous chapter, we reviewed how the study of cosmological perturbations in single-field inflation can be embedded in an EFT framework. This involved first identifying the symmetries of the problem and then systematically writing down all possible interactions that are consistent with these symmetries. However, even if EFTs constructed in this manner can be sensible theories on their own, it is known that not every EFT can be UV-completed into a full theory that respects basic physical principles, such as causality and Lorentz invariance [65].

Causality is one of the fundamental principles of any physical theory. Requiring the response of a system to be causal connects seemingly different phenomena, such as fluctuations and dissipation, or the speed and the attenuation of light in a medium. These connections are most manifest in frequency space, where causality is encoded in the analyticity of the response function. Non-trivial relations between physical observables are then simply a consequence of Cauchy's integral theorem, which relates the real and imaginary parts of the response function, as in the Kramers-Kronig relation. Similar considerations apply to scattering amplitudes: it is widely believed that (micro)causality is reflected in the analytic properties of the S-matrix. In this case, Cauchy's theorem provides a link between the low-energy, infrared (IR) limit of the scattering amplitude and its UV behavior. In this chapter, we use analyticity (causality) to derive analogous relations between cosmological observables and the underlying physics of inflation.

The outline of this chapter is as follows. In §4.1, we review the analytic properties of relativistic and non-relativistic scattering amplitudes. We derive a sum rule which relates the real part of the forward amplitude at low energies to an integral over its imaginary part. In §4.2, we assume positivity of the integral to derive constraints on a combination of the parameters of the EFT of inflation, including a consistency condition relating the quartic and cubic couplings. Moreover, armed with the full amplitude, we present an improved derivation of the critical sound speed for which the EFT admits a perturbative UV completion [42]. In §4.3, we explicitly demonstrate the validity of the sum rule for the weakly coupled completion of [66]. We show that the positivity constraints are satisfied, and argue that this is a generic feature of a large class of weakly coupled UV completions of the EFT of inflation. We also provide evidence for the conjecture that $c_s = 1$ is only compatible with slow-roll inflation. We discuss the observational implications of our results in §4.4. Technical details are relegated to Appendix A.

4.1 Analyticity and sum rules

In this section, we will review the standard analyticity arguments for relativistic scattering, see e.g. [67], and then discuss the additional subtleties that arise if the low-energy limit breaks Lorentz invariance. Some details of the discussion are relegated to Appendix A. In §4.2, we will apply the formalism to the EFT of inflation.

4.1.1 Relativistic scattering

For relativistic interactions, it is natural to consider the amplitude of $2 \to 2$ scattering to be a function of the Mandelstam variables s and t, i.e. $\mathcal{M}(s,t) \equiv \mathcal{M}(s,\theta(s,t))$, where θ is the scattering angle.[1] A minimal amount of non-analytic behavior of $\mathcal{M}(s,t)$ for complex s at fixed t is required by unitarity of the S-matrix: $SS^\dagger = 1$ [69]. In particular, for forward scattering, $t \to 0$, the *optical theorem* allows us to write the imaginary part of the amplitude as

$$2\,\mathrm{Im}[\mathcal{A}(s)] = \sum_I \int \mathrm{d}\Pi_I \, |\mathcal{M}(p_1, p_2 \to I)|^2 \,, \tag{4.4}$$

where I stands collectively for all possible intermediate states, each with a differential phase space element of $\mathrm{d}\Pi_I$. Using Hermitian analyticity, $\mathcal{A}(s) = \mathcal{A}^*(s^*)$, one may also write

$$\begin{aligned} 2i\,\mathrm{Im}[\mathcal{A}(s)] &\equiv \mathcal{A}(s+i\epsilon) - \mathcal{A}^*(s+i\epsilon) \\ &= \mathcal{A}(s+i\epsilon) - \mathcal{A}(s-i\epsilon) \equiv \mathrm{Disc}[\mathcal{A}(s)] \,, \end{aligned} \tag{4.5}$$

where $\mathrm{Disc}[\mathcal{A}(s)]$ denotes the discontinuity of $\mathcal{A}(s)$ across the real axis. The hypothesis of maximal analyticity[2] then assumes that $\mathcal{A}(s)$ is non-analytic *only* when $\mathrm{Im}[\mathcal{A}(s)] \neq 0$ along the real axis, i.e. when the right-hand side of (4.4) is non-zero above the mass thresholds for the states I. For the physical domain $s > 0$, this determines the locations of poles and branch cuts in terms of the energies of the states I. Moreover, the non-analytic behavior of $\mathcal{A}(s)$ for the unphysical values $s < 0$ is dictated by *crossing symmetry*. Specifically, there is a connection between the amplitude

[1] We follow the conventions of [68] and write the S-matrix as

$$\langle p_3 p_4 | S | p_1 p_2 \rangle = (2\pi)^4 \delta(p_1 + p_2 - p_3 - p_4)\big[1 + i\mathcal{M}(s,\theta)\big] \,, \tag{4.1}$$

where $p_a = (\omega_a, \mathbf{k}_a)$ denote four-momenta and $\cos\theta \equiv \hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3$ is the scattering angle. We denote the amplitude in the forward limit by

$$\mathcal{A}(s) \equiv \lim_{\theta \to 0} \mathcal{M}(s,\theta) \,. \tag{4.2}$$

The Mandelstam variables are defined by

$$s \equiv -(p_1 + p_2)^2 \,, \quad t \equiv -(p_1 - p_3)^2 \,, \quad u \equiv -(p_1 - p_4)^2 \,. \tag{4.3}$$

The variable s is the square of the center-of-mass energy and t is the square of the momentum transfer. The momentum conservation and on-shell condition imply that the three variables are not all independent, but related to each other by $s + t + u = \sum_{a=1}^4 m_a^2$.

[2] This hypothesis can be demonstrated in perturbation theory—see [69]. However, one cannot rule out the possibility of non-trivial analytic behavior due to non-perturbative physics (e.g. [70]).

at $s + i\epsilon$ (above the branch cuts) and that at $-s - i\epsilon$ (below the branch cuts), which may be shown to exist even for massless particles [71]. For identical particles, this implies that the forward amplitude is an even function, i.e. $\mathcal{A}(s) = \mathcal{A}(-s)$, and the singularities in the complex s-plane can all be accounted for in terms of s- and u-channel exchanges.

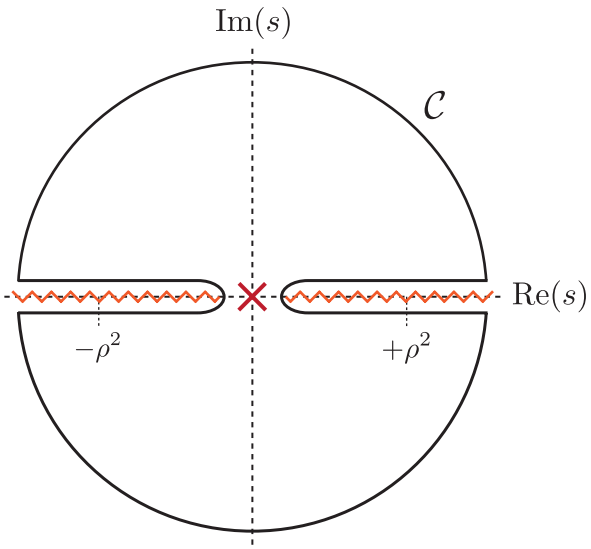


Figure 4.1: Illustration of the choice of contour in (4.6).

When the intermediate states I are massive, there is a gap between the singularities on the real axis. Considering the function $\mathcal{A}(s)/s^3$, Cauchy's theorem then implies the following sum rule

$$\frac{1}{2}\mathcal{A}''(s \to 0) = \oint_{\mathcal{C}} \frac{\mathrm{d}s}{2\pi i} \frac{\mathcal{A}(s)}{s^3}\,, \tag{4.6}$$

where $\mathcal{A}''(s) \equiv \partial_s^2 \mathcal{A}(s)$ and $\mathcal{C}$ is the contour illustrated in Fig. 4.1. The Froissart-Martin bound [72, 73], $|\mathcal{A}(s)| \leq \text{const.} \times s \ln^2 s$, for $|s| \to \infty$, lets us drop the contour at infinity, and only the discontinuities across the branch cuts, $\text{Disc}[\mathcal{A}(s)] \equiv 2i\,\text{Im}[\mathcal{A}(s)]$, contribute to the right-hand side of (4.6),

$$\mathcal{A}''(s \to 0) = \frac{2}{\pi}\left(\int_{-\infty}^{0} + \int_{0}^{\infty}\right) \mathrm{d}s\, \frac{\text{Im}[\mathcal{A}(s)]}{s^3}\,. \tag{4.7}$$

The Froissart-Martin bound may be violated when massless particles are present.[3] However, in our case, the sum rule in (4.7) still applies. As we shall see, this is because the forward scattering limit happens to be free of singularities,[4] which permits the application of standard techniques to

[3] In general the amplitude remains polynomially bounded on the physical sheet [70], so that an integral similar to (4.6) can be written for some n-th derivative of the amplitude.

[4] Let us remark that turning on gravitational interactions unavoidably induces divergences in the forward direction. While this is a general problem— gravity is universal—we expect the proper treatment of these effects to be highly suppressed in the cosmological setting, $H/M_{\rm pl} \ll 1$, and not to modify significantly the results derived

derive the number of necessary subtractions, e.g. [74].[5]

Using (4.4) to write the imaginary part of the amplitude in terms of the cross section, i.e. $\mathrm{Im}[\mathcal{A}(s)] \equiv s\sigma(s)$, and crossing symmetry which relates the integrals on the positive and negative axes, $\mathrm{Im}[\mathcal{A}(-s)] = -\mathrm{Im}[\mathcal{A}(s)]$, we get the sum rule in its final form:

$$\mathcal{A}''(s \to 0) = \frac{4}{\pi}\int_0^\infty \mathrm{d}s\, \frac{\sigma(s)}{s^2}\,. \tag{4.8}$$

The right-hand side of (4.8) is manifestly positive, which is a consequence of unitarity. Extensions to non-forward scattering are possible, even for unphysical values of t [74]. However, except for a speculative conjecture in §4.3.2, we will concentrate on forward scattering.

4.1.2 Non-relativistic scattering

We now consider the extension to non-relativistic scattering. We assume that the theory is Lorentz invariant in the UV, but allow for a non-trivial sound speed $c_s \neq 1$, as well as other Lorentz-symmetry breaking interactions, in the IR. For simplicity, we will work in the center-of-mass frame, where the forward amplitude $\mathcal{A}_{\rm cm}$ becomes a function of the square of the center-of-mass energy, $4\omega^2$, which also coincides with the Mandelstam variable s in this particular frame. To match the low-energy and high-energy behaviors of the scattering amplitude, we write $\mathcal{A}(s) \equiv \mathcal{A}_{\rm cm}(4\omega^2)$. We suppress the 'cm' subscript from now on. However, as we describe in Appendix A, away from the center-of-mass frame, the forward scattering amplitude in non-relativistic theories is typically not a function of only the Mandelstam variable s.

The argument for analyticity of the scattering amplitude off the real axis is similar to the relativistic case. However, for non-relativistic theories, the amplitude $\mathcal{A}(s)$ is not guaranteed to be symmetric under $s \to -s$. Hence, the sum rule (4.7) still applies, but the relationship between the contributions for positive and negative s needs to be reconsidered. In particular, the behavior for $s < 0$ is not directly determined (via crossing symmetry) by that at $s > 0$. We discuss the subtleties of the non-relativistic case in detail and illustrate the novel features in a specific example in Appendix A. Here, we just summarize the main results.

We will assume the existence of a high-energy scale, ρ, above which the theory becomes relativistic. As a consequence, $\mathcal{A}(s)$ satisfies the relativistic crossing symmetry for $|s| \gg \rho^2$. The contribution to the integral in (4.7) from $s \in (-\infty, -\rho^2]$ can therefore be mapped to $s \in [+\rho^2, +\infty)$, and we can write the sum rule as

$$\mathcal{A}''(s \to 0) = \underbrace{\frac{2}{\pi}\left(\int_0^\infty + \int_{\rho^2}^\infty\right)\mathrm{d}s\, \frac{\mathrm{Im}[\mathcal{A}(s)]}{s^3}}_{>0} + \underbrace{\frac{2}{\pi}\int_{-\rho^2}^0 \mathrm{d}s\, \frac{\mathrm{Im}[\mathcal{A}(s)]}{s^3}}_{?}\,. \tag{4.9}$$

from forward scattering. A similar attitude is used to ignore singularities from photon exchanges in QCD processes.

[5] Alternatively, we could introduce a small mass and later send it to zero. For scalar particles this does not modify the structure of the theory nor the UV behavior, which therefore still obeys the Froissart-Martin bound.

The integral above ρ^2 is positive definite, since it corresponds to the cross section to produce high-energy states in the theory, which we assume is dominated by relativistic interactions.[6]

Furthermore, for derivatively-coupled theories like the EFT we study in §4.2, the leading order amplitude at low energies, $0 < s \ll \rho^2$, is analytic in s. This is because particle production will be suppressed by extra factors of s over the cutoff scale of the EFT.[7] In other words, the tree level contribution dominates the amplitude since $\mathcal{A}(s) \propto s^2$ and $\text{Im}[\mathcal{A}] \propto |\mathcal{A}|^2 \propto s^4$. Therefore, at leading order, the branch cuts induced by loops of light particles do not contribute to $\text{Im}[\mathcal{A}]$. As we will see, other singularities for $0 < s < \rho^2$ (e.g. poles) do not appear unless extra light degrees of freedom are present. We therefore conclude that the first term on the right-hand side of (4.9) is manifestly positive. Only the region $-\rho^2 < s < 0$ may potentially lead to a negative contribution to the sum rule. In §4.3.1, we study an explicit example, in which an extra pole appears in the region $-\rho^2 \ll s < 0$. Nevertheless, positivity is still preserved in this example, and more generally for a large class of weakly coupled completions of small-c_s theories. In general, violations of positivity require large contributions from the u-channel whose signs are not fixed by the equivalent s-channel exchange. Although we cannot rule out such exotic (plausibly strongly coupled) possibilities, we are yet to encounter an explicit example. Nonetheless, we will take an agnostic attitude towards positivity, and in §4.2 we will derive constraints on the EFT of inflation by assuming a positive right-hand side of the sum rule. We believe these to be valuable consistency conditions on a vast class of single-field models with Lorentz-invariant UV completions. The same way a violation of the consistency condition derived in [25, 77] would require us to abandon the single-field hypothesis, violations of the positivity constraints we find here, although unlikely, would require us to incorporate the rather peculiar behavior we have identified in a full theory of inflation.

4.2 Implications for single-field inflation

In this section, we show how analyticity and unitarity of $\pi\pi \to \pi\pi$ scattering constrains the parameters of the EFT of inflation. In §4.2.1, we present the effective Lagrangian for the Goldstone boson π, at leading order in derivatives and to quartic order in fluctuations. We use this Lagrangian,[8] in §4.2.2, to compute the low-energy limit of the scattering of π-particles, and derive a positivity bound on the EFT parameters. In §4.2.3, we discuss perturbative unitarity of the scattering amplitude, in terms of its partial wave decomposition. We show that d-wave scattering leads to an improved derivation of the critical sound speed for which the EFT admits a perturbative UV completion [42].

[6] As we shall see, the forward scattering amplitude in the EFT is dominated by contact terms without long-range interactions, and therefore high energies are directly connected with short distances. In general, high-energy (e.g. super-Planckian) exchanges may still remain in a non-relativistic regime for very large impact parameters [75, 76].

[7] As we will discuss, the cutoff scale of the EFT, Λ, may be different from ρ.

[8] Following [78–80], it is in principle possible to extend our analysis to dissipative single-clock models or theories with excited initial states. However, extra care is required when computing scattering amplitudes for particle excitations in non-vacuum states. We leave this for future work.

4.2.1 Goldstone dynamics

In §3.3, we reviewed the basic elements of the EFT of inflation—an effective theory of the Goldstone boson, π, associated with the breaking of time translations in a quasi-de Sitter background. To relate the low-energy limit of the theory to its high-energy behavior, we will consider the scattering of π-particles. To perform the computations, we will exploit the natural hierarchy of scales in the problem (see Fig. 4.2). Since Goldstone bosons are derivatively coupled, their scattering amplitudes are dominated by high-energy (short-distance) processes near the cutoff scale Λ of the EFT. Moreover, at lower energies, only a handful of terms contribute since higher-order terms are suppressed by inverse powers of the cutoff. These features will allow us to work in the *flat space limit* ($H \to 0$) and compute scattering amplitudes without taking into account the cosmological expansion.[9]

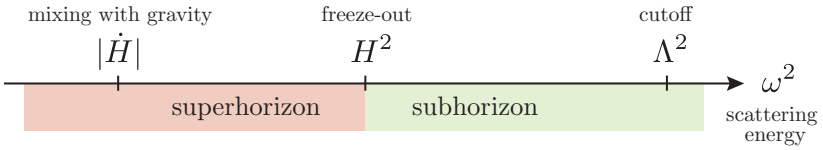


Figure 4.2: Illustration of the relevant energy scales of the EFT. The flat space approximation applies for scattering energies above the Hubble scale, $\omega^2 > H^2$. The decoupling limit captures the regime $\omega^2 > |\dot{H}|$. The hierarchy $H^2 \ll \Lambda^2$ is guaranteed by the high degree of Gaussianity of the primordial perturbations, while $|\dot{H}| \ll H^2$ holds as a condition for inflation and is supported by observations of the spectral index.

We will also take the decoupling limit ($M_{\rm pl}^2 \to \infty$, $\dot{H} \to 0$, with $M_{\rm pl}^2\dot{H}$ fixed), in which the mixing between π and gravitational perturbations vanishes. Computations performed in the decoupling limit will be accurate up to corrections that scale as $\dot{H}/H^2$ and $\omega^2/M_{\rm pl}^2$.[10] The Goldstone action at lowest order in derivatives was given in (3.37). The effective action, in principle, includes higher-derivative terms which we did not display [36]. However, in the flat space and decoupling limits, these terms are subdominant at low energies, $\omega/\Lambda \ll 1$, and will not contribute significantly to the left-hand side of the sum rule we derive in this paper. We will also be interested in the case where $\dot{H}(t) = \dot{H}$ and $M_n^4(t) = M_n^4$ are independent of time. This captures the behavior of the EFT of inflation in the limit of exact scale invariance. Deviations from scale-invariance can be treated perturbatively, but won't be relevant in this work.

Expanding the Goldstone action (3.37) up to quartic order in powers of π in the flat space

[9] This allow us to apply the sum rule (4.9) to the EFT of inflation. In a derivative expansion, the error induced in the left-hand side of (4.9) will be of order H^2/Λ^2. (This uses the fact that Goldstone bosons are only derivatively coupled.) The flat space approximation is more accurate on the right-hand side of (4.9). This is because the part of the integral which can be computed within the EFT is dominated by short-distance processes near the cutoff $\Lambda \gg H$. In addition, the rest of the integral includes contributions from higher energies, $\omega \gg \Lambda$, where the effects of the cosmological expansion are even less relevant.

[10] A somewhat unusual fact of the flat space and decoupling limits is that slow-roll inflation turns into a free theory—i.e. the inflaton potential $V(\phi)$ becomes constant and gravitational interactions are turned off. When these limits are taken, $\pi\pi \to \pi\pi$ scattering therefore has a trivial scattering amplitude for slow-roll inflation.

and decoupling limits, we get

$$\mathcal{L}_2 = M_{\rm pl}^2|\dot{H}|\Big(\dot{\pi}^2 - (\nabla\pi)^2\Big) + 2M_2^4\,\dot{\pi}^2\,, \tag{4.10}$$

$$\mathcal{L}_3 = \left(2M_2^4 - \frac{4}{3}M_3^4\right)\dot{\pi}^3 - 2M_2^4\,\dot{\pi}(\nabla\pi)^2\,, \tag{4.11}$$

$$\mathcal{L}_4 = \left(\frac{1}{2}M_2^4 - 2M_3^4 + \frac{2}{3}M_4^4\right)\dot{\pi}^4 - \Big(M_2^2 - 2M_3^4\Big)\dot{\pi}^2(\nabla\pi)^2 + \frac{1}{2}M_2^4(\nabla\pi)^4\,, \tag{4.12}$$

where $(\nabla\pi)^2 \equiv \delta^{ij}\partial_i\pi\partial_j\pi$. If $M_2 \neq 0$, then the Goldstone mode propagates with a non-trivial sound speed, c_s, defined in (3.39). Sometimes it will be convenient to rescale the spatial coordinate as $\tilde{x}^i = x^i/c_s$, so that (fake) Lorentz invariance is restored in the quadratic part of the action

$$\tilde{\mathcal{L}}_2 \equiv c_s^3\mathcal{L}_2 = -\frac{f_\pi^4}{2}(\tilde{\partial}\pi)^2\,. \tag{4.13}$$

We will find it convenient to normalize the EFT parameters M_n relative to f_π:

$$M_n^4 \equiv c_n\,\frac{f_\pi^4}{c_s^{2n-1}}\,, \tag{4.14}$$

where $c_2 \equiv \frac{1}{4}(1-c_s^2)$. The factors of c_s in (4.14) ensure that $c_n = \mathcal{O}(1)$ are natural parameter values even for $c_s \ll 1$. For instance, in DBI inflation [81, 82], all c_n are determined by c_s alone—in particular, $c_3 = -6c_2^2$ and $c_4 = 60c_2^3$. Observational constraints on the parameters (c_s, c_3, c_4) will be presented in §4.4. In the following, we will be concerned with theoretical bounds.

It will be convenient to write the effective Lagrangian in terms of the canonically normalized field $\pi_c \equiv f_\pi^2\pi$:

$$\tilde{\mathcal{L}} = -\frac{1}{2}(\tilde{\partial}\pi_c)^2 + \frac{1}{\Lambda^2}\left[\alpha_1\,\dot{\pi}_c^3 - \alpha_2\,\dot{\pi}_c(\tilde{\partial}\pi_c)^2\right] + \frac{1}{\Lambda^4}\left[\beta_1\,\dot{\pi}_c^4 - \beta_2\,\dot{\pi}_c^2(\tilde{\partial}\pi_c)^2 + \beta_3(\tilde{\partial}\pi_c)^4\right], \tag{4.15}$$

where we have introduced the cutoff scale $\Lambda \equiv f_\pi c_s$ and defined the following auxiliary parameters

$$\alpha_1 \equiv -2c_2(1-c_s^2) - \frac{4}{3}c_3\,, \quad \alpha_2 \equiv 2c_2\,, \tag{4.16}$$

$$\beta_1 \equiv \frac{1}{2}c_2(1-c_s^2)^2 + 2c_3(1-c_s^2) + \frac{2}{3}c_4\,, \quad \beta_2 \equiv -c_2(1-c_s^2) - 2c_3\,, \quad \beta_3 \equiv \frac{1}{2}c_2\,. \tag{4.17}$$

The organization of the effective Lagrangian (4.15) is somewhat unconventional: we have written all interactions in terms of the 'relativistic invariant' $(\tilde{\partial}\pi_c)^2$ and pure time derivatives $\dot{\pi}_c$. This is motivated by the analytic structure of scattering amplitudes, as discussed in Appendix A. The key point is that the 'relativistic' part will manifestly behave in a Lorentz-invariant manner, so we can trace all the subtleties of working in a non-Lorentz-invariant theory to the pure time derivatives.

4.2.2 Bounds from positivity

In what follows, we will derive a number of constraints on the Lagrangian parameters c_n (or equivalently M_n) from the requirements of analyticity and unitarity of $\pi\pi \to \pi\pi$ scattering. Details of the computations are given in Appendix A.

To gain intuition for the origin of the bounds, we first consider the special case $|c_4| \gg |c_3| \gg 1$. In [37], it was shown that this parameter regime is technically natural, so it is of a particular observational relevance. In this limit, the cubic Lagrangian is dominated by the $\dot{\pi}^3$ interaction (since $|\alpha_1| \to \frac{4}{3}|c_3| \gg \alpha_2$), and the quartic Lagrangian is domination by $\dot{\pi}^4$ (since $|\beta_1| \to \frac{2}{3}|c_4| \gg |\beta_2| \gg |\beta_3|$). The effective Lagrangian (4.15) then reduces to

$$\tilde{\mathcal{L}} \to -\frac{1}{2}(\tilde{\partial}\pi_c)^2 - \frac{4}{3}\frac{c_3}{\Lambda^2}\dot{\pi}_c^3 + \frac{2}{3}\frac{c_4}{\Lambda^4}\dot{\pi}_c^4 \,. \tag{4.18}$$

Computing the forward scattering amplitude in the center-of-mass frame, we find

$$\mathcal{A}(s) = \Big(c_4 - (2c_3)^2\Big)\frac{s^2}{\Lambda^4} \,, \tag{4.19}$$

and positivity, $\mathcal{A}'' > 0$, implies

$$\boxed{c_4 > (2c_3)^2}, \quad \text{for } |c_4| \gg |c_3| \gg 1. \tag{4.20}$$

We see that positivity simply requires that the contribution from the contact diagram ($\propto c_4$) dominates over that from the exchange diagram ($\propto c_3^2$). While either sign of c_4 is consistent with naturalness, only positive values satisfy the bound (4.20).

It is straightforward to repeat the analysis for the complete Lagrangian (4.15), i.e. without taking a special limit of the EFT parameters. From the cubic interactions, we get

$$\mathcal{M}_{\dot{\pi}^3} = -\frac{9}{4}\alpha_1^2\frac{s^2}{\Lambda^4} \,, \quad \mathcal{M}_{\dot{\pi}(\partial\pi)^2} = -4\alpha_2^2\frac{s^2}{\Lambda^4} \,, \quad \mathcal{M}_{\dot{\pi}(\partial\pi)^2\times\dot{\pi}^3} = -6\,\alpha_1\alpha_2\frac{s^2}{\Lambda^4} \,, \tag{4.21}$$

while the quartic interactions lead to

$$\mathcal{M}_{\dot{\pi}^4} = \frac{3}{2}\beta_1\frac{s^2}{\Lambda^4} \,, \quad \mathcal{M}_{\dot{\pi}^2(\partial\pi)^2} = 2\beta_2\frac{s^2}{\Lambda^4} \,, \quad \mathcal{M}_{(\partial\pi)^4} = \beta_3\,(3+\cos^2\theta)\frac{s^2}{\Lambda^4} \,. \tag{4.22}$$

Despite the fact that we have included diagrams that exchange massless particles, we see that the tree level amplitudes are analytic in s. Since these are the lowest dimension operators we could add to the EFT of inflation, any non-analytic behavior in the low-energy limit must enter at higher order in s. Notice also that the amplitude has no divergences as $\theta \to 0$ (due to the derivative nature of the Goldstone interactions) and therefore has a well-defined forward limit:

$$\mathcal{A}(s) = \sum_N \mathcal{M}_N(s,0)$$

$$= \left(-\frac{9}{4}\alpha_1^2 - 4\alpha_2^2 - 6\alpha_1\alpha_2 + \frac{3}{2}\beta_1 + 2\beta_2 + 4\beta_3\right)\frac{s^2}{\Lambda^4}$$
$$= \left(c_4 + 1 - \Big((2c_3+1) - a(c_s)\Big)^2 - b(c_s)\right)\frac{s^2}{\Lambda^4}\,, \tag{4.23}$$

where we defined

$$a(c_s) \equiv \frac{1-c_s^2}{4}(4+3c_s^2)\,, \qquad b(c_s) \equiv \frac{1-c_s^2}{16}(14+19c_s^2+15c_s^4)\,. \tag{4.24}$$

Positivity now implies that

$$\boxed{c_4 + 1 \;>\; \Big((2c_3+1) - a(c_s)\Big)^2 + b(c_s)}\,. \tag{4.25}$$

Notice that $b(c_s) \geq 0$, for all $c_s \in [0,1]$. The right-hand side of (4.25) is therefore positive semi-definite, and we conclude that

$$\boxed{c_4 + 1 > 0}\,, \quad \text{for } \textit{all} \text{ values of } c_3 \text{ and } c_s. \tag{4.26}$$

Moreover, in the limit $c_s \to 1$, (4.25) becomes

$$\boxed{c_4 + 1 \;>\; (2c_3+1)^2}\,, \quad \text{for } c_s = 1. \tag{4.27}$$

As we will see in §4.3.2, the last constraint can be reproduced by requiring the absence of superluminality around non-trivial backgrounds (with the additional requirement that $c_3 = 0$).

4.2.3 Perturbative unitarity

Given the full amplitude, $\mathcal{M}(s,\theta)$, we can learn more about the possible UV completions of the EFT by considering the perturbative unitarity of the partial wave amplitudes [42]. Perturbative unitarity will determine the scale at which the EFT becomes strongly coupled, and therefore sets an upper limit on the scale at which new physics must enter in a weakly coupled theory. These constraints are qualitatively different from the constraints from analyticity which must be satisfied at $s = 0$ for self-consistancy of the EFT. In contrast, perturbative unitarity constrains the extrapolation of the EFT to higher energies from the growth of the amplitude with s.

For this purpose, we write the amplitude in the following form

$$\mathcal{M}(s,\theta) = \left[f(c_s,c_3,c_4) + \frac{1-c_s^2}{12}P_2(\cos\theta)\right]\frac{s^2}{\Lambda^4} \;\equiv\; 16\pi\sum_\ell (2\ell+1)a_\ell(s)P_\ell(\cos\theta)\,. \tag{4.28}$$

Unitarity of the S-matrix requires that $\mathrm{Im}[a_\ell] = |a_\ell|^2$, which is only consistent if $|\mathrm{Re}[a_\ell]| < \frac{1}{2}$. When the tree level amplitude violates this condition, it means that loop corrections must be large and hence the theory is strongly coupled. We say that the theory violates 'perturbative unitarity'. Since the amplitude is a function of energy, this determines the energy scale at which perturbation

theory breaks down. For s-wave scattering, $|\mathrm{Re}[a_0]| < \frac{1}{2}$ can by achieved at all energies, by tuning the parameters in the function $f(c_s, c_3, c_4)$. However, d-wave scattering only involves the sound speed as a parameter and $|\mathrm{Re}[a_2]| < \frac{1}{2}$ implies

$$\frac{1}{60\pi}\frac{1-c_s^2}{c_s^4}\frac{\omega^4}{f_\pi^4} \;<\; \frac{1}{2}\,. \tag{4.29}$$

For a given value of c_s, perturbative unitarity will be violated at a specific energy $\omega_\star(c_s)$. Conversely, requiring the theory to be weakly coupled up to the symmetry breaking scale f_π, leads to a critical value of the sound speed

$$(c_s)_\star = 0.31\,. \tag{4.30}$$

For $c_s < (c_s)_\star$ the EFT becomes strongly coupled below the symmetry breaking scale. In other words, weakly coupled theories cannot produce $c_s \leq (c_s)_\star$ without the appearance of additional degrees of freedom *below* f_π. New physics of this type cannot occur in slow-roll inflationary models, which thus would be ruled out by a detection of $c_s < c_\star$.[11] Notice that, while our conclusions do not rely on the specific value of $c_\star$, the one in (4.30) is somewhat smaller than the value found in [42], $(c_s)_\star = 0.47$. The latter was derived from a partial answer to the s-wave amplitude, with $c_3 = c_4 = 0$. Unlike our previous result, the critical value reported here in (4.30) is more robust, and can only be modified by contributions that are higher order in ω.

4.3 Sum rule and positivity at work

The sum rule and positivity bounds discussed in the previous section are very general, but also quite abstract. At the same time, many aspects of scattering are subtle and counterintuitive in the non-relativistic context. Nevertheless, we have succeeded in deriving a sum rule relating the IR parameters of the EFT of inflation to the high-energy scattering amplitude

$$\boxed{\frac{1}{\Lambda^4}\left(c_4 + 1 - \Big((2c_3+1) - a(c_s)\Big)^2 - b(c_s)\right) \;=\; \frac{1}{\pi}\int_{-\infty}^{\infty}\mathrm{d}s\,\frac{\mathrm{Im}[\mathcal{A}(s)]}{s^3}}\,. \tag{4.31}$$

A further understanding of the physical connection between the low-energy and high-energy behaviors will require a more intuitive understanding of $\mathrm{Im}[\mathcal{A}(s)]$ in realistic theories. In this section, we will therefore study specific examples of models that UV-complete the $c_s \ll 1$ and $c_s = 1$ limits of the EFT. We will find that all examples are consistent with our positivity constraints. When $c_s \ll 1$, we will also see how the sum rule works explicitly. For $c_s = 1$, we will find that the positivity constraints are weaker than those derived from requiring subluminality around non-trivial background. We suggest that looking at non-forward scattering would lead to stronger constraints.

[11] The current bound $c_s \geq 0.024$ (95% CL) [83] still allows for either new (weakly coupled) physics or non-perturbative effects below (or at) f_π. This is similar to the situation in the pre-LHC/pre-Higgs era in particle physics. For further discussion see [42].

Based on those considerations, we will conjecture that $c_s = 1$ is always UV-completed by slow-roll inflation—i.e. a free scalar field in the flat space and decoupling limits.

4.3.1 Perturbative example with $c_s \ll 1$

The canonical example of inflation with a small sound speed is DBI inflation [81, 82]. While it is easy to show that the positivity bound (4.25) is satisfied for DBI inflation, it is less straightforward to study the high-energy scattering in this theory. To gain more intuition for how our sum rule works and how positivity arises, it will be instructive to study an example that remains perturbative up to high energies, $\omega \gg f_\pi$.

$\pi\sigma$-model. A reduced sound speed arises for fluctuations around curved trajectories in higher-dimensional field spaces. A simple two-field model that describes such dynamics is [66] (see also [64, 84–93]):

$$\mathcal{L} = -\frac{1}{2}k(\sigma)(\partial\phi)^2 - \frac{1}{2}(\partial\sigma)^2 - V(\sigma)\,, \tag{4.32}$$

where

$$k(\sigma) \equiv 1 + \frac{\sigma}{M} + \cdots\,, \qquad V(\sigma) \equiv \frac{1}{2}m^2\sigma^2 + \frac{1}{3!}\mu\sigma^3 + \cdots\,. \tag{4.33}$$

We have suppressed additional terms in the potential for σ which stabilize the second field at $\sigma_0 \ll M$ (see [84, 92]). The Lagrangian in (4.32) is itself only an EFT, valid at first order in a derivative expansion and up to energies of order M. The scale M thus becomes the new cutoff of the theory, which allows for perturbative control provided $\omega^2 < M^2$.

Perturbing around the background solution $\bar\phi(t)$, i.e. writing $\phi(t,\mathbf{x}) \equiv \bar\phi(t) + \dot{\bar\phi}\pi(t,\mathbf{x})$, we get a Lagrangian for the Goldstone fluctuations π, coupled to the additional field σ:

$$\begin{aligned}\mathcal{L} &= -\frac{1}{2}|\dot{\bar\phi}|^2\left(1+\frac{\sigma}{M}\right)\Big[-2\dot\pi + (\partial\pi)^2\Big] - \frac{1}{2}(\partial\sigma)^2 - \frac{1}{2}m^2\sigma^2 - \frac{1}{3!}\mu\sigma^3\,,\\ &= -\frac{1}{2}(\partial\bar\pi)^2 - \frac{1}{2}(\partial\sigma)^2 - \rho\sigma\dot{\bar\pi} - \frac{\sigma(\partial\bar\pi)^2}{2M} - \frac{1}{2}m^2\sigma^2 - \frac{1}{3!}\mu\sigma^3\,,\end{aligned} \tag{4.34}$$

where we have only kept the leading order terms. In the second line, we have defined $\bar\pi = |\dot{\bar\phi}|\pi$ and $\rho \equiv |\dot{\bar\phi}|/M$. In the following, we will assume the hierarchy of scales

$$\mu^2 \lesssim m^2 \ll \rho^2\,. \tag{4.35}$$

The dynamics of the Lagrangian (4.34) are discussed in detail in [64, 92]. At high energies, $\omega > \rho$, the theory describes two relativistic scalars, whose interaction can be treated as a small perturbation. Below $\omega = \rho$, the mixing becomes relevant and the theory reduces to a single propagating degree of freedom. For $m < k < \rho$, the dispersion relation of the Goldstone π is nonlinear, $\omega = k^2/\rho$. As explained in [64], integrating out the field σ produces a non-local action

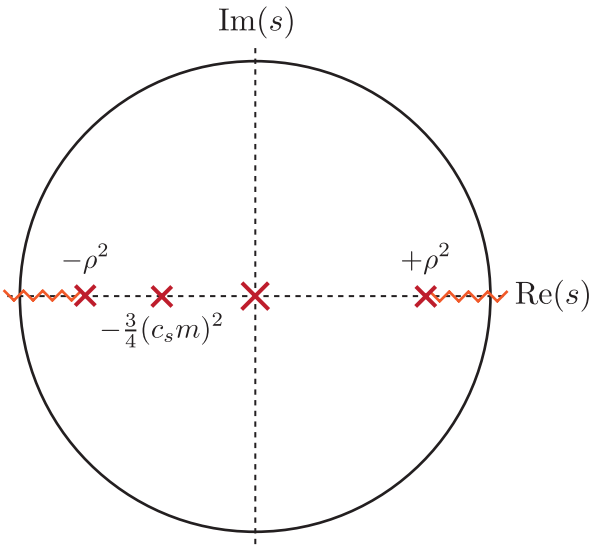


Figure 4.3: Illustration of the pole structure of the amplitude in (4.38).

for π, which is not captured by (3.37). In order to have a local description requires keeping the field σ, even though it then plays the role of an auxiliary field.[12] For $k < m$ (or $\omega \lesssim c_s m$), the dispersion relation becomes linear, and the low-energy EFT is characterized by a reduced sound speed

$$c_s^2 = \frac{m^2}{m^2 + \rho^2} \simeq \frac{m^2}{\rho^2} \,. \tag{4.36}$$

The effective theory is thus described in terms of (3.37) without reference to σ. Notice that, for $c_s \ll 1$, the range of validity of the single-field EFT description is smaller than the naive expectation, which associates the cutoff of the EFT with the mass of the particle that has been integrated out. As explained in [64], this lower scale appears as a result of the $\rho \gg m$ hierarchy, which creates the window with a nonlinear dispersion for $c_s m < \omega < \rho$. The relevance of the new scale $c_s m$ can be seen, for instance, by the presence of a pole at negative frequencies in the scattering amplitude (see Fig. 4.3).

To match the parameters of the model to the EFT parameters in §4.2, we note that

$$M_{\rm pl}^2 |\dot{H}| = \frac{1}{2} |\dot{\bar{\phi}}|^2 \,, \qquad f_\pi^4 = \frac{m}{\rho} |\dot{\bar{\phi}}|^2 \,, \qquad \Lambda^4 = \frac{m^5}{\rho^5} |\dot{\bar{\phi}}|^2 \,. \tag{4.37}$$

We now compute $\pi\pi \to \pi\pi$ scattering in the $\pi\sigma$-model and show how it fits into the analysis of the previous sections.

$\mu = 0$. Let us first consider the special case $\mu = 0$. When $s \ll c_s^2 m^2$, the $2 \to 2$ scattering amplitude for the gapless mode of the system should match the results of §4.2.2, after using (4.36)

[12] Let us emphasize that most of these features appear because Lorentz invariance is spontaneously broken, and are commonplace, for example, in non-relativistic condensed matter systems.

and (4.37). The calculation of the amplitude is technically straightforward and is performed in Appendix A. The result in the forward limit is

$$\mathcal{A} = -\frac{Z^4(\omega)}{M^2}\left\{(\omega^2+k^2)^2\left[\frac{1}{4\omega^2-m^2-\rho^2}-\frac{1}{4k^2+m^2}\right]-(\omega^2-k^2)^2\frac{1}{m^2}\right\}, \tag{4.38}$$

where $Z(\omega)$ is the relative normalization between $\bar{\pi}$ and the scattering state of the gapless mode, and is given in (A.20). In the low-energy limit, $k \ll m$, we have $Z(\omega \to 0) \to c_s$, and it is easy to check that the result in (4.38) matches[13] the scattering amplitude computed in the EFT after expanding in k/m.

For $\mu = 0$, we have $M^4_{n>2} = 0$ and the amplitude trivially satisfies the positivity constraint. Nevertheless, the analytic structure and the validity of the sum rule (4.31) arise quite non-trivially. Equation (4.38) has two poles, one at $s = m^2 + \rho^2 \simeq \rho^2$ and another at $s = -\frac{3}{4}c_s^2 m^2$ (or $k^2 = -\frac{1}{4}m^2$). These two poles are related by crossing symmetry, but the pole on the negative axis is shifted relative to the location of the new physical state on the positive axis. In the limit $c_s \ll 1$, the pole at $s = -\frac{3}{4}c_s^2 m^2$ dominates the right-hand side of (4.31). In fact, it is the only contribution at leading order in c_s (see Appendix A).

$\mu \neq 0$. For $\mu \neq 0$, we will generate non-zero M_3^4 and M_4^4 after integrating out σ. For sufficiently large μ, we expect the low-energy contributions to M_3^4 and M_4^4 to dominate over M_2^4. In the following, we will work in the same limit as at the beginning of §4.2.2, namely $|c_4| \gg |c_3| \gg c_2$. This case is particularly interesting because not every choice of c_3 and c_4 is consistent with positivity. As a result, this case offers a non-trivial test of our bounds.

The most reliable way to determine the low-energy behavior is to compute the forward amplitude and match to the EFT at low energies. This calculation is performed in Appendix A. At leading order in $c_s \ll 1$, the amplitude in the low-energy limit, $s \ll \Lambda^2$, becomes

$$\mathcal{A}_{\mu^2} \to \frac{1}{8}\frac{\mu^2}{m^6}s^2, \tag{4.39}$$

which matches the energy scaling of the EFT computation, as it should. More importantly, the result in (4.39) is manifestly positive. This means that any choice of μ will produce a combination of c_3 and c_4 which is consistent with the bound in (4.20):

$$c_4 - (2c_3)^2 = \frac{1}{8}\frac{\mu^2 M^2}{m^4} > 0. \tag{4.40}$$

Although expected, the result is non-trivial. Naively, it might have seemed possible[14] that the cubic interaction would generate large values for c_3 ($\gg 1$), while keeping $c_4 = 0$. This, however,

[13] To directly compare the results one must account for the rescaling $\tilde{x}^i = x^i/c_s$ that we used previously.

[14] Given that the potential for σ is unstable without including a quartic interaction, one might have imagined that positivity of the amplitude is enforced through stability. Perhaps unsurprisingly, positivity is a more robust feature of perturbation theory that holds for any μ.

would be inconsistent with positivity.

We have found that the $\pi\sigma$-model always produces a value of c_4 that is in agreement with our consistency condition. As we discuss in more detail in Appendix A, this is a generic feature of a large class of models. In particular, this holds for all weakly coupled theories in which the $2 \to 2$ scattering of the gapless mode is dominated by the exchange of a single heavy state at low energies.

4.3.2 Conjecture for $c_s = 1$

Single-field slow-roll inflation famously leads to $c_s = 1$ and produces little non-Gaussianity [25]. In fact, in the flat space and decoupling limits that we have been discussing, the Lagrangian for slow-roll inflation becomes that of a free field, $\mathcal{L} = -\frac{1}{2}(\partial\phi)^2$. This theory trivially saturates our positivity constraints because $M^4_{n\geq 2} = 0$ and $\mathcal{A}(s) = 0$. However, while slow-roll inflation is consistent with our bound, it is difficult to find an explicit example of a UV-complete theory with $c_s = 1$, but $c_3, c_4 \neq 0$. (For example, in DBI inflation we have $c_{n\geq 2} \to 0$ when $c_s \to 1$.) In this section, we will provide suggestive evidence for the conjecture that theories with $c_s = 1$ are always UV-completed by slow-roll inflation, without higher-order Goldstone self-interactions. If proven, such a result would allow us to directly link constraints on c_s to the unique mechanism for inflation.

First, we will show that the positivity bound from the previous section, $c_4 + 1 > (2c_3 + 1)^2$ (for $c_s = 1$), is weaker than the constraint that derives from imposing subluminal speed of propagation in non-trivial backgrounds. For this purpose, we return to the Goldstone Lagrangian in the form

$$\frac{\mathcal{L}}{f_\pi^4} = -\frac{1}{2}(\partial\pi)^2 + \sum_{n=3}^{\infty} \frac{c_n}{n!}\left[-2\dot{\pi} + (\partial\pi)^2\right]^n , \tag{4.41}$$

where we have set $c_2 = 0$ since we are concerned with the $c_s = 1$ limit. A trivial solution to the linearized equations for motion is $\pi = \alpha_\mu x^\mu + \beta$. For timelike x^μ, we can choose $\alpha_\mu = (\alpha, 0, 0, 0)$ and $\beta = 0$. At leading order in small α, the quadratic Lagrangian for the fluctuations, φ, around this background (i.e. $\pi = -\alpha t + \varphi$) is given by

$$\frac{\mathcal{L}_2}{f_\pi^4} = -\frac{1}{2}(\partial\varphi)^2 + 4\alpha c_3 \dot{\varphi}^2 + \mathcal{O}(\alpha^2) , \tag{4.42}$$

where we have dropped total derivative terms. Around the new background, the speed of propagation is

$$c^2_{s,\varphi} = 1 + 8\alpha c_3 + \mathcal{O}(\alpha^2) . \tag{4.43}$$

Since α can have either sign, we require $c_3 = 0$ to avoid superluminal speed. Going to next order

in α, we find

$$c_{s,\varphi}^2 = 1 - 4\alpha^2 c_4 + \mathcal{O}(\alpha^3)\,, \tag{4.44}$$

and superluminality is avoided iff $c_4 \geq 0$. It may be surprising that in this limit the constraint from subluminality ($c_3 = 0$, $c_4 \geq 0$) is stronger than that from positivity ($c_4 + 1 > (2c_3 + 1)^2$). However, a similar observation was made in [94]. In a relativistic EFT, it was observed that positivity of forward scattering gave qualitatively different bounds from requiring subluminal propagation around non-trivial backgrounds, and stronger results could be derived from sum rules involving non-forward scattering amplitudes. This suggests that a stronger bound may arise for fixed-angle scattering.

Inspection of the full amplitude computed in (4.21) and (4.22), shows that the only term with angular dependence is the one proportional to $\beta_3 \equiv \frac{1}{8}(1 - c_s^2)$. This d-wave contribution can be isolated for instance by decomposing the amplitude in partial waves—cf. (4.28)—such that $a_2(s) \propto (1 - c_s^2)s^2$. One may then hope to derive a sum rule for the d-wave amplitude (and hence the value of c_s):

$$\frac{1 - c_s^2}{c_s^4} \stackrel{?}{=} \int \mathrm{d}s\, f(s)\,, \tag{4.45}$$

where the function $f(s)$ would be related to the partial wave amplitudes. Isolating partial waves via non-forward dispersion relations is common in relativistic theories (e.g. [95]), so it seems feasible to derive a similar expression in the non-relativistic regime. Positivity of the sum rule (4.45), would simply correspond to subluminality of the speed of propagation at low energies, as is expected for all consistent (and Lorentz-invariant) UV theories. At the same time, provided the right-hand-side of (4.45) is positive, the vanishing of the left-hand-side for $c_s = 1$ would imply that $f(s)$ must vanish.[15] This would be true (almost by definition) for a free theory, which would then constrain all interactions of the EFT to vanish. Hence, it seems likely that a sum rule which isolates $\beta_3 = 0$ (or $c_s = 1$) would ultimately force $c_{n>2} = 0$.

Unfortunately, writing a sum rule for the partial waves introduces new challenges that are not present for the full amplitude at forward scattering. First of all, the analytic properties of the scattering amplitude are less understood for non-relativistic scattering at fixed angle (or fixed transfer momentum). Furthermore, going from the amplitude to the partial waves requires an integration over angles, which in many cases alters the (non-)analytic behavior. Some of these shortcomings may be circumvented in the relativistic context, mostly because of the extensive use of (s, t, u) crossing symmetry [95], which is not available in non-relativistic theories.

An alternative is to adapt the derivation of the Kramers-Kronig relation for the refraction index, $n(\omega) \equiv c_s^{-1}(\omega)$, to our case. If $n(\omega)$ is analytic in the upper-half plane (as it is for light in

[15] Ideally, the function $f(s)$ would be linked to the imaginary part of the partial wave amplitude which, due to unitarity and the optical theorem, carries information about the scattering and production of intermediate states in an interacting theory. A vanishing imaginary part would correspond to a free theory.

a medium), and it satisfies the limit $n(\omega \gg \Lambda) \to 1$, then the equivalent of the Kramers-Kronig dispersion relation holds

$$\mathrm{Re}[n(0)] - 1 \stackrel{?}{=} \int_0^\infty \frac{\mathrm{Im}[n(\omega)]}{\omega} \,. \tag{4.46}$$

We notice that (4.46) is qualitatively similar to (4.45). In particular, if $c_s = 1$ at low energies (i.e. $\mathrm{Re}[n(0)] \to 1$), then the dispersive term on the right-hand side again vanishes. One of the obstacles in this derivation is establishing the off-shell frequency/momentum dependence of the Green's function. Although causality guarantees certain properties for the Green's function, these are not translated as easily into the analytic behavior of $n(\omega)$ as in the electromagnetic case. While we do not think that the problems described above are insurmountable, they make the status of non-relativistic sum rules for partial wave amplitudes, or the refraction index, somewhat uncertain. We will return to these issues in future work.

4.4 Conclusions

In this chapter, we used causality (and unitarity) to link cosmological observables, and the related coefficients in the IR theory, to the unknown UV dynamics of inflation. For single-field inflation, EFT parameters that higher-order correlations can measure (or constrain) include the sound speed, c_s, as well as a cubic coupling, c_3, and a quartic coupling, c_4. The latest constraints on the parameters c_s and c_3 from the CMB bispectrum [83] are shown in Fig. 4.4. The first constraint on the parameter c_4 has recently been derived from measurements of the CMB trispectrum [83] (see also [96–98])

$$-8.3 \times 10^7 \;<\; \frac{c_4}{c_s^4} \;<\; 7.4 \times 10^7 \quad (95\%\,\mathrm{CL}). \tag{4.47}$$

Let us note that this limit assumes $c_3 = 0$, and a dedicated analysis of the CMB bispectrum and trispectrum for general c_s, c_3 and c_4 is still lacking. However, we already see that much of the parameter space remains to be observationally explored. The theoretical bounds that we discussed in this chapter are therefore very relevant.

We showed that analyticity of the $2 \to 2$ scattering amplitude for the Goldstone boson implies a sum rule that relates a combination of the parameters (c_s, c_3, c_4) to an integral over the high-energy spectrum of the scattering amplitude—cf. (4.31). Hence, the EFT parameters are connected to specific features of scattering processes in the UV completion of inflation. Assuming positivity of the sum rule, we then derived a new consistency condition which bounds the size of the four-point function in terms of the square of the three-point function for equilateral configurations. This consistency condition restricts the size and the sign of the quartic coupling c_4. While we have not been able to construct an explicit example in which our bound is violated, we have isolated the necessary ingredients. We have also argued that our consistency condition is a generic feature in weakly coupled theories. Hence, finding large negative values of c_4 would point towards less

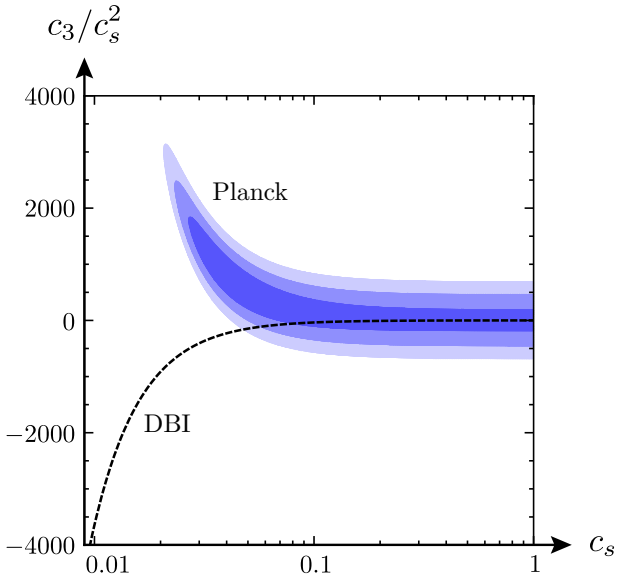


Figure 4.4: Observational constraints on the EFT parameters c_s and c_3 [83].

conventional (plausibly strongly coupled) theories of inflation, or more radically to violations of basic properties of scattering amplitudes (e.g. [69]).

We consider the analysis presented in this chapter to be only a first and modest step towards a more complete understanding of the IR/UV connections between cosmological observations and the underlying physics of inflation. Many future directions suggest themselves. For instance, we may hope to find sum rules for individual parameters of the EFT, rather than just for a special combination of several of them. This may be possible by extending our analysis to non-forward scattering, or through generalized Kramers-Kronig relations for the Green's functions. We have speculated that such an analysis would allow us to derive a sum rule for c_s, the speed of propagation of the Goldstone mode. In this case, positivity would correspond to the expected subluminality condition: $c_s < 1$. On the other hand, in the limit $c_s \to 1$, the sum rule would constrain the total amplitude to vanish. This has led us to conjecture that theories with $c_s = 1$ can only be UV completed by slow-roll inflation. While, so far, we have only given suggestive evidence for this intriguing conjecture, we hope to provide a positive answer to this question in the near future.

5. Non-Gaussianity as a particle detector

We have so far studied inflationary dynamics in the context of single-field inflation, involving two massless fields that are guaranteed to exist during inflation: the curvature perturbation, ζ (or the Goldstone boson of broken time translations, π), and the graviton, γ_{ij}. Establishing the field content during inflation beyond this minimal context, however, is a fundamental challenge of primordial cosmology. While at present there is no evidence for additional degrees of freedom [40], the imprints of extra particles can be subtle, so it remains important to fully characterize their effects and compare them to observations. Moreover, massive particles are important probes of the UV completion of inflation. For example, in string theory, massive particles in the low-energy effective theory encode physics at the string and Kaluza-Klein scales [51]. If these scales aren't too far from the inflationary Hubble scale, then their influence may be observable (although the experimental challenge could be enormous).

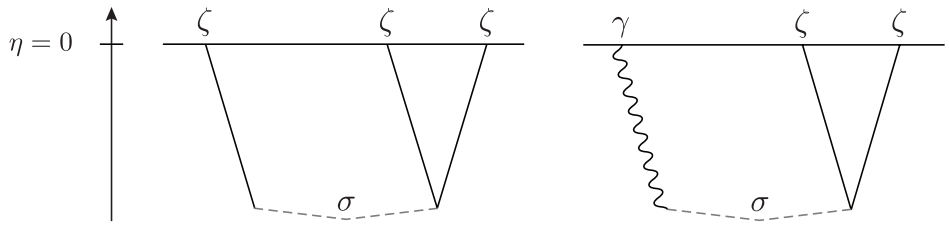


Figure 5.1: Diagrams contributing to $\langle\zeta\zeta\zeta\rangle$ and $\langle\gamma\zeta\zeta\rangle$. The solid, dashed, and wavy lines represent the curvature perturbation ζ, a massive spin-s field $\sigma_{i_1\cdots i_s}$, and the graviton γ_{ij}, respectively.

Since massive particles decay outside of the horizon during inflation, they cannot be observed directly in late-time correlation functions. Instead, the presence of massive particles has to be inferred from their indirect effects on the correlation functions of ζ and γ_{ij} (see Fig. 5.1). Some of these effects can be mimicked by adding a local vertex in the low-energy effective Lagrangian, which is the result of integrating out the heavy fields. On the other hand, massive particles may spontaneously be created in an expanding spacetime [99–101], an effect which cannot be represented by adding a local vertex to the effective Lagrangian [102]. The role of these non-local effects as a means of detecting massive particles during inflation was recently highlighted by Arkani-Hamed and Maldacena [102]: the spontaneous particle creation allows us to probe massive fields during inflation, even though we are only observing the late-time expectation values of light

fields. The rate of particle production in de Sitter space is exponentially suppressed as a function of mass, $e^{-m/T_{\rm dS}}$, with $T_{\rm dS} \equiv H/2\pi$, so their imprints will only be detectable if their masses are not too far above the Hubble rate H.[1] Since the inflationary scale may be as high as $10^{14}\,\text{GeV}$, this nevertheless provides an opportunity to probe massive particles far beyond the reach of conventional particle colliders.

In this chapter, we analyze the allowed couplings of massive particles with spin to the Goldstone boson of broken time translations and the graviton, and discuss their observational signatures. In §5.1, we first collect the equations of motion for massive fields with spin in de Sitter space, whose solutions are presented in Appendix B. In §5.2, we then construct the effective action for the leading interactions between the Goldstone boson π, the graviton γ_{ij}, and massive spinning fields $\sigma_{\mu_1\ldots\mu_s}$. We analyze under what conditions the theory is under perturbative control and discuss various constraints on the sizes of the couplings. In §5.3, we compute the correlation functions associated with the interactions of §5.2. We estimate the maximal amount of non-Gaussianity that is consistent with the constraints on the couplings of the effective theory. Details of the in-in computation are relegated to Appendix C, including the analytic results for soft limits. The conclusions are presented in §5.4.

5.1 Spin in de Sitter space

We begin by reviewing a few elementary facts about massive fields with spin in four-dimensional de Sitter space, dS_4.

Spin-1. The quadratic action of a massive spin-1 field σ_μ in de Sitter space is

$$S_1 = \int \text{d}^4x\sqrt{-g}\left[-\frac{1}{2}\nabla_\mu\sigma_\nu\nabla^\mu\sigma^\nu + \frac{1}{2}(\nabla^\mu\sigma_\mu)^2 - \frac{1}{2}m_1^2\sigma^\mu\sigma_\mu\right], \tag{5.1}$$

where $m_1^2 \equiv m^2 + 3H^2$, with m being the mass of the field.[2] The structure of the action (5.1) is uniquely fixed by requiring the absence of ghost degrees of freedom.[3] Up to integration by parts, this is equivalent to the Proca action. Variation of the action yields the equation of motion, $\Box\sigma_\mu - \nabla_\mu\nabla^\nu\sigma_\nu - m_1^2\sigma_\mu = 0$, where $\Box \equiv \nabla^\mu\nabla_\mu$ denotes the Laplace-Beltrami operator on dS_4. Taking the divergence of this equation gives the constraint $\nabla^\mu\sigma_\mu = 0$. The on-shell equation of motion then becomes

$$\left(\Box - m_1^2\right)\sigma_\mu = 0\,. \tag{5.2}$$

[1] If the extra fields have strongly time-dependent masses, whose Fourier transforms have support at a frequency $\hat{\omega}$, then non-adiabatic particle production occurs at a rate proportional to $e^{-m/\hat{\omega}}$ [103]. The scale $\hat{\omega}$ may be as large as $\dot{\bar{\phi}}^{1/2} = 58\,H$ without spoiling the slow-roll dynamics. In models with these types of time-dependent couplings, the detectable range of particle masses is somewhat enlarged.

[2] We define the mass parameter in such a way that the action acquires a gauge invariance in the massless limit, $m = 0$. This is required in order for massless spinning fields to propagate the right number of degrees of freedom. The mass defined in this way can also be identified as the mass of the field in the flat space limit [104].

[3] The ghost-free structure of the quadratic action will remain valid as long as nonlinear interactions can be treated perturbatively.

In Appendix B, we derive the solutions to this equation for the different helicity components of the field.

Spin-2. The unique ghost-free quadratic action of a massive spin-2 field $\sigma_{\mu\nu}$ in de Sitter space is [105]

$$S_2 = \int \mathrm{d}^4x\sqrt{-g}\Big[-\frac{1}{2}\nabla^\alpha\sigma^{\mu\nu}\nabla_\alpha\sigma_{\mu\nu} + \nabla^\mu\sigma_{\mu\nu}\nabla_\alpha\sigma^{\alpha\nu} - \nabla^\mu\sigma_{\mu\nu}\nabla^\nu\tilde{\sigma} + \frac{1}{2}\nabla^\mu\tilde{\sigma}\nabla_\mu\tilde{\sigma} \\ - \frac{1}{2}m_2^2(\sigma^{\mu\nu}\sigma_{\mu\nu} - \tilde{\sigma}^2) - \frac{3}{2}H^2\tilde{\sigma}^2\Big], \tag{5.3}$$

where $m_2^2 \equiv m^2 + 2H^2$ and $\tilde{\sigma} \equiv \sigma^\mu{}_\mu$ denotes the trace. Varying the action with respect to $\sigma_{\mu\nu}$, we obtain

$$\Box\sigma_{\mu\nu} - 2\nabla_{(\mu}\nabla^\alpha\sigma_{\nu)\alpha} + \nabla_\mu\nabla_\nu\tilde{\sigma} + g_{\mu\nu}(\nabla^\alpha\nabla^\beta\sigma_{\alpha\beta} - \Box\tilde{\sigma}) - m_2^2\sigma_{\mu\nu} + (m_2^2 - 3H^2)g_{\mu\nu}\tilde{\sigma} = 0\,. \tag{5.4}$$

Taking the divergence gives $\nabla^\mu\sigma_{\mu\nu} = \nabla_\nu\tilde{\sigma}$, and plugging this back into the equation yields $(m^2 - 2H^2)\tilde{\sigma} = 0$. For $m^2 \neq 2H^2$, the equation of motion and the constraints satisfied by the field $\sigma_{\mu\nu}$ are[4]

$$\left(\Box - m_2^2\right)\sigma_{\mu\nu} = 0\,, \quad \nabla^\mu\sigma_{\mu\nu} = 0\,, \quad \tilde{\sigma} = 0\,. \tag{5.5}$$

In Appendix B, we derive the solutions to the on-shell conditions (5.5).

Spin-s. The Lagrangian for massive fields with arbitrary spin in flat space was constructed in [107, 108], and generalized to (A)dS spaces in [109]. For massive fields with spin greater than 2, the action is rather complex and requires introducing auxiliary fields of lower spins. An alternative, which we will follow, is to use a group theoretical approach to find the equations of motion directly [110]. A massive bosonic spin-s field is described by a totally symmetric rank-s tensor, $\sigma_{\mu_1\cdots\mu_s}$, subject to the constraints

$$\nabla^{\mu_1}\sigma_{\mu_1\cdots\mu_s} = 0\,, \quad \sigma^\mu{}_{\mu\mu_3\cdots\mu_s} = 0\,. \tag{5.6}$$

The conditions in (5.6) project out the components of the tensor which transform as fields with lower spins. The Casimir eigenvalue equation of the de Sitter group then gives the wave equation satisfied by these fields:

$$\left(\Box - m_s^2\right)\sigma_{\mu_1\cdots\mu_s} = 0\,, \tag{5.7}$$

where $m_s^2 \equiv m^2 - (s^2 - 2s - 2)H^2$. The shift in the mass arises from the mismatch between the Casimir and Laplace-Beltrami operators in de Sitter space and is necessary to describe the correct

[4] For $m^2 = 2H^2$, the system enjoys a (partial) gauge invariance $\sigma_{\mu\nu} \to \sigma_{\mu\nu} + \nabla_{(\mu}\nabla_{\nu)}\xi$, and the longitudinal (helicity-0) mode becomes non-dynamical [106].

representations for massless fields. Equivalently, it is required by imposing gauge invariance in the massless limit, $m = 0$. Solutions to equation (5.7) are obtained in Appendix B.

Following Wigner [111], we identify the spectrum of particles by the unitary irreducible representations of the spacetime isometry group. For the de Sitter group SO(1, 4), these representations fall into three distinct categories [112, 113]:

principal series	complementary series	discrete series
$\frac{m^2}{H^2} \geq \left(s-\frac{1}{2}\right)^2$	$s(s-1) < \frac{m^2}{H^2} < \left(s-\frac{1}{2}\right)^2$	$\frac{m^2}{H^2} = s(s-1) - t(t+1)$

for $t = 0, 1, 2, ..., s-1$. Masses that are not associated with one of the above categories are forbidden and correspond to non-unitary representations. At the specific mass values corresponding to the discrete series, the system gains an additional gauge invariance and some of the lowest helicity modes become pure gauge modes—this phenomenon is called partial masslessness [114]. The spectrum of massive particles is contained in the principal and complementary series. We see that unitarity demands the existence of a lower bound, $m^2 > s(s-1)H^2$, on the masses of fields that belong to this spectrum. For $s = 2$, this is known as the Higuchi bound [105].

In the late-time limit, the generators of the de Sitter isometries form the 3-dimensional conformal group. The asymptotic scaling of a spin-s field is

$$\lim_{\eta\to 0} \sigma_{i_1\cdots i_s}(\eta, \mathbf{x}) = \sigma^{+}_{i_1\cdots i_s}(\mathbf{x})\, \eta^{\Delta_s^+ - s} + \sigma^{-}_{i_1\cdots i_s}(\mathbf{x})\, \eta^{\Delta_s^- - s}\,, \tag{5.8}$$

where the conformal weight of the field is defined as[5]

$$\Delta_s^{\pm} = \frac{3}{2} \pm i\mu_s\,, \quad \text{with} \quad \mu_s \equiv \sqrt{\frac{m^2}{H^2} - \left(s-\frac{1}{2}\right)^2}\,. \tag{5.9}$$

In this chapter, we will deal mostly with particles belonging to the principal series which covers the largest mass range and corresponds to $\mu_s \geq 0$. For real μ_s, the asymptotic scaling is given by a complex-conjugate pair, resulting in a wavefunction that oscillates logarithmically in conformal time. The complementary series has imaginary μ_s and corresponds to the interval $-i\mu_s \in (0, 1/2)$. In that case, only the growing mode survives in the late-time limit.

5.2 Spin in the effective theory of inflation

In this section, we will construct the leading interactions between the Goldstone boson of broken time translations π, the graviton γ_{ij}, and massive spinning fields $\sigma_{\mu_1\ldots\mu_s}$.[6] We will also consider

[5] Notice that for $s = 0$, the case $m = 0$ corresponds to a conformally coupled scalar field. For a minimally-coupled massless scalar, one should instead use $m^2 \to m^2 - 2H^2$ in (5.9).

[6] For constructions of the EFT of inflation with extra scalar fields, see [115, 116].

self-interactions of the massive spinning fields, and focus on terms which contribute to the correlation functions $\langle\zeta\zeta\zeta\rangle$ and $\langle\gamma\zeta\zeta\rangle$ at tree level and at leading order in derivatives. Moreover, we will restrict our presentation to the subset of interactions which give rise to a distinctive angular dependence due to the exchange of the spinning fields.

In §5.2.1 and §5.2.2, we introduce the couplings of massive particles with spin to the Goldstone and the graviton, respectively; first for the special cases $s = 1$ and 2, and then for arbitrary spin. In §5.2.3, we discuss how large the mixing interactions can be made while keeping the effective theory under theoretical control.

5.2.1 Couplings to the Goldstone

We reviewed the construction of the effective actions for the Goldstone boson and the graviton in Chapter 3. The construction of the effective action for couplings between the Goldstone and spinning fields proceeds similarly. We first write down all operators consistent with the symmetries. Amongst them will be tadpole terms, which must add up to zero. In unitary gauge, the basic building blocks involving spinning fields are $\sigma^{0\cdots0}$ and all Lorentz-invariant self-interactions, e.g. $\sigma^{\mu_1\cdots\mu_s}\sigma_{\mu_1\cdots\mu_s}$. The latter are invariant under all diffeomorphisms, so they don't lead to a coupling to π after the Stückelberg trick, whereas the former transform as

$$\sigma^{0\cdots0} \to (\delta^0_{\mu_1} + \partial_{\mu_1}\pi)\cdots(\delta^0_{\mu_s} + \partial_{\mu_s}\pi)\,\sigma^{\mu_1\cdots\mu_s}\,. \tag{5.10}$$

We may also have contractions with the curvature tensors, which appear at higher order in derivatives.

Spin-1. We first analyze the couplings between a massive spin-1 field σ_μ and the Goldstone boson π. In unitary gauge, the operators of the effective action involve g^{00} and σ^0. In order not to alter the background solution, these operators have to start at quadratic order in fluctuations.

- At leading order in derivatives and to linear order in σ_μ, the mixing Lagrangian is[7]

$$\mathcal{L}^{(1)}_{\pi\sigma} = \omega_1^3\,\delta g^{00}\sigma^0 + \omega_2^3\,(\delta g^{00})^2\sigma^0\,. \tag{5.11}$$

  Introducing π using (3.36) and (5.10), we get

$$\mathcal{L}^{(1)}_{\pi\sigma} = \omega_1^3\,a^{-2}\big(2\partial_i\pi\sigma_i - (\partial_i\pi)^2\sigma_0 - 2\dot\pi\partial_i\pi\sigma_i\big) + (3\omega_1^3 + 4\omega_2^3)\,\dot\pi^2\sigma_0 + \cdots\,, \tag{5.12}$$

  where we have taken the decoupling limit so that couplings to metric fluctuations become irrelevant.[8] We also used the constraint $\nabla^\mu\sigma_\mu = 0$, which we assume to hold at the back-

[7] Note that there are no terms involving $\delta g^{0\mu}\sigma_\mu$ in the effective action. This is because this operator does not satisfy the symmetries of the EFT, since the background value $\bar g^{0\mu}\sigma_\mu = -\sigma_0$ transforms nontrivially under spatial diffeomorphisms (and so does the fluctuation).

[8] The decoupling limit is not affected by the inclusion of mixing interactions, provided that we are in the perturbative regime. This can be shown by an ADM analysis of the metric perturbations [116, 117].

ground level, to replace $\dot{\pi}\sigma_0$ by $\partial_i\pi\sigma_i$. Since only the cubic mixing $\dot{\pi}\partial_i\pi\sigma_i$ will lead to the characteristic angular structure in the resulting correlation functions (see §5.3.2), we will focus on the bispectrum created by the combination of $\dot{\pi}\partial_i\pi\sigma_i$ and $\partial_i\pi\sigma_i$. Note that there is a single parameter ω_1 controlling the size of these two interactions. This is a consequence of the nonlinearly-realized time translation symmetry.

- At quadratic order in σ_μ, the mixing Lagrangian is

$$\mathcal{L}^{(1)}_{\pi\sigma^2} = \omega_3^2\,\delta g^{00}(\sigma^0)^2 + \omega_4^2\,\delta g^{00}\sigma^\mu\sigma_\mu \tag{5.13}$$

$$\to\ -2(\omega_3^2-\omega_4^2)\dot{\pi}\sigma_0\sigma_0 - 2\omega_4^2 a^{-2}\dot{\pi}\sigma_i\sigma_i\,, \tag{5.14}$$

  where in the second line we have introduced the Goldstone and taken the decoupling limit. We see that, this time, the size of the cubic interaction $\dot{\pi}\sigma_i\sigma_i$ is independent from the quadratic mixing term.

Combining the above, we can write

$$\mathcal{L}^{(1)}_{\rm mix} = \frac{1}{a^2}\left(\rho_1\,\partial_i\pi_c\sigma_i + \frac{1}{\Lambda_1}\dot{\pi}_c\partial_i\pi_c\sigma_i + \lambda_1\,\dot{\pi}_c\sigma_i\sigma_i\right), \tag{5.15}$$

where $\pi_c \equiv f_\pi^2\pi$ is the canonically normalized Goldstone boson, and we defined

$$\rho_1 \equiv \frac{2\omega_1^3}{f_\pi^2}\,,\quad \Lambda_1 \equiv -\frac{f_\pi^2}{\rho_1}\,,\quad \lambda_1 \equiv -\frac{2\omega_4^2}{f_\pi^2}\,. \tag{5.16}$$

We note that ρ_1 and Λ_1 are correlated, since they are both determined by the parameter ω_1.

Spin-2. Next, we consider the mixing between a massive spin-2 field and the Goldstone boson.

- At linear order in $\sigma_{\mu\nu}$, the mixing Lagrangian is

$$\mathcal{L}^{(2)}_{\pi\sigma} = \tilde{\omega}_1^3\,\delta g^{00}\sigma^{00} + \tilde{\omega}_2^3\,(\delta g^{00})^2\sigma^{00} + \tilde{\omega}_3^2\,\delta K_{\mu\nu}\sigma^{\mu\nu} + \tilde{\omega}_4^2\,\delta g^{00}\delta K_{\mu\nu}\sigma^{\mu\nu}\,, \tag{5.17}$$

  where it was necessary to include higher-derivative operators to get the relevant interactions for the spatial components σ_{ij}. In the decoupling limit, the mixing with the Goldstone boson is

$$\begin{aligned}\mathcal{L}^{(2)}_{\pi\sigma} = \tilde{\omega}_1^3\left[-2\dot{\pi}\sigma_{00} + a^{-2}(\partial_i\pi)^2\sigma_{00} + 4a^{-2}\dot{\pi}\partial_i\pi\sigma_{0i}\right] - (5\tilde{\omega}_1^3 - 4\tilde{\omega}_2^3)\,\dot{\pi}^2\sigma_{00}\\ - \tilde{\omega}_3^2\,a^{-4}\partial_i\partial_j\pi\sigma_{ij} + 2\tilde{\omega}_4^2\,a^{-4}\dot{\pi}\partial_i\partial_j\pi\sigma_{ij} + \cdots\,.\end{aligned} \tag{5.18}$$

  We will focus on the traceless part of σ_{ij}, which we denote by $\hat{\sigma}_{ij}$. Only the cubic mixing $\dot{\pi}\partial_i\partial_j\pi\hat{\sigma}_{ij}$ will lead to the characteristic angular structure in the bispectrum. Since the quadratic mixing does not affect the angular structure, we will simply choose $\partial_i\partial_j\pi\sigma_{ij}$ as a representative example. Unlike the spin-1 case, the sizes of the quadratic and cubic mixing operators are controlled by two independent parameters, $\tilde{\omega}_3$ and $\tilde{\omega}_4$.

- At quadratic order in $\sigma_{\mu\nu}$, the mixing Lagrangian is

$$\mathcal{L}^{(2)}_{\pi\sigma^2} = \tilde{\omega}_5^2\,\delta g^{00}(\sigma^{00})^2 + \tilde{\omega}_6^2\,\delta g^{00}\sigma^{\mu\nu}\sigma_{\mu\nu}\,, \tag{5.19}$$

$$\to\ -(2\tilde{\omega}_5^2 + 2\tilde{\omega}_6^2)\,\dot{\pi}\sigma_{00}^2 - 2\tilde{\omega}_6^2\left[2a^{-2}\dot{\pi}\sigma_{0i}\sigma_{0i} + a^{-4}\dot{\pi}\sigma_{ij}\sigma_{ij}\right] + \cdots\,, \tag{5.20}$$

  where the last term in (5.20) will lead to the angular structure that we are interested in.

We will study the following mixing Lagrangian

$$\mathcal{L}^{(2)}_{\rm mix} = \frac{1}{a^4}\left(\rho_2\,\partial_i\partial_j\pi_c\hat{\sigma}_{ij} + \frac{1}{\Lambda_2^2}\,\dot{\pi}_c\partial_i\partial_j\pi_c\hat{\sigma}_{ij} + \lambda_2\,\dot{\pi}_c\hat{\sigma}_{ij}\hat{\sigma}_{ij}\right), \tag{5.21}$$

where we defined

$$\rho_2 \equiv -\frac{\tilde{\omega}_3^2}{f_\pi^2}\,, \quad \Lambda_2 \equiv \frac{f_\pi^2}{\sqrt{2}\tilde{\omega}_4}\,, \quad \lambda_2 \equiv -\frac{2\tilde{\omega}_6^2}{f_\pi^2}\,. \tag{5.22}$$

This is similar to the spin-1 mixing Lagrangian (5.15), except that the quadratic and cubic mixing parameters, ρ_2 and Λ_2, are now independent.

Spin-s. Performing the same analysis for a field with arbitrary spin $s > 2$, we find the following mixing Lagrangian

$$\mathcal{L}^{(s)}_{\rm mix} = \frac{1}{a^{2s}}\left(\rho_s\,\partial_{i_1\cdots i_s}\pi_c\hat{\sigma}_{i_1\cdots i_s} + \frac{1}{\Lambda_s^s}\dot{\pi}_c\partial_{i_1\cdots i_s}\pi_c\hat{\sigma}_{i_1\cdots i_s} + \lambda_s\,\dot{\pi}_c\hat{\sigma}^2_{i_1\cdots i_s}\right), \tag{5.23}$$

where $\partial_{i_1\cdots i_s} \equiv \partial_{i_1}\cdots\partial_{i_s}$. As in the case of spin-2, these interactions generically arise from independent operators, i.e. ρ_s, Λ_s, and λ_s are independent parameters.

The mixing in (5.23) can convert hidden non-Gaussianity in the σ-sector into visible non-Gaussianity in the π-sector. To allow for this possibility, we add cubic self-interactions to the action for σ, which schematically we can write as

$$a^{3s}\mathcal{L}^{(s)}_{\sigma^3} \equiv \begin{cases} \xi_s\,\hat{\sigma}\cdot\hat{\sigma}\cdot\hat{\sigma} & s \text{ even}, \\ \xi_s\,\hat{\sigma}\cdot\hat{\sigma}\cdot(\partial\hat{\sigma}) & s \text{ odd}, \end{cases} \tag{5.24}$$

with suitable symmetric contractions of spatial indices.

5.2.2 Couplings to the graviton

We will also be interested in the couplings between massive spinning fields and the graviton, γ_{ij}. For simplicity, we will only consider linear couplings to γ_{ij}, but the generalization to higher orders will essentially be straightforward.

Spin-1. The leading couplings to the graviton arise from

$$\mathcal{L}^{(1)}_{\gamma\pi\sigma} = \omega_5^2\,\delta g^{00} g^{\mu\nu}\nabla_\mu\sigma_\nu - \frac{m_1^2}{2}\sigma^{\mu\nu}\sigma_{\mu\nu}\,. \tag{5.25}$$

Note that, in our perturbative treatment, the on-shell conditions for σ_μ hold at the background level, so that $\bar{g}^{\mu\nu}\nabla_\mu\sigma_\nu = 0$. The first term in (5.25) is thus proportional to $g^{\mu\nu}\nabla_\mu\sigma_\nu = \delta g^{\mu\nu}\nabla_\mu\sigma_\nu$ and starts at cubic order in fluctuations. In terms of π and γ_{ij}, the mixing Lagrangian becomes

$$\mathcal{L}^{(1)}_{\gamma\pi\sigma} = \frac{1}{a^2}\frac{1}{M_{\rm pl}}\left(\tau_1\,\dot{\pi}_c\gamma^c_{ij}\partial_i\sigma_j + m_1^2\gamma^c_{ij}\sigma_i\sigma_j\right)\,, \tag{5.26}$$

where $\tau_1 \equiv 4\omega_5^2/f_\pi^2$ and $\gamma^c_{ij} \equiv \frac{1}{2}M_{\rm pl}\gamma_{ij}$ denotes the canonically normalized graviton. While the first term in (5.26) is higher order in derivatives than $\gamma_{ij}\partial_i\pi\sigma_j$, the latter only arises from the tadpole σ^0, and is therefore required to have a vanishing coefficient. Moreover, a quadratic mixing between the spin-1 field and the graviton is forbidden by kinematics: any such mixing will involve spatial gradients and hence must vanish because the graviton is transverse, $\partial_i\gamma_{ij} = 0$.

Spin-2. The couplings between a massive spin-2 field and the graviton follow from

$$\mathcal{L}^{(2)}_{\gamma\pi\sigma} = \tilde{\omega}_3^2\,\delta K_{\mu\nu}\sigma^{\mu\nu} + \tilde{\omega}_7^3\,\delta g^{00} g^{\mu\nu}\sigma_{\mu\nu} - \frac{m_2^2}{2}\sigma^{\mu\nu}\sigma_{\mu\nu}\,. \tag{5.27}$$

Note that we have already encountered the operator $\delta K_{\mu\nu}\sigma^{\mu\nu}$ in (5.17). In our perturbative treatment, the on-shell traceless condition holds at the background level, $\bar{g}^{\mu\nu}\sigma_{\mu\nu} = 0$, which implies that $g^{\mu\nu}\sigma_{\mu\nu} = \delta g^{\mu\nu}\sigma_{\mu\nu}$, so that the second term in (5.27) starts at cubic order in fluctuations. The cubic operator $\delta g^{00}\delta g^{\mu\nu}\nabla_\mu\sigma_{\nu 0}$ will not be considered, since its effects are indistinguishable from those of the first term in (5.26). Introducing π and γ_{ij}, the mixing Lagrangian becomes

$$\mathcal{L}^{(2)}_{\gamma\pi\sigma} = \frac{1}{a^2}\frac{1}{M_{\rm pl}}\left(\tilde{\rho}_2\,\dot{\gamma}^c_{ij}\hat{\sigma}_{ij} + \tau_2\,\dot{\pi}_c\gamma^c_{ij}\hat{\sigma}_{ij} + \frac{m_2^2}{a^2}\,\gamma^c_{ij}\hat{\sigma}_{ik}\hat{\sigma}_{kj}\right)\,, \tag{5.28}$$

where $\tilde{\rho}_2 \equiv -\rho_2 f_\pi^2$ and $\tau_2 \equiv 4\tilde{\omega}_7^3/f_\pi^2$. Note that we have only kept the spatial components in the coupling to the mass term. Unlike in the spin-1 case, there is now a quadratic mixing between the spin-2 field and the graviton, whose size is correlated with the π-σ mixing in (5.21). The other possible form of mixing $\gamma_{ij}\sigma_{ij}$ comes from the tadpole $\tilde{\sigma}$ and is thus absent.

Spin-s. For arbitrary spin $s > 2$, the leading interactions with the graviton and the Goldstone take the following form

$$\mathcal{L}^{(s)}_{\gamma\pi\sigma} = \frac{1}{a^{2s-2}}\frac{1}{M_{\rm pl}}\left(\tilde{\rho}_s\,\partial_{i_3\cdots i_s}\dot{\gamma}^c_{i_1 i_2}\hat{\sigma}_{i_1\cdots i_s} + \tau_s\,\gamma^c_{i_1 i_2}\partial_{i_3\cdots i_s}\pi_c\hat{\sigma}_{i_1\cdots i_s} + \frac{m_s^2}{a^2}\gamma^c_{i_1 j_1}\hat{\sigma}_{i_1\cdots i_s}\hat{\sigma}_{j_1\cdots i_s}\right)\,, \tag{5.29}$$

where $\tilde{\rho}_s \equiv -\rho_s f_\pi^2$. Again, we have only kept interactions that involve the purely spatial components of the field. In practice, there are other low-dimensional operators that can also contribute to the correlator $\langle\gamma\zeta\zeta\rangle$ with the same angular structure, such as $\dot{\gamma}_{ij}\sigma_{ij0\cdots 0}$.

5.2.3 Bounds on mixing coefficients

It is important to determine how large the mixing interactions of the previous section can be made while keeping the effective theory under theoretical control. In this section, we will discuss bounds arising from *i*) the requirement that the mixing interactions can be treated perturbatively and *ii*) the absence of superluminal propagation. Finally, we also consider what range of coefficients yields a technically natural effective field theory, in the sense of stability under radiative corrections. In §5.3, we will consider the implications of these constraints on the size of non-Gaussianities.

Perturbativity. We wish to treat the mixing interactions as perturbative corrections to the free-field actions for the Goldstone boson and the massive spinning fields. Since massive particles decay outside the horizon and oscillate rapidly inside the horizon, the dominant contributions to correlation functions will occur at horizon crossing of the Goldstone boson, corresponding to frequencies of order H. Consistency of the perturbative description therefore requires that the sizes of the mixing interactions at $\omega \sim H$ are smaller than the terms in the free-field actions. This puts constraints on the couplings in the mixing Lagrangians discussed in the previous section. For $c_\pi = 1$,[9] the criteria for a consistent perturbative treatment require little more than dimensional analysis. The dimensionful couplings of relevant interactions have to be less than H, while those of irrelevant interactions have to be greater than H. The dimensionless couplings of marginal interactions have to be less than unity. For example, for the couplings appearing in (5.21), we require $\{\rho_2, \lambda_2\} < 1$ and $\Lambda_2 > H$. Similar considerations apply for the couplings in (5.15) and (5.23). For $c_\pi \neq 1$, determining the perturbativity constraints on the mixing parameters requires a more careful analysis. Spatial gradients of the Goldstone mode are enhanced and the correlation functions can receive contributions from a second time scale, the time of crossing of the sound horizon. We will return to this complication in §5.3.

Superluminality. The breaking of time diffeomorphism invariance can modify the actions for spinning fields of §5.1 by introducing additional non-Lorentz-invariant interactions. For concreteness, we will confine our discussion in this subsection to the case of spin one, but we expect similar results to hold for higher spins. In unitary gauge, the most general quadratic action for a spin-1 field is

$$S_\sigma = \int \mathrm{d}^4x\sqrt{-g}\left[-\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{a_1}{2}F^{0\mu}F^0{}_\mu - \frac{1}{2}m^2(\sigma^\mu\sigma_\mu - a_0\,\sigma^0\sigma^0)\right], \tag{5.30}$$

where $F_{\mu\nu} \equiv \partial_\mu\sigma_\nu - \partial_\nu\sigma_\mu$, and the structure of the kinetic part is enforced by gauge invariance in the massless limit. The departure from the Lorentz-invariant action is characterized by the parameters a_0 and a_1, which lead to nontrivial sound speeds for the longitudinal mode, $c_0 \equiv 1/\sqrt{1+a_0}$, and for the transverse mode, $c_1 \equiv 1/\sqrt{1+a_1}$. To see this, we consider the on-shell equations of

[9] In this chapter and Appendices B and C only, we will use the notation c_π (instead of c_s) for the sound speed of the Goldstone boson to avoid possible confusion of identifying s as a spin label.

motion in the flat-space limit,

$$\ddot{\sigma}_0 - c_0^2\nabla^2\sigma_0 + m^2\sigma_0 = 0\,, \tag{5.31}$$

$$\ddot{\sigma}_i^{\rm t} - c_1^2\nabla^2\sigma_i^{\rm t} + m^2\sigma_i^{\rm t} = 0\,, \tag{5.32}$$

where $\nabla^2 \equiv \partial_i\partial_i$ is the spatial Laplacian, and $\sigma_i^{\rm t}$ denotes the transverse mode, $\partial_i\sigma_i^{\rm t} = 0$. The components of the spin-1 field propagate subluminally with no gradient instability as long as $\{a_0, a_1\} \geq 0$. A tachyonic instability is avoided for $m^2 > 0$.

Even if the spin-1 field propagates subluminally, the mixing with the Goldstone boson π can lead to superluminal propagation in the coupled system. Requiring the absence of superluminality imposes a constraint on the size of the quadratic mixing term in (5.15). To derive this constraint, we consider the on-shell equations of motion for the coupled system,

$$\ddot{\sigma}_0 - c_0^2\nabla^2\sigma_0 + m^2\sigma_0 = -c_0^2\rho_1 m^{-2}\nabla^2\dot{\pi}_c\,, \tag{5.33}$$

$$\ddot{\sigma}_i^{\rm t} - c_1^2\nabla^2\sigma_i^{\rm t} + m^2\sigma_i^{\rm t} = 0\,, \tag{5.34}$$

$$\ddot{\pi}_c - c_\pi^2\,\nabla^2\pi_c = -c_\pi^3\rho_1\,(c_0^{-2}\dot{\sigma}_0 - \rho_1 m^{-2}\nabla^2\pi_c)\,. \tag{5.35}$$

We see that the transverse mode does not mix with π, and hence its dispersion relation is unmodified. After diagonalizing the coupled π-σ system, the dispersion relations obeyed by the normal modes are

$$\omega_\pm^2 = \frac{1}{2}\left[\big(c_0^2 + c_\pi^2(1+2\delta^2)\big)k^2 + m^2 \pm \sqrt{\big[(c_0^2 - c_\pi^2)k^2 + m^2\big]^2 + 4c_\pi^4 k^4\delta^2(1+\delta^2)}\right], \tag{5.36}$$

where $\delta^2 \equiv c_\pi^3\rho_1^2/m^2$. For large k, subluminality implies the following constraint

$$\frac{\rho_1^2}{m^2} \;\leq\; \frac{1-c_0^2}{2-c_0^2}\,\frac{1-c_\pi^2}{c_\pi^3}\,. \tag{5.37}$$

Note that the mixing is required to vanish if either c_π or c_0 are equal to 1. (A similar result for the mixing with a scalar field was found in [118].) However, even a relatively small deviation of c_π and c_0 from 1 is sufficient to allow ρ_1 to be of order H (i.e. of order the maximal size allowed by perturbativity). For simplicity, we will therefore work with $c_0 = c_1 \approx 1$, but incorporating nontrivial sound speeds for the spin-1 field could be done straightforwardly using modifications of the mode functions given in Appendix B. Similarly, we will assume that all spinning fields obey a relativistic dispersion relation.

Naturalness. Finally, we will consider constraints arising from the radiative stability of the effective theory. This is more of a philosophic criterion rather than a strict consistency condition.

- Let us consider the interaction $\dot{\pi}_c\partial_{i_1\cdots i_s}\pi_c\sigma_{i_1\cdots i_s}$ in (5.23), suppressed by the scale Λ_s. At one

loop, this term generates the following correction to the non-Lorentz-invariant mass term,

$$\delta m^2_{\sigma_{i_1\cdots i_s}} \sim \frac{1}{\Lambda_s^{2s}} \int \mathrm{d}\omega\, \mathrm{d}^3k\, \frac{\omega^2 k^{2s}}{[c_\pi^{-3}(\omega^2 - c_\pi^2 k^2)]^2} \sim c_\pi^{3-2s} \frac{\Lambda^{2s+2}}{\Lambda_s^{2s}}\,. \tag{5.38}$$

Naturalness of the mass of the spinning field requires $\delta m^2_{\sigma_{i_1\cdots i_s}} \lesssim m_s^2 \sim H^2$. To estimate the size of (5.38), we take the cutoff of the π-loop to be of order the strong coupling scale of the Goldstone sector. For $c_\pi = 1$, we can, in principle, extend the π-loop up to the symmetry breaking scale, i.e. $\Lambda \sim f_\pi$, while, for $c_\pi \ll 1$, the effective theory of the Goldstone becomes strongly coupled at $\Lambda \sim f_\pi c_\pi$. We will therefore use $\Lambda \sim f_\pi c_\pi$ for all values of c_π. The condition for radiative stability then becomes

$$\left(\frac{H}{\Lambda_s}\right)^s \lesssim \left(\frac{(2\pi\Delta_\zeta)^{s+1}}{c_\pi^5}\right)^{1/2}. \tag{5.39}$$

Typically, this constraint requires Λ_s to be slightly larger than f_π.

- Next, we consider the interaction $\lambda_s \dot\pi_c \sigma^2_{i_1\cdots i_s}$ in (5.23). This leads to the following radiative correction to the non-Lorentz-invariant mass term,

$$\delta m^2_{\sigma_{i_1\cdots i_s}} \sim \lambda_s^2 \int \mathrm{d}\omega\, \mathrm{d}^3k\, \frac{\omega^2}{c_\pi^{-3}(\omega^2 - c_\pi^2 k^2)(\omega^2 - k^2 - m^2)} \sim c_\pi^2 \lambda_s^2 \Lambda^2\,. \tag{5.40}$$

Cutting off the loop at $\Lambda \sim f_\pi c_\pi$, we obtain the following constraint for radiative stability:

$$\lambda_s \lesssim \frac{(2\pi\Delta_\zeta)^{1/2}}{c_\pi^2}\,. \tag{5.41}$$

The interaction $\lambda_s \dot\pi_c \sigma^2_{i_1\cdots i_s}$ can also give a correction to the kinetic term for the Goldstone. However, on dimensional grounds, it is easy to see that this interaction only contributes a negligible correction to the sound speed of π.

- Lastly, the radiative correction generated by the cubic self-interaction of the spinning fields is

$$\delta m^2_{\sigma_{i_1\cdots i_s}} \sim \begin{cases} \xi_s^2 & s \text{ even}, \\ \xi_s^2 \Lambda^2 & s \text{ odd}, \end{cases} \tag{5.42}$$

where the couplings ξ_s are of dimensions zero and one for odd and even spins, respectively—cf. (5.24). For even spins, we only get a fixed finite correction to the mass term. Since we require $\xi_s < H$ for perturbative control, the loop contribution is guaranteed to be small. For odd spins, it is natural to take the cutoff for the σ-loop to be Λ_s. We then get

$$\xi_s^2 \lesssim \frac{H^2}{\Lambda_s^2} \lesssim \left(\frac{(2\pi\Delta_\zeta)^{s+1}}{c_\pi^5}\right)^{1/s}, \tag{5.43}$$

where we have used the naturalness constraint (5.39) on Λ_s in the second inequality.

5.3 Imprints on cosmological correlators

We will now compute the effects of massive particles with spin on the correlation functions of the Goldstone boson and the graviton. Following [102], we will study separately the contributions from local and non-local processes. Local processes are, by definition, those whose imprint can be mimicked by adding a local operator in the low-energy effective theory of the light fields alone. Non-local processes, on the other hand, capture particle production effects which cannot be mimicked by additional local operators. While the latter are the distinctive signature of extra particles during inflation, the amplitude of such effects is exponentially suppressed for masses above the Hubble scale. We will discover that the sound speed of the Goldstone boson plays a crucial role in controlling the relative size of the local and non-local processes.

5.3.1 $\langle\zeta\zeta\rangle$

Before discussing a potentially richer structure in the bispectra, we will gain some useful insights by first examining the effect of massive particles on the power spectrum $\langle\zeta\zeta\rangle$ (see Fig. 5.2). We will separate the correlation function into distinct contributions coming from local and non-local processes.

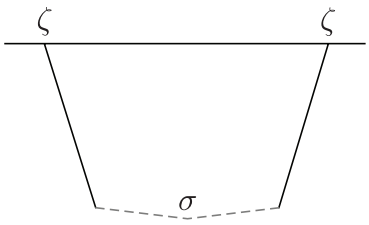


Figure 5.2: Tree-level diagram contributing to the two-point function $\langle\zeta\zeta\rangle$. The solid and dashed lines represent the curvature perturbation ζ and a massive spin-s field $\sigma_{i_1\cdots i_s}$, respectively.

Spin doesn't play a big role in the correction to the power spectrum, so for simplicity we will consider a minimally-coupled massive scalar field σ, whose two-point function in de Sitter space is

$$\langle\sigma_{\mathbf{k}}(\eta)\sigma_{-\mathbf{k}}(\eta')\rangle' = \frac{\pi}{4}H^2(\eta\eta')^{3/2}e^{-\pi\mu}H_{i\mu}(-k\eta)H^*_{i\mu}(-k\eta')\,, \tag{5.44}$$

where $H_{i\mu} \equiv H^{(1)}_{i\mu}$ is the Hankel function of the first kind and $\mu \equiv \sqrt{m^2/H^2 - 9/4}$. We will focus on massive particles belonging to the principal series, so that $\mu \geq 0$. The local part of the two-point function has support only at coincident points in position space, while the non-local part describes correlations over long distances. In Fourier space, the local and non-local parts of the two-point function are analytic and non-analytic in the momentum k, respectively. In the late-time limit, we have

$$\lim_{\eta,\eta'\to 0}\langle\sigma_{\mathbf{k}}(\eta)\sigma_{-\mathbf{k}}(\eta')\rangle'_{\rm local} = \frac{H^2(\eta\eta')^{3/2}}{4\pi}\Gamma(-i\mu)\Gamma(i\mu)\left[e^{\pi\mu}\left(\frac{\eta}{\eta'}\right)^{i\mu} + e^{-\pi\mu}\left(\frac{\eta}{\eta'}\right)^{-i\mu}\right], \tag{5.45}$$

$$\lim_{\eta,\eta'\to 0}\langle\sigma_{\mathbf{k}}(\eta)\sigma_{-\mathbf{k}}(\eta')\rangle'_{\text{non-local}} = \frac{H^2(\eta\eta')^{3/2}}{4\pi}\left[\Gamma(-i\mu)^2\Big(\frac{k^2\eta\eta'}{4}\Big)^{i\mu} + \Gamma(i\mu)^2\Big(\frac{k^2\eta\eta'}{4}\Big)^{-i\mu}\right]. \tag{5.46}$$

Away from the late-time limit, we use a series expansion of the Hankel function,

$$H_{i\mu}(x) = \sum_{n=0}^{\infty}\sum_{\pm} c_n^{\pm}(\mu,x)\,, \quad c_n^{\pm}(\mu,x) \equiv \pm\frac{(-1)^n}{n!}\frac{e^{\pi\mu(1\pm 1)/2}}{\sinh\pi\mu}\frac{(x/2)^{2n\pm i\mu}}{\Gamma(n+1\pm i\mu)}\,, \tag{5.47}$$

to decompose the two-point function (5.44) into its local and non-local pieces. Summing over the set of local and non-local contributions, the two-point function can be split into[10]

$$\langle\sigma_{\mathbf{k}}(\eta)\sigma_{-\mathbf{k}}(\eta')\rangle'_{\text{local}} = \frac{\pi}{4}\frac{H^2(\eta\eta')^{3/2}}{\sinh^2\pi\mu}\left[e^{\pi\mu}J_{i\mu}(-k\eta)J^*_{i\mu}(-k\eta') + e^{-\pi\mu}J^*_{i\mu}(-k\eta)J_{i\mu}(-k\eta')\right], \tag{5.48}$$

$$\langle\sigma_{\mathbf{k}}(\eta)\sigma_{-\mathbf{k}}(\eta')\rangle'_{\text{non-local}} = \frac{\pi}{4}\frac{H^2(\eta\eta')^{3/2}}{\sinh^2\pi\mu}\left[J_{i\mu}(-k\eta)J_{i\mu}(-k\eta') + J^*_{i\mu}(-k\eta)J^*_{i\mu}(-k\eta')\right], \tag{5.49}$$

where $J_{i\mu}$ denotes the Bessel function of the first kind.

To illustrate the distinct roles played by local and non-local parts, let us consider a coupling between π and σ of the form $\int \mathrm{d}^4x\, a^3\rho\dot\pi_c\sigma$ [91, 119]. At tree level, this produces the following correction to the power spectrum of ζ:

$$\langle\zeta_{\mathbf{k}}\zeta_{-\mathbf{k}}\rangle' = P_\zeta(k)\left[1 + \frac{c_\pi^2\rho^2}{H^2}(\mathcal{C}_1+\mathcal{C}_2)\right], \tag{5.50}$$

where

$$\mathcal{C}_1 \equiv \frac{\pi}{4}e^{-\pi\mu}\left|\int_0^\infty\frac{\mathrm{d}x}{x}\sqrt{x}H_{i\mu}(x)\,e^{ic_\pi x}\right|^2, \tag{5.51}$$

$$\mathcal{C}_2 \equiv -\frac{\pi}{2}e^{-\pi\mu}\,\mathrm{Re}\left[\int_0^\infty\frac{\mathrm{d}x}{\sqrt{x}}\,H_{i\mu}(x)e^{-ic_\pi x}\int_x^\infty\frac{\mathrm{d}y}{\sqrt{y}}\,H^*_{i\mu}(y)\,e^{-ic_\pi y}\right]. \tag{5.52}$$

These represent the non-time-ordered and time-ordered integrals that arise from performing the in-in calculation. The integral in (5.51) can be evaluated analytically to give

$$\mathcal{C}_1 = \frac{\pi^2}{2\cosh^2\pi\mu}\,{}_2F_1\left(\frac{1}{2}-i\mu\,,\frac{1}{2}+i\mu\,,1\,,\frac{1-c_\pi}{2}\right)^2, \tag{5.53}$$

where ${}_2F_1$ is the hypergeometric function. It is instructive to consider the $c_\pi\to 1$ and $c_\pi\to 0$ limits of this result:

- For $c_\pi = 1$, the hypergeometric function becomes unity, and (5.53) scales as $e^{-2\pi\mu}$ for large μ, as expected for the pair-production of massive particles.

[10] Away from the late-time limit, we are summing an infinite series of local/non-local elements for the propagator, in which case the distinction between the local and non-local parts is not as sharp. Nevertheless, we will see that this decomposition still leads to some useful insights.

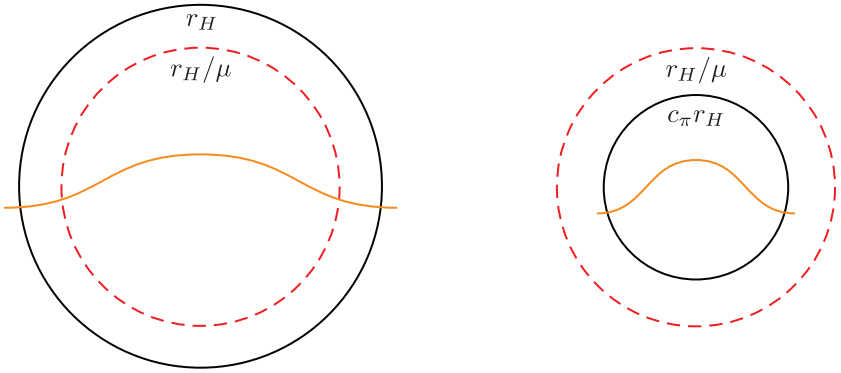


Figure 5.3: Pictorial representations of the horizon crossing scale of the Goldstone boson (solid) and the scale associated with the turning point in the dynamics of a massive particle (dashed), with the left (right) diagram corresponding to $c_\pi = 1$ ($c_\pi < \mu^{-1}$). The Hubble radius is denoted by $r_H \equiv H^{-1}$. We see that for $c_\pi < \mu^{-1}$ the horizon crossing of the Goldstone boson occurs before the turning point of the massive particles, while for $c_\pi = 1$ it occurs after.

- In the limit $c_\pi \to 0$, we instead get

$$\lim_{c_\pi \to 0} \mathcal{C}_1 = \frac{\pi^2}{2\cosh^2 \pi\mu} \times \frac{\pi}{\Gamma(\frac{3}{4} + \frac{i\mu}{2})^2 \, \Gamma(\frac{3}{4} - \frac{i\mu}{2})^2} \,, \tag{5.54}$$

  which scales as $e^{-\pi\mu}$ for large μ instead of the usual Boltzmann factor $e^{-2\pi\mu}$.

To see why the exponential suppression of $\mathcal{C}_1$ changes for $c_\pi \ll 1$, we need to consider the change in the dynamics of σ and π. There are two relevant timescales in the problem:

i) at the *turning point*, $|k\eta| \sim \mu$, the mode function of the massive particle starts to decay,

ii) at the *sound horizon crossing*, $|k\eta| \sim c_\pi^{-1}$, the Goldstone boson freezes.

For $c_\pi = 1$, event *i*) occurs before *ii*), while for $c_\pi < \mu^{-1}$, the order is reversed (see Fig. 5.3). As a consequence, the integral in (5.51) is dominated at the horizon crossing of π for $c_\pi = 1$, while it is dominated by the turning point of σ for $c_\pi < \mu^{-1}$. This is illustrated in Fig. 5.4, where we show the Wick-rotated integrand of the integral in (5.51) as a function of $x = |k\eta|$. A notable feature is the peak at $x \sim \mu$, which increases for small c_π. For $c_\pi = 1$, the turning point occurs before horizon crossing and the overlap between π and σ is suppressed. For $c_\pi < \mu^{-1}$, on the other hand, the turning point occurs after the freeze-out of the Goldstone, which enhances the feature at $x \sim \mu$. This qualitatively explains the boost in the amplitude of $\mathcal{C}_1$ for small c_π.

Let us now consider the time-ordered integral $\mathcal{C}_2$ in (5.52). For general c_π, it cannot be evaluated analytically, but some insights can be obtained by taking the limits $c_\pi \to 1$ and $c_\pi \to 0$:

- For $c_\pi = 1$, the above decomposition of the σ-propagator into local and non-local pieces leads to [91]

$$\mathcal{C}_2|_{\rm local} = \frac{e^{\pi\mu}}{8\sinh \pi\mu} \mathrm{Re}\left[\psi^{(1)}\left(\frac{3}{4} + \frac{i\mu}{2}\right) - \psi^{(1)}\left(\frac{1}{4} + \frac{i\mu}{2}\right)\right] - e^{-2\pi\mu}\,(i\mu \leftrightarrow -i\mu)\,, \tag{5.55}$$

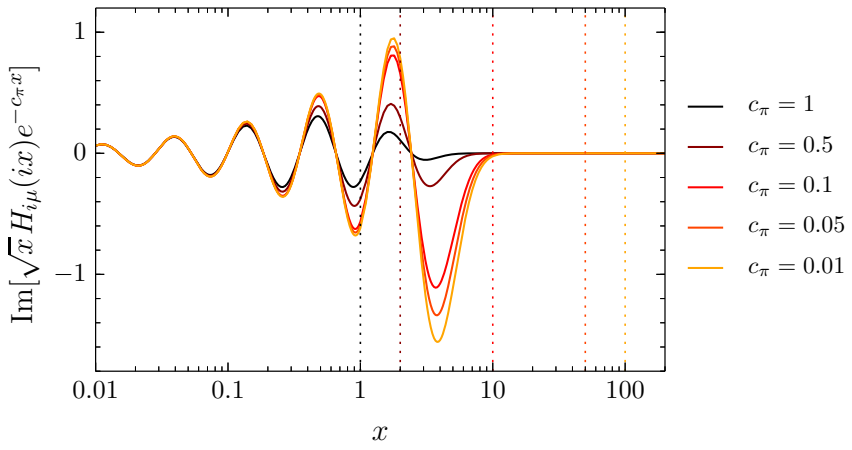


Figure 5.4: Wick-rotated integrand of the integral in (5.51) as a function of $x = |k\eta|$ and for $\mu = 5$. The vertical dotted lines indicate the times of sound horizon crossing of π, i.e. $x = c_\pi^{-1}$, for each value of c_π. The solid vertical line marks the turning point of σ, i.e. $x = \mu$.

$$\mathcal{C}_2|_{\text{non-local}} = 0\,, \tag{5.56}$$

where $\psi^{(1)}(z) = \partial_z^2 \ln\Gamma(z)$ is the polygamma function of order 1.[11] For large μ, the first term in (5.55) scales as μ^{-2}, which has a simple interpretation: a heavy field contributes to non-renormalizable interactions in the low-energy effective theory of the light fields with coefficients given by inverse powers of the mass of the heavy field. The second term is instead suppressed by $e^{-2\pi\mu}$, describing an effect which cannot be captured by a local Lagrangian of the light fields alone. Finally, we see that the non-local part of the σ-propagator does not contribute to the correction to the power spectrum.

- In the limit $c_\pi \to 0$, we find

$$\lim_{c_\pi\to 0} \mathcal{C}_2|_{\text{local}} = 0\,, \tag{5.57}$$

$$\lim_{c_\pi\to 0} \mathcal{C}_2|_{\text{non-local}} = -\frac{\pi^2}{2\cosh^2\pi\mu} \times \frac{\pi}{\Gamma(\frac{3}{4}+\frac{i\mu}{2})^2\,\Gamma(\frac{3}{4}-\frac{i\mu}{2})^2}\,. \tag{5.58}$$

We wish to highlight several features of this result. First, the local contribution to $\mathcal{C}_2$ vanishes. This follows from the simple fact that the Goldstone bosons become non-propagating when $c_\pi = 0$; hence, they can only communicate to each other through non-local effects. Second, the non-local contributions to $\mathcal{C}_1$ and $\mathcal{C}_2$ precisely cancel each other, implying that the correction to the two-point function (5.50) vanishes faster than c_π^2 in the limit $c_\pi \to 0$. This is the result of the cancellation between the contributions from the forward and back-

[11] There are also logarithmically divergent terms within the separate integrals for the local and non-local parts. These are the result of an imperfect decomposition between the two terms away from the late-time limit and the fact that we are integrating over time. However, these terms exactly cancel in the sum over all contributions, so that the final result remains finite.

ward branches of the integration contour. A way to see this is to drop the exponentials in c_π in (5.51) and (5.52), and notice that $\mathcal{C}_1 + \mathcal{C}_2$ is now proportional to the sum of all Schwinger-Keldysh propagators for the σ field; these propagators add up to zero. Of course, for small (but finite) c_π, we do not expect this cancellation to be exact.

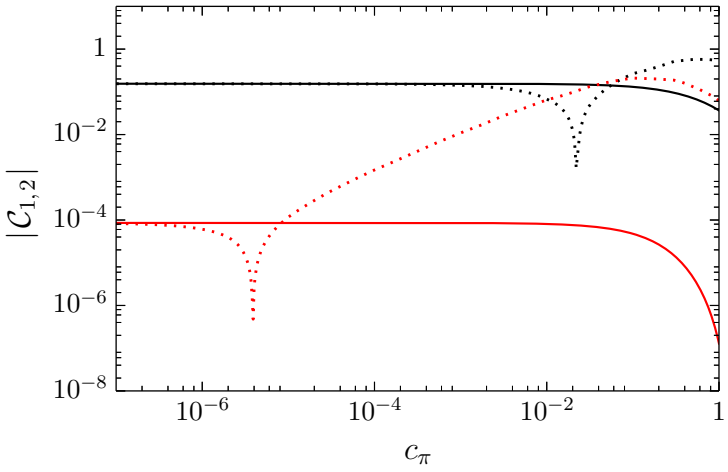


Figure 5.5: $\mathcal{C}_1$ and $\mathcal{C}_2$ as functions of c_π for $\mu = 1$ (black) and $\mu = 3$ (red). The solid and dotted lines denote $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively.

To understand how the result for general c_π interpolates between these two limiting behaviours, we evaluate $\mathcal{C}_2$ numerically. Figure 5.5 shows the analytical result (5.53) for $\mathcal{C}_1$ and a numerical computation of $\mathcal{C}_2$, both as functions of c_π. As c_π is lowered, the exponential dependence on μ for both of the integrals changes. For $\mathcal{C}_1$, this happens relatively quickly when $c_\pi \lesssim \mu^{-1}$, agreeing with the intuition that reversing the ordering of the turning point of σ and the horizon exit of π changes the solution qualitatively. On the other hand, the transition in the exponential behavior for $\mathcal{C}_2$ only occurs for very small c_π, typically much smaller than the lower limit required for perturbative control of the non-renormalizable interaction $\dot{\pi}(\partial_i\pi)^2$ associated with c_π. This implies that, while for $\mathcal{C}_2$ the dependence on $\mu > 1$ will not change much within the allowed range of $c_\pi > 10^{-2}$, the exponential suppression $e^{-2\pi\mu}$ of $\mathcal{C}_1$ can be reduced to $e^{-\pi\mu}$ when $c_\pi < \mu^{-1}$.

5.3.2 $\langle\zeta\zeta\zeta\rangle$

Next, we consider the imprints of massive spinning particles on the three-point function $\langle\zeta\zeta\zeta\rangle$. In single-field inflation, a long-wavelength curvature perturbation locally corresponds to a rescaling of the background experienced by short-wavelength fluctuations. As a result, the bispectrum $\langle\zeta\zeta\zeta\rangle$ satisfies a consistency relation for the squeezed limit [25, 62, 77]. In particular, we can write a Taylor expansion around the squeezed limit,

$$\lim_{k_1 \ll k_3} \langle \zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle' = P_\zeta(k_1)P_\zeta(k_3)\sum_{n=0}^{\infty} b_n \left(\frac{k_1}{k_3}\right)^n , \tag{5.59}$$

where the leading coefficient is determined by the tilt of the scalar power spectrum, $b_0 = -(n_s - 1)$. The consistency condition furthermore fixes the coefficient of the linear term, b_1, and partially constrains higher-order coefficients [26, 60, 120–123]. Since the contributions coming from b_0 and b_1 cannot be measured by a local observer [124, 125], any physical effect will only appear at order $(k_1/k_3)^2$ [126]. A crucial consequence of the consistency relation is the existence of the Taylor expansion (5.59) with only integer powers of k_1/k_3. Interesting non-analytic deviations from (5.59), however, are known to arise in the presence of additional fields. For example, fractional powers $(k_1/k_3)^\nu$ can be present in quasi-single-field inflation [84], with scaling $0 < \nu \leq 3/2$ in between the fully constrained $(k_1/k_3)^0$ term and the physical $(k_1/k_3)^2$ term. In this section, we will study such deviations for additional fields that carry spin.

Figure 5.6 shows all possible tree-level contributions to $\langle\zeta\zeta\zeta\rangle$. The three diagrams share many qualitative features, so to avoid repetition we will mostly concentrate on the analysis of the single-exchange diagram [(a)], and only highlight the differences that arise for the other two diagrams [(b,c)]. We will split the contributions to the bispectrum into its local and non-local parts. To avoid confusion with the alternative usage of "local non-Gaussianity", we will refer to these contributions as *analytic* and *non-analytic*, respectively. (This terminology highlights the distinctive scaling behavior in the squeezed limit.) Although we will ultimately be interested in the behavior of the latter, the observability of the signal will depend on the full bispectrum, so we will present the results for both types of contributions. As before, we will mostly restrict our analysis to particles in the principal series, with $\mu_s \geq 0$.

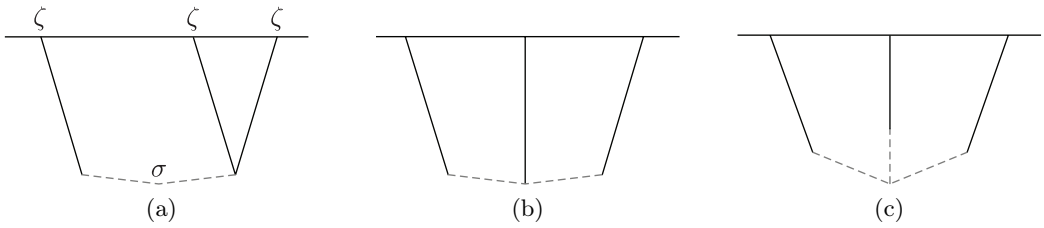


Figure 5.6: Tree-level diagrams contributing to $\langle\zeta\zeta\zeta\rangle$. The solid and dashed lines represent the curvature perturbation ζ and a spinning field $\sigma_{i_1\cdots i_s}$, respectively.

Single-exchange diagram. We will first compute the bispectrum associated with the exchange of a single spinning field (Fig. 5.6a). The relevant interaction Lagrangian is (cf. (5.23))

$$\mathcal{L}_I = \frac{1}{a^{2s}}\left(\rho_s\,\partial_{i_1\cdots i_s}\pi_c\hat{\sigma}_{i_1\cdots i_s} + \frac{1}{\Lambda_s^s}\dot{\pi}_c\partial_{i_1\cdots i_s}\pi_c\hat{\sigma}_{i_1\cdots i_s}\right). \tag{5.60}$$

We obtain the following bispectrum

$$\frac{\langle\zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle'}{\Delta_\zeta^4} = \alpha_s\Delta_\zeta^{-1}\times P_s(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3)\times\mathcal{I}^{(s)}(\mu_s, c_\pi, k_1, k_2, k_3) + 5\text{ perms.}, \tag{5.61}$$

where an integral representation of the function $\mathcal{I}^{(s)}$ is given in Appendix C. The dimensionless

parameters α_s are

$$\alpha_s \equiv \frac{1}{c_\pi^{s-3/2}} \frac{\rho_s}{H^{2-s}} \left(\frac{H}{\Lambda_s}\right)^s . \tag{5.62}$$

where we have included powers of c_π in α_s, so that the function $\mathcal{I}^{(s)}$ does not scale parametrically with c_π. By this we mean that $\mathcal{I}^{(s)}$ saturates to a constant value in the limit of small c_π, similar to the behavior of the integrals (5.51) and (5.52). The requirement of a perturbative treatment of non-Gaussianity implies that

$$\alpha_s < 1 . \tag{5.63}$$

Notice that we have a stronger perturbativity condition on the bare parameters ρ_s and Λ_s for subluminal c_π, which takes into account the fact that the dispersion relation, $\omega = c_\pi k$, is non-relativistic.

Size of NG. It is customary to quantify the size of non-Gaussianity by the parameter

$$f_{\rm NL} \equiv \frac{5}{18} \frac{\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle'}{P_\zeta^2(k)} , \tag{5.64}$$

where the bispectrum is evaluated in the equilateral configuration, $k_1 = k_2 = k_3 \equiv k$. The overall size of the non-Gaussianity can only partially be read off from the prefactor in (5.61), since there is a hidden dependence on μ_s in the function $\mathcal{I}^{(s)}$. An estimate for the size of non-Gaussian signal is

$$f_{\rm NL} \sim f(\mu_s) \alpha_s \Delta_\zeta^{-1} , \tag{5.65}$$

where $f(\mu_s)$ gives the appropriate mass suppressions for the analytic and non-analytic parts[12,13]

$$f(\mu_s) \equiv \begin{cases} \mu_s^{-2} & \text{analytic,} \\ e^{-\pi\mu_s} & \text{non-analytic,} \quad c_\pi = 1, \\ e^{-\pi\mu_s/2} & \text{non-analytic,} \quad c_\pi < \mu_s^{-1}. \end{cases} \tag{5.66}$$

We see that there are two sources of suppression in the signal: the mass suppression as a function of μ_s and the mixing efficiency parameterized by α_s. At the same time, there is a $\Delta_\zeta^{-1} \approx 10^5$ enhancement in the signal. It is this large factor that can, in principle, allow for observable non-Gaussianity even in the presence of the above suppressions. The size of the analytic part is only power-law suppressed and thus dominates for large mass, whereas the non-analytic part is always accompanied by an exponential Boltzmann suppression. For $c_\pi = 1$, the dominant non-analytic term is suppressed by $e^{-\pi\mu_s}$. As explained in [102, 127], this arises from the quantum interference of two wavefunctions: $\Psi[2\sigma] \propto e^{-\pi\mu_s}$ for pair-produced massive particles and $\Psi[0\sigma]$ for the wave-

[12] The displayed μ_s scalings are the asymptotic behaviors for large μ_s. There is also a polynomial dependence in μ_s for the non-analytic part which competes with the exponential suppression for intermediate values of μ_s.

[13] It is more useful to consider this separation of the signal in the squeezed limit, where the distinction between the analytic and non-analytic parts becomes sharp, as we will show below.

function involving no spontaneously created massive particles.[14] This interference contribution is larger than the probability of pair-producing massive particles, which is $|\Psi[2\sigma]|^2 \propto e^{-2\pi\mu_s}$. For $c_\pi < \mu_s^{-1} \ll 1$, the exponential suppression of the non-analytic part changes to $e^{-\pi\mu_s/2}$. We have already encountered this phenomenon in §5.3.1: for small c_π the horizon crossing of the Goldstone boson occurs *before* the turning point in the mode function of the massive particle. In this case, we are picking out the contribution of the wavefunction for a pair of massive particles not in the late-time limit, but at the turning point, which comes with a different exponential factor.

In §5.2.3, we derived naturalness constraints on the mixing parameters of the effective theory. For the parameter α_s in (5.62), the radiative stability of the mass (5.39) implies

$$\alpha_s \lesssim \left(\frac{2\pi\Delta_\zeta}{c_\pi^2}\right)^{(s+1)/2} . \tag{5.67}$$

For $c_\pi = 1$, this naturalness constraint is rather strong, implying that large non-Gaussianity, $f_{\rm NL} > 1$, is only possible if additional physics, such as supersymmetry, stabilizes the mass of the spinning particle, or if the mass term is fine-tuned. For $c_\pi \neq 1$, the current observational constraint $c_\pi \geq 0.024$ [35] still allows for naturally large non-Gaussianity, although within a rather narrow range in the small c_π regime.

Some comments are in order concerning the observability of particles with odd spins. In [102], it was shown that the diagram due to the exchange of an odd-spin particle vanish exactly at leading order in the weak breaking of conformal symmetry. At subleading orders, however, there are non-zero contributions from odd-spin particles.[15] When conformal symmetry is strongly broken, these terms become as important as the leading ones, and odd-spin particles can leave an equally relevant imprint on the correlation function $\langle\zeta\zeta\zeta\rangle$. Nevertheless, the amplitude of the bispectrum with an intermediate spin-1 particle is

$$f_{\rm NL} \sim f(\mu_1)\sqrt{c_\pi}\,\frac{\rho_1^2}{H^2} . \tag{5.70}$$

As long as the mixing is perturbative, $\rho_1 < H$, this non-Gaussianity is constrained to be less than

[14] Potential tests of the quantum nature of cosmological fluctuations have also been discussed in [128–132].

[15] When the approximate conformal invariance is valid, we can think of this in terms of correlation functions of the inflaton $\phi(t,\mathbf{x}) = \bar\phi(t) + \delta\phi(t,\mathbf{x})$, where $\dot{\bar\phi} \neq 0$ characterizes the weak breaking of conformal symmetry. The leading three-point function for the inflaton perturbation $\delta\phi$ will be given by the four-point function of Φ with one external leg set to $\dot{\bar\phi}$:

$$\langle\delta\phi\delta\phi\delta\phi\rangle' \propto \langle\delta\phi\delta\phi\sigma\rangle'\langle\sigma\delta\phi\dot{\bar\phi}\rangle' \propto \dot{\bar\phi}\langle\delta\phi\delta\phi\sigma\rangle'\langle\sigma\delta\phi\rangle'_{\rm inf} , \tag{5.68}$$

where $\langle\cdots\rangle_{\rm inf}$ denotes an inflationary correlation function which breaks conformal symmetry [133]. However, in the conformally symmetric case, $\langle\delta\phi\delta\phi\sigma\rangle$ vanishes when σ has odd spin [134]. The next-to-leading order result is given by the six-point function with three insertions of $\dot{\bar\phi}$,

$$\langle\delta\phi\delta\phi\delta\phi\rangle' \propto \langle\dot{\bar\phi}\delta\phi\delta\phi\sigma\rangle'\langle\sigma\delta\phi\dot{\bar\phi}^2\rangle' \propto \dot{\bar\phi}^3\langle\delta\phi\delta\phi\sigma\rangle'_{\rm inf}\langle\sigma\delta\phi\rangle'_{\rm inf} . \tag{5.69}$$

This is suppressed by an additional factor of $\dot{\bar\phi}^2$, but notice that the correlator $\langle\delta\phi\delta\phi\sigma\rangle_{\rm inf}$, not being constrained by conformal symmetry, does not have to vanish for odd-spin σ.

unity. We see that a spin-1 particle cannot lead to large non-Gaussianity because the size of the cubic vertex in (5.15) is tied to the quadratic mixing coefficient. In fact, the same reasoning applies to the coupling to scalar fields, which is why the single-exchange diagram has been neglected in the context of quasi-single-field inflation [84, 118]. This fact, however, is only tied to spins zero and one, and the bispectrum does not have to be suppressed for higher odd-spin particles. Moreover, we will see that the diagrams involving more than a single exchange can allow for observable non-Gaussianity, even for spin one.

Shape of NG. Before considering the general shape of the bispectrum, we will first analyze the singular behavior of the bispectrum in the squeezed limit, mainly concentrating on particles with even spins. We will quote results whose derivations can be found in Appendix C.

- For the analytic part of the bispectrum, we get

$$\lim_{k_1\ll k_3}\langle\zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle' \propto \frac{1}{k_1^3k_3^3}\left(\frac{k_1}{k_3}\right)^2 . \tag{5.71}$$

  We see that the local effects of massive particles lead to the same squeezed limit behavior as for single-field inflation—cf. (5.59). This is expected, since the massive particle can be integrated out for large μ_s, producing an effective cubic vertex of the form $\dot\pi(\hat\partial_{i_1\cdots i_s}\pi)^2$. The presence of extra particles therefore cannot be inferred from this part of the signal. Although the analytic part of the non-Gaussianity is itself interesting and more information can be gained by analyzing its shape for general momentum configurations, we have to treat it as an effective noise in the squeezed limit as far as the detection of extra particles is concerned.

- For the non-analytic part, we find

$$\lim_{k_1\ll k_3}\langle\zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle' \propto \frac{1}{k_1^3k_3^3}\left(\frac{k_1}{k_3}\right)^{3/2} P_s(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3)\cos\left[\mu_s\ln\left(\frac{k_1}{k_3}\right)+\phi_s\right], \tag{5.72}$$

  where the phase ϕ_s is uniquely fixed in terms of μ_s and c_π (see Appendix C). The suppression factor $(k_1/k_3)^{3/2}$ represents the dilution of the physical particle number density due to the volume expansion. This non-analytic scaling in the squeezed limit, corresponding to an intrinsically non-local process, cannot be mimicked by a local interaction within the effective theory of a single field. The signal contains oscillations in $\ln(k_1/k_3)$, with a frequency set by the mass of the spinning particle. This is due to the fact that the wavefunctions of massive particles oscillate logarithmically in time on superhorizon scales. The spin of the extra particle is reflected in the angular dependence, which is given by a Legendre polynomial of the angle between the short and long momenta.

  The above behavior applies for particles in the principal series, for which $\mu_s \geq 0$. For particles in the complementary series, μ_s becomes imaginary and the scaling of the squeezed

bispectrum changes to

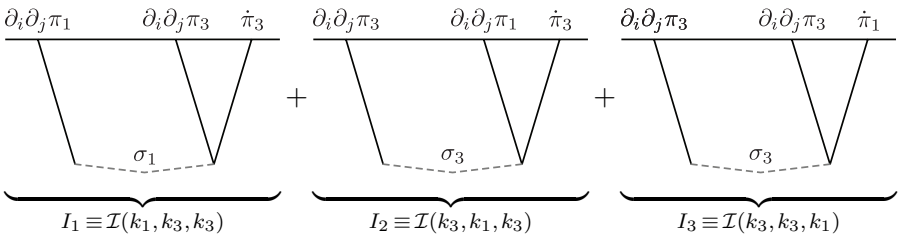
$$\lim_{k_1 \ll k_3} \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle' \propto \frac{1}{k_1^3 k_3^3} \left(\frac{k_1}{k_3} \right)^{3/2-\nu_s} P_s(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3) \,, \tag{5.73}$$

with $\nu_s \equiv -i\mu_s$ real. For $s \geq 2$, unitarity implies $\nu_s \in [0, 1/2)$, and the singular behavior in the squeezed limit is suppressed by at least k_1/k_3 compared to the leading term in the consistency relation (5.59).

The fact that the polarization tensors corresponding to odd-spin particles are odd under the exchange of two short momenta, together with momentum conservation, implies that the signal will gain an extra suppression factor of k_1/k_3 in the squeezed limit compared to the case of even spin. This means that the non-analytic part due to odd-spin particles scales as $(k_1/k_3)^{5/2}$ in the squeezed limit, which is more suppressed than the analytic part that scales as $(k_1/k_3)^2$. The latter, however, have an analytic dependence on momenta and correspond to local correlations in position space. Thus, the presence of odd-spin particles could still be inferred from long-distance correlations, although it might be subdominant compared to other non-local effects.

It is possible to understand the different behaviors in the squeezed limit intuitively. For concreteness, let us consider the exchange of a spin-2 field involving the interactions $\partial_i\partial_j\pi\hat{\sigma}_{ij}$ and $\dot{\pi}\partial_i\partial_j\pi\hat{\sigma}_{ij}$. The bispectrum in the isosceles-triangle configuration, $k_2 = k_3$, consists of three different permutations of the external legs:

$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle' \propto$



, (5.74)

where $\pi_n \equiv \pi(\mathbf{k}_n)$, $\sigma_n \equiv \sigma_{ij}(\mathbf{k}_n)$ and $\mathcal{I}(k_1, k_2, k_3) \propto P_2(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3)\, \mathcal{I}^{(2)}(\mu_2, c_\pi, k_1, k_2, k_3)$. The non-analytic squeezed limit (5.72) arises if the massive exchange particle carries the soft momentum, corresponding to the contribution I_1 in (5.74). This describes a non-local conversion process between the massive particle and the Goldstone boson between the horizon crossing times of the long and short modes. However, when the mass of the extra particle becomes large, it can be integrated out and the same effect will be captured by a local vertex. In that case, the bispectrum should become indistinguishable from that produced by a self-interaction of π, namely $\dot{\pi}(\hat{\partial}_{ij}\pi)^2$. Note, in particular, that this interaction is symmetric under the exchange of the momenta associated with the two external legs with spatial gradients. This allows us to gauge how well the interaction is approximated by a local vertex by looking at how similar the terms I_1 and I_2 are. Both I_2 and I_3 will lead to analytic scalings in the squeezed limit, where the latter produces (5.71).

To analyze the shape of the bispectrum for general momentum configurations, we proceed numerically. For this purpose, it is convenient to define a dimensionless shape function

$$S(k_1, k_2, k_3) \equiv \frac{k_1^2 k_2^2 k_3^2}{(2\pi)^4} \frac{\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle'}{\Delta_\zeta^4} \,. \tag{5.75}$$

Figure 5.7 shows two-dimensional projections of the shape function for spin 2 with $\mu_2 = 3, 5, 7$ and $c_\pi = 1, 0.1$ in the isosceles-triangle configuration, $k_2 = k_3$. For the reasons explained in the previous paragraph, in Fig. 5.7 we have shown separately the shape functions corresponding to the contributions I_1 and I_2 in (5.74).[16] As anticipated, these contributions exhibit different scalings in the squeezed limit. The plots show $(k_3/k_1) \times S$, so that the analytic part is expected to approach a constant in the squeezed limit, while the non-analytic part grows as $(k_1/k_3)^{-1/2}$ for small k_1. We see that the shape of the bispectrum is mostly governed by the non-analytic part for small mass, giving almost pure oscillations. The amplitude of this effect, however, goes as $e^{-\pi\mu_2}$ for large μ_2. The analytic part, being power-law suppressed, therefore takes over in size as the mass increases, and the shape approaches the equilateral form in the limit of large mass. For large mass, it is clear that the non-Gaussianity is dominated by the analytic piece, with small oscillations coming from the non-analytic piece indicating the presence of a heavy mode. For $c_\pi = 1$, the contributions I_1 and I_2 lead to the same shape of the bispectrum for $\mu_2 = 7$, indicating that the π-σ conversion process has become local. Indeed, in this case the bispectrum precisely overlaps with that of the local interaction $\dot{\pi}(\hat{\partial}_{ij}\pi)^2$. For small c_π, we have argued that the exponential suppression is instead $e^{-\pi\mu_2/2}$. The fact that we see more pronounced oscillations for $c_\pi = 0.1$ is a consequence of this. Moreover, for small c_π, the shapes of the contributions I_1 and I_2 are no longer identical. Note that, in order for the massive particle to be integrated out, the time of its turning point should be much earlier than the time at which the Goldstone boson crosses its sound horizon, which translates into the condition $c_\pi > \mu_2^{-1}$. For $c_\pi = 0.1$, this condition is not satisfied for the list of mass parameters used in the figure, which is the reason why we do not see the convergence to the local behavior. We have checked that the convergence does indeed happen for sufficiently large $\mu_2 > c_\pi^{-1}$.

Another characteristic of the signal due to spinning particles is its angular dependence. Figure 5.8 shows the shape function of the total signal as a function of the angle between the long and short momenta, $\theta \equiv \cos^{-1}(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3)$, for a range of momentum configurations with fixed k_1/k_3. For visualization purposes, the plot has been rescaled so that it can be compared more easily to the Legendre polynomial $P_2(\cos\theta)$. As expected, the angular dependence converges to the pure Legendre behavior as the triangle becomes squeezed, $k_1/k_3 \ll 1$. The non-zero offset is due to the analytic part which doesn't carry any angular dependence. We also see that the angular dependence deviates from the pure Legendre behavior as the triangle approaches the equilateral shape. Still, the peak around the flat triangle ($\theta = 180°$) remains prominent regardless of the

[16] We have omitted I_3 in the plots, which has the same analytic scaling as in (5.71) and thus shows no interesting features. Of course, this contribution should be added in order to obtain the full bispectrum.

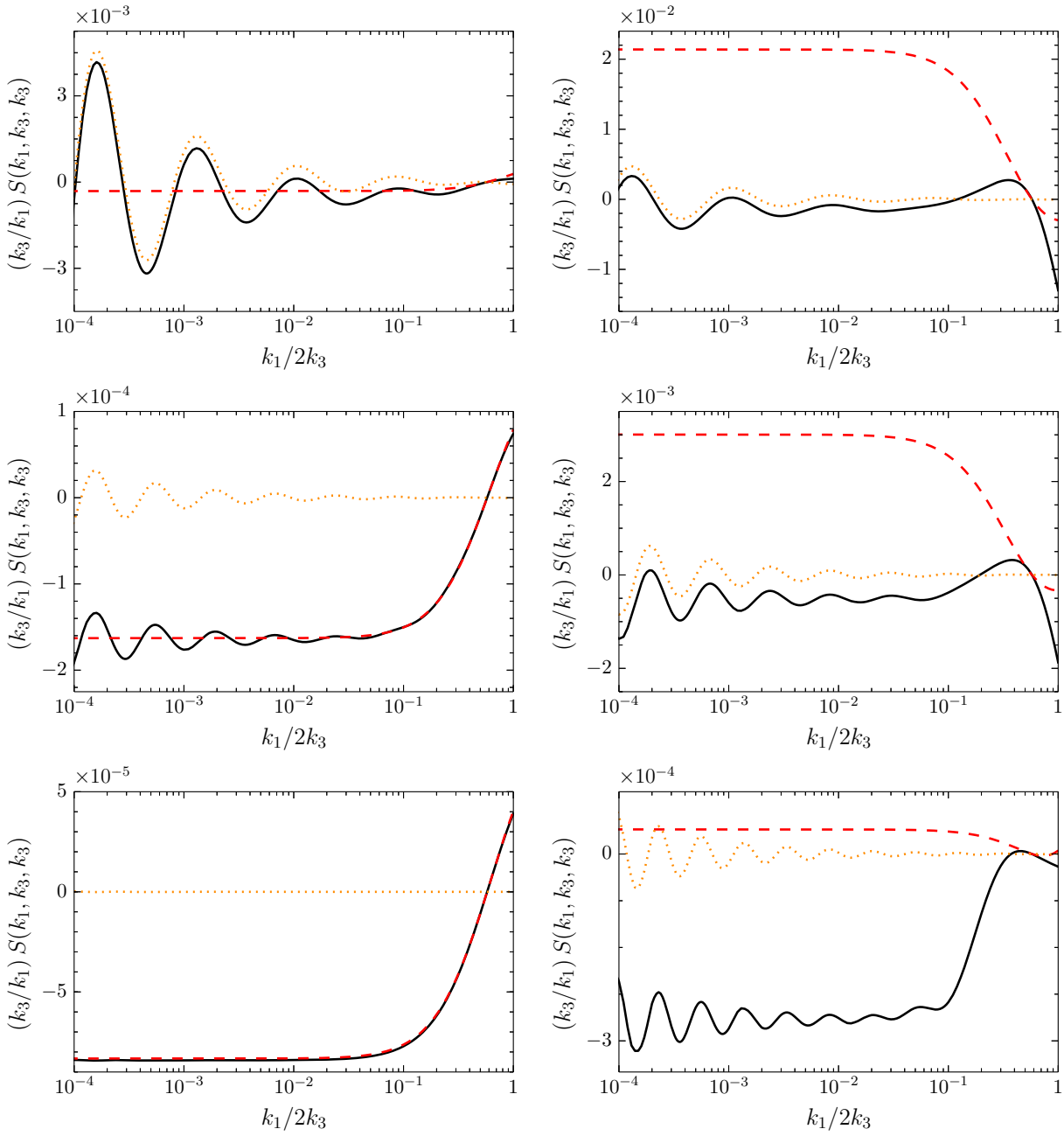


Figure 5.7: Shape functions (in units of $\alpha_2 \Delta_\zeta^{-1}$) for the spin-2 single-exchange diagram in the isosceles-triangle configuration, $k_2 = k_3$, with $\mu_2 = 3$ (*top*), $\mu_2 = 5$ (*middle*), and $\mu_2 = 7$ (*bottom*) for $c_\pi = 1$ (*left*) and $c_\pi = 0.1$ (*right*). The solid and dashed lines correspond to the numerical results for the parts of the signal corresponding to the terms I_1 and I_2 in (5.74), respectively. Not shown in the figure is the term I_3, which produces an analytic scaling in the squeezed limit and is needed to obtain the full bispectrum. Convergence of the solid and dashed lines indicates that the same effect can be captured by a local vertex $\dot{\pi}(\hat{\partial}_{ij}\pi)^2$ in the single-field EFT. The dotted lines show the analytical predictions for the non-analytic part.

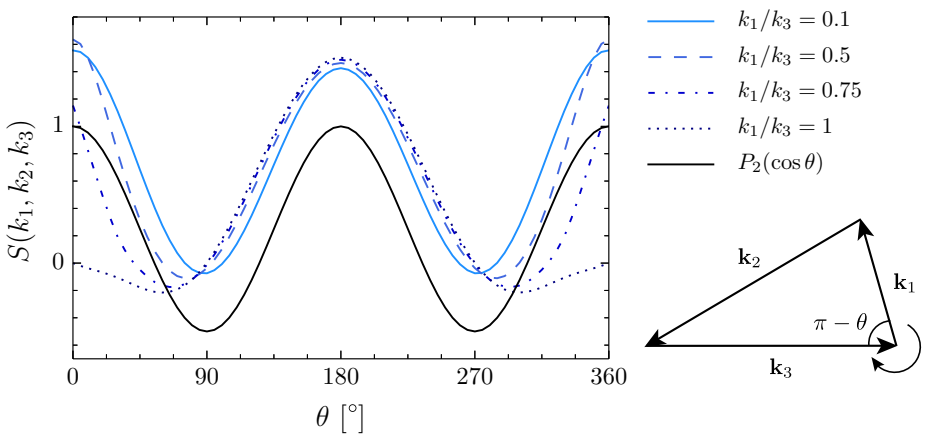


Figure 5.8: Shape functions of the spin-2 single-exchange diagram with $\mu_2 = 5$ and $c_\pi = 1$ as a function of the base angle $\theta = \cos^{-1}(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3)$ for fixed ratios of k_1/k_3. For easy comparison, the plot has been normalized such that the height difference between $\theta = 90°$ and $\theta = 180°$ of each curve is fixed to $3/2$.

momentum configuration. This suggests that the information about a particle's spin can still be inferred without necessarily going to very squeezed momentum configurations, since the width of the peak is still fixed by the polarization tensor of the spinning particle. This property can serve as an important tool for detecting odd-spin particles, whose signal in the squeezed limit necessarily gains an extra suppression in the soft momentum.

Double-exchange diagram. The bispectrum for the double-exchange diagram (Fig. 5.6b) is

$$\frac{\langle \zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle'}{\Delta_\zeta^4} = \tilde{\alpha}_s \Delta_\zeta^{-1} \times P_s(\hat{\mathbf{k}}_2 \cdot \hat{\mathbf{k}}_3) \times \mathcal{J}^{(s)}(\mu_s, c_\pi, k_1, k_2, k_3) + 5 \text{ perms.}\,, \tag{5.76}$$

where the function $\mathcal{J}^{(s)}$ is given explicitly in Appendix C, and the dimensionless parameters $\tilde{\alpha}_s$ are

$$\tilde{\alpha}_s \equiv \lambda_s \left(\frac{\rho_s}{H^{2-s}}\right)^2 < 1\,, \tag{5.77}$$

with λ_s and ρ_s defined in (5.23).

The size of the non-Gaussianity associated with the double-exchange diagram can be read off from (5.65) after replacing α_s by $\tilde{\alpha}_s$, but with an extra suppression of μ_s^{-2}, because this diagram involves another particle exchange. The condition for radiative stability (5.41) imposes the following upper limit on the size of the mixing parameter:

$$\tilde{\alpha}_s \lesssim \frac{(2\pi\Delta_\zeta)^{1/2}}{c_\pi^2}\,. \tag{5.78}$$

Notice that this is a much weaker constraint than the corresponding constraint for the single-exchange diagram (5.67). Depending on the values of c_π, this may or may not be stronger than the requirement for perturbativity, $f_{\rm NL} < \Delta_\zeta^{-1}$. This diagram can thus naturally produce

detectable levels of non-Gaussianity, even for $c_\pi = 1$.

Note that this diagram involves two π-σ conversion processes. When one of these processes becomes local, the double-exchange diagram becomes essentially equivalent to the single-exchange diagram. This can be seen by replacing one of the $\hat{\sigma}_{i_1\cdots i_s}$ legs in the cubic vertex $\dot{\pi}\hat{\sigma}^2_{i_1\cdots i_s}$ by $\partial_{i_1\cdots i_s}\pi$, after which the interaction becomes the same as the cubic vertex for the single-exchange diagram. As a result, the squeezed-limit behavior for this diagram is essentially the same as that of the single-exchange diagram. Hence, the analysis we have presented for the single-exchange diagram applies also to the double-exchange diagram.

Triple-exchange diagram. As indicated in (5.23), there is a slight difference between the form of the cubic self-interaction of spinning fields for even and odd spins. For concreteness, we will present the results for the former. The bispectrum for the triple-exchange diagram (Fig. 5.6c) is

$$\frac{\langle\zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle'}{\Delta^4_\zeta} = \hat{\alpha}_s\Delta_\zeta^{-1}\times P(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2,\hat{\mathbf{k}}_3)\times\mathcal{K}^{(s)}(\mu_s,c_\pi,k_1,k_2,k_3)+5\text{ perms.}\,, \tag{5.79}$$

where $P(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2,\hat{\mathbf{k}}_3)\equiv\varepsilon^0(\hat{\mathbf{k}}_1)\cdot\varepsilon^0(\hat{\mathbf{k}}_2)\cdot\varepsilon^0(\hat{\mathbf{k}}_3)$ is a symmetric contraction of the longitudinal polarization tensors $\varepsilon^0_{i_1\cdots i_s}$ (see Appendix B for the precise definition of the polarization tensor) that reduces to $P_s(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3)$ in the squeezed limit. The couplings $\hat{\alpha}_s$ are

$$\hat{\alpha}_s\equiv\xi_s\left(\frac{\rho_s}{H^{2-s}}\right)^2<1\,, \tag{5.80}$$

where ξ_s was introduced in (5.23). The function $\mathcal{K}^{(s)}$ can be found in Appendix C.

The size of the non-Gaussianity associated with this diagram can, again, be read off from (5.65), with α_s replaced by $\hat{\alpha}_s$, and taking into account an extra suppression of μ_s^{-4}. Although the qualitative features of the non-analytic signal will be similar to that of the other diagrams, there are some relevant differences. First, as shown in §5.2.3, naturalness does not constrain the size of the coupling ξ_s, so the triple-exchange diagram allows for a naturally large non-Gaussianity. This is to be contrasted especially with the single-exchange diagram, where the naturalness criterion imposed a strong constraint on the size of the corresponding non-Gaussianity. Second, when the mass of the particle becomes large, the bispectrum is well-captured by a local vertex, namely $(\hat{\partial}_{i_1\cdots i_s}\pi)^3$ with symmetric contraction of indices. Notice that, due to the number of spatial gradients, for $s>2$ the squeezed-limit bispectrum is suppressed by more than $(k_1/k_3)^2$ for small k_1. This makes the non-analytic part, scaling as $(k_1/k_3)^{3/2}$, a rather clean signal in the squeezed limit.

Summary. All diagrams in Fig. 5.6, except for the single-exchange diagram for spin one, can yield sizable non-Gaussianities within the perturbative regime. In order for this to be natural, the single-exchange diagram requires new physics or fine-tuning to stabilize the mass of the spinning particle, whereas both the double- and triple-exchange diagrams can naturally produce large non-Gaussianities. The non-analytic part of the bispectrum is suppressed by $e^{-\pi\mu_s}$ for $c_\pi = 1$, but

only by $e^{-\pi\mu_s/2}$ for small c_π. Typically, we find that $f_{\rm NL} \gtrsim \mathcal{O}(1)$ from the non-analytic part is possible if $\mu_s \lesssim 5$ for $c_\pi = 1$ and $\mu_s \lesssim 10$ for $c_\pi \ll 1$.

5.3.3 $\langle\gamma\zeta\zeta\rangle$

Lastly, we consider the tensor-scalar-scalar correlation function $\langle\gamma\zeta\zeta\rangle$. In single-field inflation, a long-wavelength tensor fluctuation is locally equivalent to a spatially anisotropic coordinate transformation. Again, we can Taylor expand the expectation value around the squeezed limit, thus obtaining

$$\lim_{k_1\ll k_3} \langle\gamma^\lambda_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle' = P_\gamma(k_1)P_\zeta(k_3)\sum_{n=0}^{\infty} d_n \left(\frac{k_1}{k_3}\right)^n , \tag{5.81}$$

where γ^λ, with $\lambda = \pm 2$, denotes the positive or negative helicity components of the graviton. As in the case of the scalar bispectrum, the leading coefficients are determined by the single-field consistency relation [25] (see also [26, 60]). In particular, d_0 in (5.81) is given by

$$d_0 = \frac{1}{16}\mathcal{E}^\lambda_2(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3)\big[3-(n_s-1)\big] , \tag{5.82}$$

where $\mathcal{E}^\lambda_2(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3) \equiv \hat{k}^i_3\hat{k}^j_3\varepsilon^\lambda_{ij}(\hat{\mathbf{k}}_1)$, with $\varepsilon^\lambda_{ij}\varepsilon^{\lambda *}_{ij} = 4$. When the consistency relation holds, it also completely fixes the linear term d_1 in (5.81), and physical effects appear at order $(k_1/k_3)^2$. The presence of new particles during inflation invalidates the Taylor expansion and leads to non-analytic scalings in (5.81). Our goal in this section is to study these characteristic signatures of massive spinning particles.

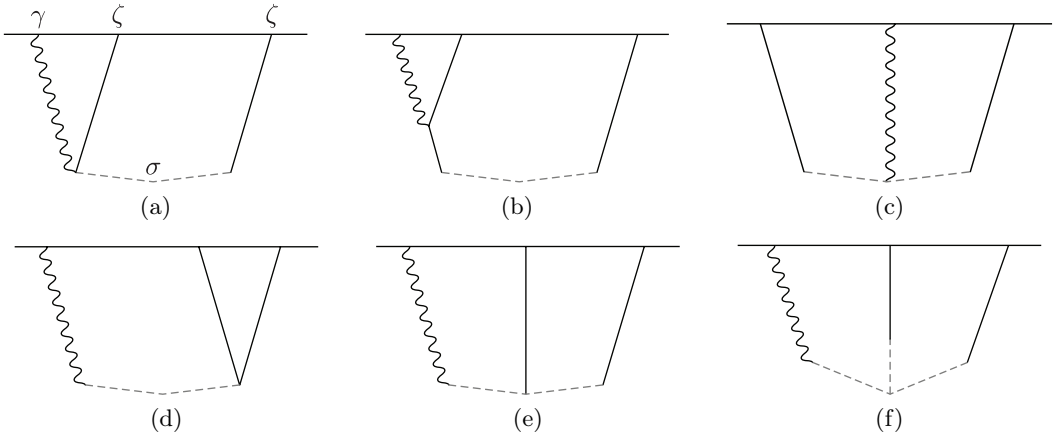


Figure 5.9: Tree-level diagrams contributing to $\langle\gamma\zeta\zeta\rangle$. The solid, dashed, and wavy lines represent the curvature perturbation ζ, a spinning field $\sigma_{i_1\cdots i_s}$, and the graviton γ_{ij}, respectively.

All tree-level diagrams contributing to $\langle\gamma\zeta\zeta\rangle$ are shown in Fig. 5.9. Not all of these diagrams can lead to a nontrivial deviation from the consistency relation. For the diagrams [(a-c)] the same symmetry that generates the tensor consistency relation enforces corrections to the power

spectrum, so that the relation in (5.81) and (5.82) is preserved [135]. Only the diagrams [(d-f)], which involve a quadratic mixing between the graviton and the intermediate particle, can lead to such a deviation. These diagrams have the same structure as those in Fig. 5.6, except that one of the legs in the quadratic mixing is replaced by an external graviton, so that the exchanging particle must carry the same helicity as the graviton. In the following, we will present results for the diagrams [(d-f)], mostly focusing on the single-exchange diagram [(d)] to avoid repetition. The quadratic γ-σ mixing vanishes for spins 0 and 1, so only particles with $s \geq 2$ will contribute.

Single-exchange diagram. We first consider the single-exchange diagram (Fig. 5.9d). The relevant interaction Lagrangian is (cf. (5.23) and (5.29))

$$\mathcal{L}_I = \frac{1}{a^{2s}}\left(-\frac{f_\pi^2}{M_{\rm pl}}\rho_s a^2\,\partial_{i_3\cdots i_s}\dot{\gamma}^c_{i_1 i_2}\hat{\sigma}_{i_1\cdots i_s} + \frac{1}{\Lambda_s^s}\,\dot{\pi}_c\partial_{i_1\cdots i_s}\pi_c\hat{\sigma}_{i_1\cdots i_s}\right). \tag{5.83}$$

Using (3.40) and (3.45), we can write the coefficient of the quadratic mixing term as $-\rho_s\sqrt{r/8}\,H$. The perturbativity condition on the π-σ mixing, $\rho_s < 1$, implies that the γ-σ mixing carries an extra suppression factor of $\sqrt{r/8}$. The bispectrum corresponding to the single-exchange diagram is

$$\frac{\langle\gamma^\lambda_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle'}{\Delta_\gamma\Delta_\zeta^3} = \alpha_2\sqrt{r}\,\Delta_\zeta^{-1}\times\mathcal{E}_2^\lambda(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3)\,\hat{P}_s^\lambda(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3)\times\mathcal{B}^{(s)}(\mu_s,c_\pi,k_1,k_2,k_3) + (\mathbf{k}_2\leftrightarrow\mathbf{k}_3)\,, \tag{5.84}$$

where $\hat{P}_s^\lambda \equiv (1-x^2)^{-\lambda/2}P_s^\lambda$, with P_s^λ being the associated Legendre polynomial. The function $\mathcal{B}^{(s)}$ is given explicitly in Appendix C.

Size of NG. We quantify the size of the tensor-scalar-scalar bispectrum by

$$f_{\rm NL}^{\gamma\zeta\zeta} \equiv \frac{6}{17}\sum_{\lambda=\pm2}\frac{\langle\gamma^\lambda_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle'}{P_\gamma^{1/2}(k)P_\zeta^{3/2}(k)}\,, \tag{5.85}$$

where the bispectrum is evaluated in the equilateral configuration, $k_1 = k_2 = k_3 \equiv k$, with vectors maximally aligned with the polarization tensor. This choice of normalization agrees with that adopted in [136] and implies $f_{\rm NL}^{\gamma\zeta\zeta} = \sqrt{r}/16$ for single-field slow-roll inflation [25]. An estimate of the size of the non-Gaussianity from the single-exchange diagram is

$$f_{\rm NL}^{\gamma\zeta\zeta} \;\sim\; g(\mu_s)\,\alpha_s\sqrt{r}\,\Delta_\zeta^{-1}\,, \tag{5.86}$$

where $g(\mu_s)$ denotes the appropriate mass suppressions for the analytic and non-analytic parts, which in the large μ_s limit scale as[17]

$$g(\mu_s) \;\equiv\; \begin{cases} \mu_s^{-2} & \text{analytic,} \\ e^{-\pi\mu_s} & \text{non-analytic.} \end{cases} \tag{5.87}$$

[17]The exponential suppression of the non-analytic part of the signal applies to particles in the principal series. For particles belonging to the complementary series, the non-analytic part of the signal would not be exponentially suppressed.

Unlike the scalar case, this exponential suppression cannot be reduced to $e^{-\pi\mu_s/2}$, since the graviton propagates with $c_t = 1$. The enhancement of $f_{\rm NL}^{\gamma\zeta\zeta}$ by the large factor Δ_ζ^{-1} means that, in principle, the signal could be significantly larger than the one predicted from single-field slow-roll inflation, $f_{\rm NL}^{\gamma\zeta\zeta} \gg \sqrt{r}/16$, even in the perturbative regime. As in the scalar case, the condition for radiative stability gives a rather strong constraint on the naturally allowed size of the bispectrum associated with the single-exchange diagram—cf. (5.67). Future constraints on $f_{\rm NL}^{\gamma\zeta\zeta}$ from observations of the $\langle BTT\rangle$ correlator of CMB anisotropies were discussed in [136]. The proposed CMB Stage IV experiments [137] will have the sensitivity to reach $\sigma(\sqrt{r} f_{\rm NL}^{\gamma\zeta\zeta}) \sim 0.1$, which suggests that the tensor non-Gaussianity due to massive spinning particles might be detectable for $r \gtrsim 10^{-6}\,[g(\mu_s)\alpha_s]^{-1}$.[18]

Shape of NG. In the squeezed limit, $\langle\gamma\zeta\zeta\rangle$ behaves in the following ways:

- The analytic part scales as

$$\lim_{k_1\ll k_3} \langle \gamma^\lambda_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3}\rangle' \propto \frac{1}{k_1^3 k_3^3} \left(\frac{k_1}{k_3}\right)^s \mathcal{E}_2^\lambda(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3)\, \hat{P}_s^\lambda(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3)\,. \tag{5.88}$$

  Notice that the suppression of the analytic part in the squeezed limit increases with spin. This can be understood by looking at the form of the local vertex after integrating out the massive particle, which becomes $\dot\pi\partial_{i_1\cdots i_s}\pi\partial_{i_3\cdots i_s}\dot\gamma_{i_1 i_2}$. As we will see below, this means that the analytic part of the signal will be subdominant compared to its non-analytic counterpart in the soft graviton limit.

- For $\mu_s \geq 0$, the squeezed limit of the non-analytic part of the bispectrum scales as

$$\lim_{k_1\ll k_3} \langle \gamma^\lambda_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3}\rangle' \propto \frac{1}{k_1^3 k_3^3} \left(\frac{k_1}{k_3}\right)^{3/2} \mathcal{E}_2^\lambda(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3)\, \hat{P}_s^\lambda(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3) \cos\left[\mu_s \ln\left(\frac{k_1}{k_3}\right) + \tilde\phi_s\right], \tag{5.89}$$

  where the phase $\tilde\phi_s$ is a function of μ_s and c_π (see Appendix C). Coupling to a particle with spin greater than two induces an extra angular structure. For imaginary μ_s, we instead have

$$\lim_{k_1\ll k_3} \langle \gamma^\lambda_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3}\rangle' \propto \frac{1}{k_1^3 k_3^3} \left(\frac{k_1}{k_3}\right)^{3/2-\nu_s} \mathcal{E}_2^\lambda(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3)\, \hat{P}_s^\lambda(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3)\,, \tag{5.90}$$

  with $\nu_s \equiv -i\mu_s \in [0, 1/2)$. This gives a non-analytic $(k_1/k_3)^{3/2-\nu_s}$ correction to the leading term of the consistency relation (5.81). Since unitarity implies $\nu_s < 1/2$, the squeezed-limit bispectrum due to massive spinning particles will be suppressed by at least k_1/k_3 compared to the leading term in the tensor consistency relation.[19]

[18] Producing a large tensor contribution while keeping the scalar contribution small may require some fine-tuned cancellation between interactions in the scalar sector. This is because the interaction vertices in (5.83) and (5.60) arise from the same operators in unitary gauge. Suppressing the effects of the interactions in (5.60) would require balancing them against additional interactions such as $\dot\pi\sigma_{0\cdots 0}$.

[19] A deviation from the leading term of the consistency relation due to spinning particles can arise in a number of ways: First, the unitarity bound can be evaded if the de Sitter isometries are not fully respected in the quadratic

Other diagrams. The extensions to the diagrams [(e,f)] are completely analogous to the scalar case. These diagrams have the advantage that they are less constrained by naturalness considerations.

5.4 Conclusions

In this chapter, we have studied the imprints of massive particles with spin on cosmological correlators using the framework of the effective field theory of inflation [36]. This generalizes the work of Arkani-Hamed and Maldacena (AHM) [102] to cases where conformal symmetry is strongly broken. Let us summarize our results and contrast them with the conclusions of AHM:

- In AHM's more conservative analysis, the overall size of non-Gaussianity was too small to be observable even in the most optimistic experimental scenarios. Our results are cautiously more optimistic. Within the regime of validity of the effective field theory, we can accommodate observable non-Gaussianity as long as the masses of the new particles aren't too far above the Hubble scale during inflation.

- The key spectroscopic features of massive particles with spin do not rely on conformal invariance and therefore continue to hold in our analysis. As explained in [102], the masses and spins of extra particles during inflation can be extracted by measuring the momentum dependence in the squeezed limit.

- Our systematic effective field theory treatment of massive spinning particles during inflation allows for a complete characterization of their effects on non-Gaussian cosmological correlators, including their imprints beyond the squeezed limit. We showed that the characteristic angular dependence resulting from the presence of particles with spin persists even for more general momentum configurations. Having access to the complete correlation functions will be valuable for future data analysis.

- We also studied the effects of an explicit breaking of special conformal symmetry by introducing a sound speed c_π for the Goldstone fluctuations. We found that, for $c_\pi < \mu_s^{-1}$, the exponential suppression in the production of the massive particles, $e^{-\pi\mu_s}$, is changed to $e^{-\pi\mu_s/2}$. For a given mass, the size of non-Gaussianity is therefore enhanced (or less suppressed) for small c_π.

- Finally, we showed that particles with spin greater than or equal to two lead to a signature in the squeezed limit of $\langle\gamma\zeta\zeta\rangle$. This signal may be observable in the $\langle BTT\rangle$ correlator of CMB anisotropies [136].

action of the spinning field [138, 139]. Another possibility involves partially massless fields with spin greater than two, since the late-time behavior of these fields does not obey the same restrictions as for the massive case. It would be interesting to explore these possibilities further [140].

6. Tensors beyond Einstein gravity

Our primary goal in Chapters 4 and 5 was to look for signatures of UV physics encoded in correlation functions of scalar perturbations. Another fundamental prediction of inflation, which is the subject of this chapter, is a stochastic background of primordial tensor modes. If the inflationary scale is sufficiently high, tensor modes may be sensitive to higher-curvature corrections from the UV completion of gravity such as string theory. In particular, in [141], it was shown that higher-curvature corrections give rise to a new structure in the tensor three-point function. While the analysis of the tensor three-point function is particularly clean and model-insensitive, it is also hard to verify in observations, since tensor non-Gaussianities are likely to be very small. In this chapter, we will discuss a potentially larger signature of higher-curvature corrections in the tensor two-point function.

The tensor power spectrum is characterized by an amplitude (r) and a tilt (n_t). In single-field slow-roll inflation, minimally coupled to Einstein gravity, these two parameters are related by a consistency condition [142], $r = -8n_t$. We will show that the leading higher-curvature corrections to the gravitational action lead to a violation of this consistency condition (see also [143]). We will arrive at this conclusion from two different perspectives: from the inflationary action and the wavefunction of the universe.

The outline of the chapter is as follows. In §6.1, we introduce a simple action in which higher-curvature corrections lead to a violation of the tensor consistency condition. We emphasize that the effect arises at leading order in the breaking of the de Sitter isometries. In §6.2, we confirm this conclusion with an analysis of the stress tensor two-point function in a perturbed conformal field theory. Our conclusions are summarized in §6.3, and some technical details are contained in Appendix D.

6.1 High-scale inflation

In this section, we motivate a scenario in which the inflaton breaks the isometries of de Sitter space by a minimal amount, and analyze the consequences for the scalar and tensor two-point functions. We show that a coupling to the Weyl tensor leads to a violation of the tensor consistency condition.

6.1.1 Weakly broken conformal symmetry

In a realistic inflationary model, the de Sitter symmetries need to be broken. For this purpose, we introduce the dynamical inflaton field ϕ. We give it a potential, $M_{\rm pl}^2 H^2 \to V(\phi)$, so that the field acquires a time-dependent expectation value $\phi = \bar{\phi}(t)$. This provides a natural clock measuring the time to the end of inflation. The expansion rate is now time dependent, $H \to H(t)$, and the inflationary slow-roll parameter can then be written as $\varepsilon = \dot{\bar{\phi}}^2/2M_{\rm pl}^2 H^2$. The size of ε controls the breaking of the conformal symmetries of de Sitter, with the symmetries being restored in the limit $\varepsilon \to 0$.

We will assume that inflaton self-interactions are suppressed by a relatively large mass scale, $\Lambda^2 \gg \dot{\bar{\phi}}$, while gravitational interactions are controlled by a lower scale, $\dot{\bar{\phi}} > M^2 \gtrsim H^2$. The hierarchy $\Lambda \gg M$ is protected because the gravitational coupling to the scalar sector is small.[1] In that case, the leading breaking of the conformal symmetry comes from the inflaton potential, while higher-derivative interactions, like $(\partial\phi)^4/\Lambda^4$, will be suppressed[2] by powers of $\dot{\bar{\phi}}/\Lambda^2 \ll H^2/M^2$. In addition, we may have functions of ϕ coupled to curvature tensors. These couplings were discussed systematically by Weinberg in [46] and was reviewed in §3.2. Like Weinberg, we consider these terms to be perturbative corrections to the Einstein-Hilbert action. This ensures that any ghost instabilities are moved outside the regime of validity of the effective theory. Using the field equations of the leading terms in the action, all inflaton-curvature couplings can be written in terms of couplings to the Weyl tensor. We will therefore study the following action

$$S = \int \mathrm{d}^4x \left(\mathcal{L}_\phi + \mathcal{L}_g\right), \qquad \text{with} \qquad \begin{aligned} \mathcal{L}_\phi &= \sqrt{-g}\left[-\frac{1}{2}(\partial\phi)^2 - V(\phi)\right], \\ \mathcal{L}_g &= \sqrt{-g}\,\frac{M_{\rm pl}^2}{2}\left[R + f(\phi)\frac{W^2}{M^2} + h(\phi)\frac{W\widetilde{W}}{M^2}\right]. \end{aligned} \tag{6.1}$$

Higher powers of the Weyl tensor will contribute to higher-point correlation functions, and are not relevant for the considerations of this chapter. Notice that we have factored out the scale $M_{\rm pl}^2$ in $\mathcal{L}_g$. This is consistent with the structure expected in string effective actions [51], with M playing the role of the string scale or the Kaluza-Klein scale. Since the Weyl tensor vanishes for any homogeneous FLRW metric, the background slow-roll solution is still determined by the Einstein-Hilbert part of $\mathcal{L}_g$.

The effects of the parity-violating term $h(\phi)W\widetilde{W}$ have been studied in [144–147]. Since $W\widetilde{W}$ is a total derivative, this term vanishes if $h(\phi)$ is a constant. The correction to the tensor two-point

[1] To see this, consider, for instance, higher-curvature terms of the form $\Delta\mathcal{L} \equiv M_{\rm pl}^2 R^2/M^2$. Using the leading-order equation of motion [46], $M_{\rm pl}^2 R_{\mu\nu} = -\partial_\mu\phi\partial_\nu\phi - V(\phi)g_{\mu\nu}$, one finds $\Delta\mathcal{L} \subset (\partial\phi)^4/\Lambda^4$, with $\Lambda \equiv \sqrt{M_{\rm pl}M} \gg M$. We see that the effective cutoff of the scalar sector is enhanced by the large ratio $M_{\rm pl}/M$.

[2] If the scale controlling inflaton interactions is smaller than $\dot{\bar{\phi}}$, then the power counting of the EFT changes significantly [36]. In this limit, inflaton fluctuations can propagate with a non-trivial sound speed, $c_s \ll 1$. In Appendix D, we show that a small sound speed induces a much stronger breaking of conformal symmetry than we wish to consider in this chapter. Conversely, if we demand that the conformal symmetry is only broken by effects of order ε, then these symmetry-breaking operators have to be highly suppressed.

function therefore comes from the field-dependent variation of $h(\phi)$. This leads to a difference in the amplitudes of the two chiralities of the tensor modes of order $\sqrt{\varepsilon}H^2/M^2$. In this chapter, we will be interested in the effects of the coupling $f(\phi)W^2$. We will show that the field-dependent variation[3] of $f(\phi)$ induces a sound speed for tensor fluctuations and a contribution to the tensor tilt of order $\sqrt{\varepsilon}H^2/M^2$.

Interestingly, the $f(\phi)W^2$ term in (6.1) is similar to a term in the effective action of the original Starobinsky model [149]. In these models, inflation is driven by a large number of conformally coupled fields whose stress tensor is induced by the conformal anomaly [150], $\langle T^\mu_\mu \rangle \supset cW^2$. The effective action that reproduces the conformal anomaly [151] includes the Weyl-squared term. It is not hard to imagine that variations of the model could contain a term of the form $f(\phi)W^2$. For example, such a term arises if one introduces the dilaton. It would be interesting to make this connection more precise [152].[4]

6.1.2 Violation of the consistency condition

We will now compute the scalar and tensor power spectra resulting from the action (6.1). We will use the standard ADM decomposition of the metric (2.30) and work in comoving gauge, defined in (2.31). At leading order, the curvature perturbation, ζ, and the tensor mode, γ_{ij}, decouple and can be treated separately.

Scalars. We first consider the spectrum of the curvature perturbation ζ. Since $W\widetilde{W}$ vanishes for scalar fluctuations, only $f(\phi)W^2$ contributes. At linear order in ζ, the non-zero components of the Weyl tensor are[5]

$$\begin{aligned} W^{0i}{}_{0j} &= \frac{1}{2}\left(\partial_i\partial_j - \frac{\delta_{ij}}{3}\partial_k^2\right)\left(\frac{\varepsilon\zeta}{a^2} + \partial_l^{-2}\frac{\mathrm{d}}{\mathrm{d}t}(a\varepsilon\dot{\zeta})\right) \equiv F_{ij}\,, \\ W^{ij}{}_{kl} &= \delta_{ik}F_{jl} + \delta_{jl}F_{ik} - \delta_{il}F_{jk} - \delta_{jk}F_{il}\,, \end{aligned} \tag{6.2}$$

and hence $W^2 = 8F_{ij}^2$. To eliminate terms with second-order time derivatives in the Weyl tensor, we use the leading-order equation of motion (2.33). The quadratic action can then be written in the following form

$$\frac{\mathcal{L}_\zeta}{M_{\rm pl}^2} = \frac{a^3\varepsilon}{c_s^2}\left[\dot{\zeta}^2 - a^{-2}c_s^2(\partial_i\zeta)^2\right], \tag{6.3}$$

[3] For constant $f(\phi)$, the Weyl-squared term, W^2, can be put into the Gauss-Bonnet form, $R^2_{\mu\nu\rho\sigma} - 4R^2_{\mu\nu} + R^2$, via the field redefinition $g_{\mu\nu} \to g_{\mu\nu} + fM_{\rm pl}^{-2}(-2R_{\mu\nu} + \frac{5}{3}g_{\mu\nu}R)$ [148]. Since the Gauss-Bonnet term is a total derivative, it only contributes a boundary term. However, the field redefinition also changes the inflaton kinetic term and the normalization of the Einstein-Hilbert action, so the constant part of the function $f(\phi)$ is still physical.

[4] We thank Juan Maldacena for this suggestion.

[5] To arrive at (6.2), we have used the linearized solutions to the Einstein constraint equations [25]: $\delta N = \dot{\zeta}/H$ and $\partial_i N^i = \varepsilon\dot{\zeta} - a^{-2}\partial_i^2\zeta/H$.

where we have introduced the sound speed

$$\frac{1}{c_s^2} - 1 \equiv \frac{8}{3}\varepsilon f(\bar{\phi})\frac{H^2}{M^2}\,. \tag{6.4}$$

If we had kept the couplings of the inflaton to R^2 and $R_{\mu\nu}^2$, we would have found additional corrections to c_s of the same order. Since the deviation from $c_s = 1$ is suppressed by a factor of $\varepsilon \ll 1$, it will not play a significant role for the rest of this chapter. For simplicity, we will therefore take $c_s \approx 1$, and write the power spectrum of ζ in the standard slow-roll form

$$\Delta_\zeta^2 \approx \frac{1}{8\pi^2}\frac{1}{\varepsilon}\frac{H^2}{M_{\rm pl}^2}\,, \tag{6.5}$$

where the right-hand side is evaluated at horizon crossing, $k = aH$. We conclude that the coupling to the Weyl tensor has very little effect on the scalar power spectrum, and its main effect is a correction to the tensors.

Tensors. The linearized equation of motion for tensor fluctuations in Einstein gravity is

$$\ddot{\gamma}_{ij} + 3H\dot{\gamma}_{ij} - a^{-2}\partial_k^2\gamma_{ij} = 0\,. \tag{6.6}$$

We will use this to simplify some of the perturbative corrections to the quadratic action for γ. At linear order in γ, the components of the Weyl tensor are

$$\begin{aligned}
W^{0i}{}_{0j} &= \frac{1}{4}\left(\ddot{\gamma}_{ij} + H\dot{\gamma}_{ij} + a^{-2}\partial_k^2\gamma_{ij}\right),\\
W^{0i}{}_{jk} &= \frac{1}{2a}\left(\dot{\gamma}_{ik,j} - \dot{\gamma}_{ij,k}\right),\\
W^{jk}{}_{0i} &= \frac{1}{2a}\left(\dot{\gamma}_{ij,k} - \dot{\gamma}_{ik,j}\right),\\
W^{ij}{}_{kl} &= \frac{1}{2}\Big\{a^{-2}\left(\gamma_{il,jk} + \gamma_{jk,il} - \gamma_{ik,jl} - \gamma_{jl,ik}\right)\\
&\quad + \frac{1}{2}\Big[\delta_{il}\left(\ddot{\gamma}_{jk} + H\dot{\gamma}_{jk} - a^{-2}\partial_m^2\gamma_{jk}\right) + \delta_{jk}\left(\ddot{\gamma}_{il} + H\dot{\gamma}_{il} - a^{-2}\partial_m^2\gamma_{il}\right) - (i\leftrightarrow j)\Big]\Big\}\,.
\end{aligned} \tag{6.7}$$

Substituting this into W^2, we get

$$\begin{aligned}
W^{\mu\nu}{}_{\rho\sigma}W^{\rho\sigma}{}_{\mu\nu} &= 4W^{0i}{}_{0j}W^{0j}{}_{0i} + 4W^{0i}{}_{jk}W^{jk}{}_{0i} + W^{ij}{}_{kl}W^{kl}{}_{ij}\\
&= 2\left(\ddot{\gamma}_{ij} + H\dot{\gamma}_{ij} + a^{-2}\partial_k^2\gamma_{ij}\right)^2 + 2a^{-2}\partial_k^2(\dot{\gamma}_{ij}^2)\,.
\end{aligned} \tag{6.8}$$

Using (6.6), this can be brought into the form of eq. (21) in [46]. After a few integrations by parts, we obtain

$$\frac{\mathcal{L}_\gamma}{M_{\rm pl}^2} = \frac{a^3}{8c_t^2}\Big[\dot{\gamma}_{ij}^2 - c_t^2 a^{-2}(\partial_k\gamma_{ij})^2\Big]$$

$$- 2\frac{a^3}{M^2} f(\bar{\phi})\Big[\gamma_{ij} a^{-2}\partial_k^2\big(\ddot{\gamma}_{ij} + 3H\dot{\gamma}_{ij} - a^{-2}\partial_l^2\gamma_{ij}\big)\Big]$$
$$- 2\frac{a}{M^2}\frac{\mathrm{d}f(\bar{\phi})}{\mathrm{d}t}\gamma_{ij}\partial_k^2\dot{\gamma}_{ij} - \frac{4}{M^2}\frac{\mathrm{d}h(\bar{\phi})}{\mathrm{d}t}\,\epsilon^{ijk0}\,\gamma_{il}\partial_j\partial_m^2\gamma_{kl}\,. \tag{6.9}$$

The second line in (6.9) vanishes after using the equation of motion (6.6) once more. The last line is proportional to $\dot{\bar{\phi}}$ and hence is suppressed in the slow-roll limit.[6] This leaves the first line, which is the quadratic action for tensors with a non-trivial sound speed

$$\frac{1}{c_t^2} - 1 \equiv 8f(\bar{\phi})\frac{H^2}{M^2}\,. \tag{6.10}$$

In [153], it was shown that a tensor sound speed can always be set to unity by a disformal transformation [154], followed by a Weyl rescaling to take the action to Einstein frame. These two metric transformations trade the non-trivial tensor sound speed for a scalar sound speed, $\tilde{c}_s$, and a modified Hubble rate, $\tilde{H}(\tilde{t})$. In the new frame, the tensor spectrum takes the standard form, $\Delta_\gamma^2 \propto \tilde{H}^2/M_{\rm pl}^2$, but the scalar spectrum is modified. Of course, predictions for observables are frame-independent [153, 155], so the choice of frame is simply a matter of convenience. In particular, the violation of the consistency condition that we will find is a frame-independent conclusion. We leave the details to Appendix D, but one point is worth emphasizing here. The violation of the consistency condition in the new frame is not the same as that found in $P(X)$-theories [156]. Rather, it is the time derivative of $\tilde{c}_s$ that modifies the tensor-to-scalar ratio.

The tensor sound speed leads to a simple rescaling of the standard tensor power spectrum. Summing over the two graviton polarizations, we obtain (cf. (3.45))

$$\Delta_\gamma^2 = \frac{2}{\pi^2}\frac{H^2}{M_{\rm pl}^2}\frac{1}{c_t}\,, \tag{6.11}$$

where the right-hand side is evaluated at $c_t k = aH$. If the inflaton-Weyl coupling is a small correction to the leading gravitational action—as we are assuming in order to avoid propagating ghost degrees of freedom—then c_t can't deviate much from unity. The main effect is not the size of c_t, but its time dependence.[7] In particular, the tensor-to-scalar ratio approximately still takes the form predicted by standard slow-roll inflation (cf. (2.54))

$$r \equiv \frac{\Delta_\gamma^2}{\Delta_\zeta^2} = \frac{16\varepsilon}{c_t} \approx 16\varepsilon\,. \tag{6.12}$$

However, the tensor tilt can still receive an important correction due to the time dependence of c_t. Crucially, the evolution of $c_t(t)$ is coupled to that of $\bar{\phi}(t)$. This will induce a tilt of the tensor

[6] The last term in (6.9), although slow-roll suppressed, is phenomenologically interesting because it leads to chiral gravitational waves [144–147]. Incidentally, the size of the chiral splitting, $\sqrt{\varepsilon}\,H^2/M^2$, is of the same order as the correction to the tensor tilt that we will get from the rest of the action.

[7] The fact that c_t is never allowed to deviate too far from unity puts a constraint on the time dependence of c_t, and hence on the function $f(\phi)$ in (6.10).

spectrum proportional to $\dot{\bar{\phi}} \propto \sqrt{\varepsilon}$. To see this, let us define a slow-variation parameter for the tensor sound speed, ε_t, and express it in terms of the slow-roll parameter ε:

$$\varepsilon_t \equiv \frac{\dot{c}_t}{H c_t} = \mp 4 c_t^2 b \sqrt{2\varepsilon} \frac{H^2}{M^2} + (1 - c_t^2)\varepsilon \,, \tag{6.13}$$

where $b \equiv M_{\rm pl} f'$ is a dimensionless constant. Taking the scale of variation of $f(\phi)$ to be of order Λ, we get $b \sim M_{\rm pl}/\Lambda$, which may be large if $\Lambda \ll M_{\rm pl}$. The fractional change of (6.11) per Hubble time then determines the tensor tilt

$$n_t \equiv \frac{\mathrm{d}\ln \Delta_\gamma^2}{\mathrm{d}\ln k} = -2\varepsilon - \varepsilon_t \approx -2\varepsilon \pm 4b\sqrt{2\varepsilon}\frac{H^2}{M^2} \,, \tag{6.14}$$

where we have ignored a small shift in the coefficient of the standard contribution, -2ε. Notice that the correction has undetermined sign, so it seems to allow a blue tilt for the tensor spectrum, even without a violation of the NEC.[8] The second term in (6.14) leads to a modification of the tensor consistency condition

$$\boxed{-\frac{8 n_t}{r} = 1 \mp \frac{4b}{\sqrt{2\varepsilon}}\frac{H^2}{M^2}} \,. \tag{6.15}$$

We see that the violation of the relation $n_t = -r/8$ is enhanced for small ε (and large b), but suppressed by H^2/M^2. In the stringy regime of inflation, H^2/M^2 can be of order one[9] and our proposed modification of the tensor spectrum could be a significant effect.

Testing the tensor consistency condition observationally is challenging (see [158, 159] for a recent discussion). Naturally, the observational prospects improve for large tensor-to-scalar ratio and if a large range of scales can be accessed (maybe with futuristic direct detection experiments [160–162]). A blue tensor spectrum would be easier to detect.

6.2 Dual interpretation

The freeze-out of quantum fluctuations during inflation allows us to recast cosmological expectation values in terms of the 'wavefunction of the universe', $\Psi[\zeta, \gamma]$. This wavefunction computes late-time expectation values of superhorizon fluctuations. The isometries of de Sitter space imply that the coefficients of the wavefunction can be interpreted as correlation functions of the stress tensor in a putative conformal field theory [25, 163]. The small breaking of conformal symmetry during inflation is modelled as a small deformation of the CFT, which can be treated perturbatively [164]. In this section, we will show that this alternative point of view reproduces the results of the previous section.

[8] In the Einstein frame, with $\tilde{c}_t = 1$, a blue tilt still corresponds to a violation of the NEC, $\dot{\tilde{H}} > 0$, but without inducing the gradient instability that this usually implies [53, 153].

[9] When M is the string scale, the ratio H/M is constrained by the fact that we require the Hagedorn temperature to remain above the de Sitter temperature in order to avoid a phase transition of the system [157].

6.2.1 Wavefunction of the universe

The wavefunction of the universe can be computed by a saddle-point approximation, $\Psi \approx e^{iS_{cl}}$, where the action S_{cl} is evaluated for a classical solution with certain Dirichlet boundary conditions [25]. The result takes the following form

$$\Psi = e^{iS_{div}} e^{W_0[\zeta,\gamma]}\,, \qquad \text{with} \qquad W_0 = \frac{1}{2}\int \mathrm{d}^3k \left(\zeta_{\mathbf{k}}\,\zeta_{-\mathbf{k}}\,\langle T_{\mathbf{k}}\,T_{-\mathbf{k}}\rangle' + \sum_s \gamma^s_{\mathbf{k}}\,\gamma^s_{-\mathbf{k}}\,\langle T^s_{\mathbf{k}}\,T^s_{-\mathbf{k}}\rangle' \right). \tag{6.16}$$

The local divergent piece, $e^{iS_{div}}$, is a pure phase factor, and thus drops out of expectation values. The coefficient functions $\langle T_{\mathbf{k}}T_{-\mathbf{k}}\rangle'$ and $\langle T^s_{\mathbf{k}}T^s_{-\mathbf{k}}\rangle'$ may be interpreted as the correlation functions of the trace and the trace-free part of the stress tensor T_{ij} of a dual field theory [25, 163, 165].[10] The power spectra of ζ and γ are then computed by a simple Gaussian integration

$$\langle \zeta_{\mathbf{k}}\,\zeta_{-\mathbf{k}}\rangle' = \int \mathcal{D}\zeta\;\zeta_{\mathbf{k}}\,\zeta_{-\mathbf{k}}\,|\Psi[\zeta]|^2 = -\frac{1}{2\,\mathrm{Re}\langle T_{\mathbf{k}}\,T_{-\mathbf{k}}\rangle'}\,, \tag{6.17}$$

$$\langle \gamma^s_{\mathbf{k}}\,\gamma^s_{-\mathbf{k}}\rangle' = \int \mathcal{D}\gamma^s\;\gamma^s_{\mathbf{k}}\,\gamma^s_{-\mathbf{k}}\,|\Psi[\gamma^s]|^2 = -\frac{1}{2\,\mathrm{Re}\langle T^s_{\mathbf{k}}\,T^s_{-\mathbf{k}}\rangle'}\,. \tag{6.18}$$

The diffeomorphism invariance of gravitational theories implies that the generators of coordinate transformations act as constraints on the wavefunction [166]. These constraint equations are the conformal Ward identities of the coefficient functions $\langle T_{\mathbf{k}}\,T_{-\mathbf{k}}\rangle'$ and $\langle T^s_{\mathbf{k}}\,T^s_{-\mathbf{k}}\rangle'$. In a CFT, these constraints imply

$$\langle T_{\mathbf{k}}\,T_{-\mathbf{k}}\rangle' = 0\,, \qquad \langle T^s_{\mathbf{k}}\,T^s_{-\mathbf{k}}\rangle' = c_T k^3\,, \tag{6.19}$$

where c_T is the central charge. We see that there are no ζ-fluctuations and the gravitational sector consists only of gravitons. In terms of bulk quantities, the central charge is

$$c_T = -\frac{1}{4}\frac{M_{\rm pl}^2}{H^2}\,. \tag{6.20}$$

In a quasi-de Sitter background, with finite slow-roll parameter ε, some of the conformal symmetries are softly broken (see Appendix D). The effects of this weak symmetry breaking can be treated perturbatively.

6.2.2 Conformal perturbation theory

A suitable framework for analyzing field theories that are almost conformal is conformal perturbation theory [164]. We now wish to show that such an analysis reproduces the results of §6.1.

[10] The CFT that describes the de Sitter cosmology is not unitary and has some unusual features, mostly related to the spectrum of the dimensions of operators. Our analysis will only use the mapping of the symmetries between the bulk and the boundary, and does not rely on a deeper meaning of dS/CFT.

We deform the CFT by a local primary operator [167][11]

$$S = S_{\rm CFT} + \beta \int \mathrm{d}^3 z\, \mathcal{O}(\mathbf{z})\,, \tag{6.21}$$

where β is a small coupling.[12] We take the perturbing operator to be marginally relevant, so its dimension is $\Delta \equiv 3 - \lambda$, with $0 < \lambda \ll 1$. The small expansion parameter dual to $\sqrt{\varepsilon}$ will be a combination of β and λ. For convenience, we normalize the two-point function of $\mathcal{O}$ by the central charge

$$\langle \mathcal{O}_{\mathbf{k}}\, \mathcal{O}_{-\mathbf{k}} \rangle' = c_T\, k^{3-2\lambda}\,. \tag{6.22}$$

For small β, the two-point function of the stress tensor can be computed perturbatively as

$$\begin{aligned} \langle T_{ij} T_{kl} \rangle &= \big\langle T_{ij} T_{kl}\, e^{-\beta \int \mathrm{d}^3 z\, \mathcal{O}} \big\rangle_0 \\ &= \big\langle T_{ij} T_{kl} \big\rangle_0 - \beta \int \mathrm{d}^3 z\, \big\langle T_{ij} T_{kl}\, \mathcal{O}(\mathbf{z}) \big\rangle_0 \\ &\quad + \frac{\beta^2}{2} \int \mathrm{d}^3 z\, \mathrm{d}^3 w\, \big\langle T_{ij} T_{kl}\, \mathcal{O}(\mathbf{z}) \mathcal{O}(\mathbf{w}) \big\rangle_0 + \cdots\,, \end{aligned} \tag{6.23}$$

where the expectation values $\langle T_{ij} T_{kl} \ldots \mathcal{O} \rangle_0$ are computed using the CFT operator algebra, and in general are constrained by Ward identities. We will use the following trace Ward identities obeyed by the stress tensor

$$\big\langle T^i{}_i(\mathbf{x}) T_{kl}(\mathbf{y}) \mathcal{O}(\mathbf{z}) \big\rangle_0 = \lambda\, \delta(\mathbf{x} - \mathbf{z}) \big\langle T_{kl}(\mathbf{y}) \mathcal{O}(\mathbf{z}) \big\rangle_0 = 0\,, \tag{6.24}$$

$$\big\langle T^i{}_i(\mathbf{x}) T_{kl}(\mathbf{y}) \mathcal{O}(\mathbf{z}) \mathcal{O}(\mathbf{w}) \big\rangle_0 = \lambda \left[\delta(\mathbf{x} - \mathbf{z}) \big\langle T_{kl}(\mathbf{y}) \mathcal{O}(\mathbf{z}) \mathcal{O}(\mathbf{w}) \big\rangle_0 + (\mathbf{z} \leftrightarrow \mathbf{w}) \right], \tag{6.25}$$

$$\big\langle T^i{}_i(\mathbf{x}) T^k{}_k(\mathbf{y}) \mathcal{O}(\mathbf{z}) \mathcal{O}(\mathbf{w}) \big\rangle_0 = \lambda^2 \left[\delta(\mathbf{x} - \mathbf{z})\, \delta(\mathbf{y} - \mathbf{w}) \big\langle \mathcal{O}(\mathbf{z}) \mathcal{O}(\mathbf{w}) \big\rangle_0 + (\mathbf{z} \leftrightarrow \mathbf{w}) \right], \tag{6.26}$$

where, in the last identity, we have dropped an irrelevant contact term, with support when $\mathbf{x} = \mathbf{y}$. In a reparametrization invariant theory, $\nabla^i \langle T_{ij} \rangle = 0$, we furthermore have $k^i \langle T_{ij}(\mathbf{k}) T_{kl}(-\mathbf{k}) \rangle' = 0$. Imposing this constraint implies that $\langle T_{ij} T_{kl} \rangle$ has the following form

$$\langle T_{ij,\mathbf{k}} T_{kl,-\mathbf{k}} \rangle' = \frac{1}{4} \left[\delta^\perp_{ij} \delta^\perp_{kl} \langle T_{\mathbf{k}} T_{-\mathbf{k}} \rangle' + \left(\delta^\perp_{ik} \delta^\perp_{jl} + \delta^\perp_{il} \delta^\perp_{jk} - \delta^\perp_{ij} \delta^\perp_{kl} \right) \langle T^s_{\mathbf{k}} T^s_{-\mathbf{k}} \rangle' \right], \tag{6.27}$$

where $\delta^\perp_{ij} \equiv \delta_{ij} - k_i k_j / k^2$.

In general, $\langle T^s_{\mathbf{k}} T^s_{-\mathbf{k}} \rangle'$ and $\langle T_{\mathbf{k}} T_{-\mathbf{k}} \rangle'$ are arbitrary functions of k, but in a theory with approxi-

[11] Of course, CFTs are characterized by a set of correlation functions rather than by an action. Here, $S_{\rm CFT}$ is simply a metaphoric way of characterizing the content of the original CFT. In practice, calculations in conformal perturbation theory are always performed at the level of correlation functions.

[12] In conformal perturbation theory, one usually tunes the perturbation so that the beta function vanishes and the theory flows to a new conformal fixed point [164]. Since we are mainly interested in the parametric scaling of the corrections, we will not perform this additional step and thus we do not worry about the particular renormalization group flow.

mate conformal symmetry, we expect them to be approximately scale invariant. The breaking of scale invariance can be studied in powers of β:

- First, let us look at the two-point function of the trace, $\langle T_{\mathbf{k}} T_{-\mathbf{k}}\rangle$. It follows from the trace Ward identity (6.24) that this vanishes at order β. The scalar two-point function is therefore only generated at order β^2. Using (6.25) and (6.26), we find

$$\langle T_{\mathbf{k}} T_{-\mathbf{k}}\rangle' = \beta^2\lambda^2\langle \mathcal{O}_{\mathbf{k}}\mathcal{O}_{-\mathbf{k}}\rangle' \,=\, c_T\,\beta^2\lambda^2\,k^{3-2\lambda}\,. \tag{6.28}$$

  The tensor-to-scalar ratio therefore is

$$r \equiv \frac{\langle T_{\mathbf{k}} T_{-\mathbf{k}}\rangle'}{\langle T^s_{\mathbf{k}} T^s_{-\mathbf{k}}\rangle'} = \beta^2\lambda^2\,. \tag{6.29}$$

  Comparing this to the bulk result, $r = 16\varepsilon$, we identify the duality map $\beta\lambda \leftrightarrow \pm 4\sqrt{\varepsilon}$.

- Next, we consider the correction to $\langle T^s_{\mathbf{k}} T^s_{-\mathbf{k}}\rangle'$. At $O(\beta)$, we require the integral of the three-point function $\langle T^s T^s \mathcal{O}\rangle_0$. In position space, we have

$$\big\langle T_{ij}(\mathbf{x}) T_{kl}(\mathbf{y})\mathcal{O}(\mathbf{z})\big\rangle_0' = c_T\, f_{TT\mathcal{O}}\,\mathcal{T}_{ijkl}(\mathbf{x}-\mathbf{y},\mathbf{y}-\mathbf{z},\mathbf{z}-\mathbf{x})\,, \tag{6.30}$$

  where an explicit expression for the tensor structure $\mathcal{T}_{ij\,kl}$ can be found in the classic work of Osborn and Petkou [168]—cf. eqs. (3.3)–(3.6). We have identified $c_T f_{TT\mathcal{O}}$ with the coefficient a of eq. (3.4) in [168]. Integrating (6.30) over $\mathbf{z}$, and transforming to momentum space, we get

$$\big\langle T^s_{\mathbf{k}} T^s_{-\mathbf{k}}\,\mathcal{O}(\mathbf{0})\big\rangle_0' = c_T\,\alpha(\lambda) f_{TT\mathcal{O}}\,k^{3-\lambda}\,, \tag{6.31}$$

  where $\alpha(\lambda)$ is a numerical coefficient. We have confirmed by explicit integration that $\alpha(\lambda)$ is finite, even in the limit $\lambda \to 0$.[13] This implies that T_{ij} doesn't need to be renormalized, and also means that perturbing the CFT by an exactly marginal operator simply shifts the coefficient of the stress tensor two-point function. Substituting (6.31) into (6.23), we get

$$\langle T^s_{\mathbf{k}} T^s_{-\mathbf{k}}\rangle' = c_T k^3\big(1 - n_t \ln k + \cdots\big)\,, \tag{6.32}$$

  where we have dropped a small $O(\beta)$ shift of the amplitude, and defined

$$n_t \equiv -\beta\lambda\,\alpha(\lambda)\,f_{TT\mathcal{O}}\,. \tag{6.33}$$

  This reproduces the $O(\sqrt{\varepsilon})$ contribution in the bulk result (6.14), if we make the following identification: $\alpha(\lambda) f_{TT\mathcal{O}} \leftrightarrow \sqrt{2}\,b H^2/M^2$. The $O(\beta^2)$ term in the $\langle T^s_{\mathbf{k}} T^s_{-\mathbf{k}}\rangle$ correlator contains the standard tensor tilt proportional to ε. This contribution depends on the details of the CFT and its various OPE coefficients.

[13] We also found that $\alpha(\lambda)$ vanishes for a two-dimensional CFT. This is to be expected from the c-theorem in two dimensions. It is also consistent with our bulk interpretation, since the Weyl tensor vanishes identically in three dimensions.

The main result of this section was the confirmation that a tensor tilt is generated at $O(\beta)$, while the tensor-to-scalar ratio is only non-zero at $O(\beta^2)$. The tensor tilt comes from a specific three-point function, whose size is set by $f_{TT\mathcal{O}}$ in the boundary CFT, and by H^2/M^2 in the bulk action. The standard result of Einstein gravity is recovered for $f_{TT\mathcal{O}} \to 0$, or $M \to \infty$.

6.3 Conclusions

The weak breaking of conformal symmetry during inflation can be used to constrain the predictions for cosmological correlators—e.g. [102, 141, 169]. This is especially relevant in inflationary models in which the scale suppressing higher-curvature corrections is close to the Hubble scale. In this chapter, we have studied the coupling of the inflaton field to higher-curvature tensors in models with a minimal breaking of conformal symmetry. We showed that the most general correction to the tensor two-point function is captured by a coupling to the square of the Weyl tensor. This interaction modifies the consistency condition of single-field slow-roll inflation

$$-\frac{8n_t}{r} = 1 \mp \frac{4b}{\sqrt{2\varepsilon}}\frac{H^2}{M^2}\,. \tag{6.34}$$

The correction can have either sign, and may dominate over the prediction from Einstein gravity if H/M is not too small. We consider this an interesting signature of higher-curvature corrections during inflation. We have left a few open questions for future work:

- What is the precise connection between our effective action (6.1) and the original Starobinsky model [149]? Both models rely on softly broken conformal invariance, and the effective actions even contain some terms of the same functional form. Making this relationship more precise would be very interesting [152].

- How naturally does our scenario arise in explicit string compactifications? Under which circumstances is a weakly broken conformal symmetry maintained in the four-dimensional effective theory? Is there a relation to conformal supergravity [170–172]?

- If a violation of the tensor consistency condition were to be observed, how would we convince ourselves that it comes from higher-curvature effects? In particular, it is well-known that a violation of $r = -8n_t$ can also arise from modifications of the scalar spectrum in models with a non-trivial scalar sound speed [156] and/or isocurvature fluctuations [173]. However, in that case we also expect strong interactions in the scalar sector. In contrast, in our proposal we do not predict a strong counterpart in scalar non-Gaussianity. A positive test of our scenario would be looking for correlated signatures of a low string scale, such as angular dependence in the scalar bispectrum [102] and specific forms of tensor non-Gaussianity [141].

- Can we get blue tensors? Our analysis determines neither the sign of the coupling to the Weyl tensor, nor its time dependence. While the sign may be constrained by requiring tensors to propagate subluminally, we see no a priori way to constrain the rate of change of the coupling. At the moment, blue tensors therefore seem to be a legitimate possibility.

7. Detecting primordial tensors

A distinguished feature of inflationary perturbations is the fact that they are correlated over apparently acausal scales. For scalar perturbations, this leads to a distinctive cross-correlation in the CMB between temperature perturbations and E-mode polarization [174]. The detection of superhorizon TE correlations by WMAP [175] is arguably the most convincing piece of evidence that the observed density perturbations were generated during inflation [34, 176]. In [177], it was pointed out that an analogous causality test can be performed for inflationary tensor modes. These leave a distinct swirl pattern (B-mode) in the polarization of the cosmic microwave background. In this chapter, we revisit and refine this proposal.

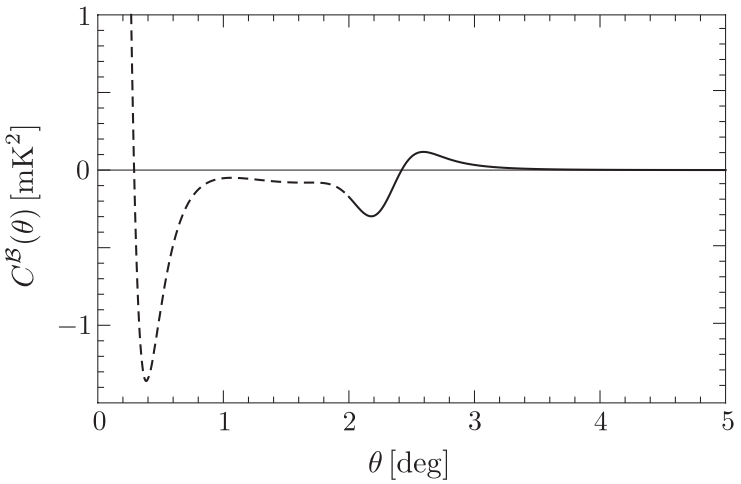


Figure 7.1: Local $\mathcal{B}$-mode correlation function predicted by inflation for $r = 0.13$. The dashed and solid parts of the curve represent subhorizon and superhorizon scales, respectively.

The conventional E- and B-modes [178, 179] are ill-suited to address questions of causality, since they are defined non-locally in terms of the Stokes parameters of the radiation field. We will therefore work with a local alternative [180] to the standard E- and B-modes which we will denote by $\mathcal{E}$ and $\mathcal{B}$. In the flat-sky limit, we have $\mathcal{E} = \nabla^2 E$ and $\mathcal{B} = \nabla^2 B$, where ∇^2 is the two-dimensional Laplacian in the plane orthogonal to the line-of-sight. Figure 7.1 shows the inflationary prediction for the $\mathcal{B}$-mode correlation function in real space. Unlike the correlation functions sourced by scalar fluctuations, the signal does not have a peak at the acoustic scale ($2\theta_a \sim 1.2^\circ$). Instead the tensor-induced signal peaks around the horizon scale ($\theta_c \equiv 2\theta_H \sim 2.3^\circ$) corresponding to the time

when the inflationary gravitational waves re-entered the horizon and started oscillating. Causality forbids superhorizon correlations above the horizon scale for any post-inflationary mechanism for gravitational wave production [176, 181], such as phase transitions [182] or defects [183, 184]. A measurement of $\mathcal{B}$-mode correlations above 2 degrees therefore constitutes an important test for the inflationary origin of the signal.

This chapter is organized as follows. In §7.1, we review the concept of the local $\mathcal{B}$-modes and present the superhorizon $\mathcal{B}$-mode signal predicted by inflation. In §7.2, we examine all potential sources of noise. We show that subhorizon modes can contaminate the superhorizon signal, especially if smoothing is applied to the data to suppress small-scale noise. In §7.4, we introduce an estimator of the superhorizon part of the signal and define the signal-to-noise ratio. We also present a measure for the amount of contamination from spurious subhorizon modes and describe ways to minimize their effects. In §7.5, we provide forecasts for the detectability of the superhorizon nature of inflationary B-modes for both current and future CMB polarization experiments. Our conclusions are stated in §7.6, and additional reference material is contained in Appendix E.

7.1 Superhorizon B-modes

We begin with a brief review of the superhorizon signature of inflationary B-modes [177].

Local B-modes. The polarization of the CMB is characterized by a symmetric, traceless rank-2 tensor defined in the plane perpendicular to the line-of-sight $\hat{\mathbf{n}}$:

$$P_{ij} = U\sigma_{ij}^{(1)} + Q\sigma_{ij}^{(3)}\,, \tag{7.1}$$

where $\sigma_{ij}^{(I)}$ denotes the Pauli matrices. Since the Stokes parameters Q and U transform non-trivially under rotations of the coordinates, it is more convenient to work with two invariants that can be constructed from the polarization tensor: a scalar $\mathcal{E} \equiv \partial_i\partial_j P_{ij}$ and a pseudo-scalar $\mathcal{B} \equiv \epsilon_{kj}\partial_k\partial_i P_{ij}$, corresponding to the gradient and curl parts of the polarization tensor, respectively. In the flat-sky limit, these $\mathcal{E}$- and $\mathcal{B}$-modes are related to the Stokes parameters and the ordinary E- and B-modes by [185]

$$\mathcal{E}(\mathbf{x}) = \nabla^2 E(\mathbf{x}) = (\partial_x^2 - \partial_y^2)Q(\mathbf{x}) + 2\partial_x\partial_y U(\mathbf{x})\,, \tag{7.2}$$

$$\mathcal{B}(\mathbf{x}) = \nabla^2 B(\mathbf{x}) = (\partial_x^2 - \partial_y^2)U(\mathbf{x}) - 2\partial_x\partial_y Q(\mathbf{x})\,. \tag{7.3}$$

By construction, $\mathcal{E}$ and $\mathcal{B}$ are local functions of the Stokes parameters, whereas E and B are defined non-locally in terms of Q and U. Being just a linear transformation of the conventional B-modes, the local $\mathcal{B}$-modes are also a signature of tensor (and vector) modes in the initial conditions.

Any scalar field on the celestial sphere can be expanded in terms of spherical harmonics, so

we write

$$X(\hat{\mathbf{n}}) \equiv \sum_{\ell m} a_{X,\ell m} Y_{\ell m}(\hat{\mathbf{n}}) \,, \tag{7.4}$$

where $X = \{T, \mathcal{E}, \mathcal{B}\}$. Assuming statistical isotropy, the two-point statistics of the multipole moments are described in terms of the angular power spectrum:

$$\langle a_{X,\ell m} a^*_{X,\ell' m'} \rangle = C^X_\ell \delta_{\ell\ell'} \delta_{mm'} \,, \tag{7.5}$$

where the angle brackets denote the ensemble average. The late-time power spectrum, C^X_ℓ, can be related to quantum zero-point fluctuations in both the spacetime metric and the matter fields during inflation [186]. Given a measurement of the harmonic coefficients $a_{X,\ell m}$, we define estimators of the angular power spectra as

$$\widehat{C}^X_\ell \equiv \frac{1}{2\ell+1} \sum_m a_{X,\ell m} a^*_{X,\ell m} \,. \tag{7.6}$$

The power spectrum of the local $\mathcal{B}$-modes is related to that of the conventional B-modes by [177, 187]

$$C^{\mathcal{B}}_\ell = n^2_\ell C^B_\ell \,, \tag{7.7}$$

where $n_\ell \equiv \sqrt{(\ell+2)!/(\ell-2)!}$. A harmonic transformation gives the corresponding correlation function in real space:

$$C^{\mathcal{B}}(\theta) = \sum_\ell \frac{2\ell+1}{4\pi} C^{\mathcal{B}}_\ell P_\ell(\cos\theta) \,, \tag{7.8}$$

where θ is the angle between pairs of line-of-sight directions $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$, i.e. $\cos\theta \equiv \hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2$. The relation between (7.8) and the correlation function of the conventional B-modes is

$$C^{\mathcal{B}}(\theta) = \nabla^2(\nabla^2+2) C^B(\theta) \,. \tag{7.9}$$

Again, the non-local nature of the ordinary B-modes is manifest: (7.9) implies that $C^{\mathcal{B}}$ vanishes for any C^B living in the kernel of $\nabla^2(\nabla^2+2)$, even if C^B is non-zero.

Superhorizon signal. Having defined the local $\mathcal{B}$-modes, we can analyze causality constraints on their correlation functions. The superhorizon part of the two-point correlation function is identified most directly in real space:[1]

$$S^{\mathcal{B}}(\theta) \equiv \Theta(\theta - \theta_c) C^{\mathcal{B}}(\theta) \,, \tag{7.10}$$

where Θ is the Heaviside step function and $\theta_c \simeq 2.3^\circ$ is (twice) the angle subtended by the particle

[1] This is the unique definition of the superhorizon signal, defined in reference to the horizon scale at recombination. In processing the data, it is possible to introduce a smoother filter to this definition, though not mandatory since there are effectively no ringing artifacts when the signal (7.10) is transformed to harmonic space.

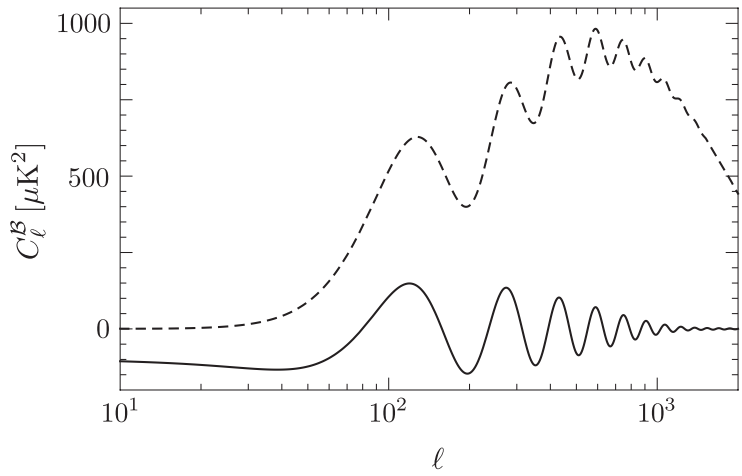


Figure 7.2: The superhorizon $\mathcal{B}$-mode power spectrum (solid) and the full $\mathcal{B}$-mode power spectrum (dashed) for $r = 0.13$ (without lensing).

horizon at recombination. The corresponding signal in harmonic space is

$$\begin{aligned} S_\ell^{\mathcal{B}} &= 2\pi \int_{-1}^{1} \mathrm{d}\cos\theta \; S^{\mathcal{B}}(\theta)\, P_\ell(\cos\theta) \\ &= \sum_{\ell'} M_{\ell\ell'} C_{\ell'}^{\mathcal{B}}\,, \end{aligned} \tag{7.11}$$

where the mode-coupling matrix $M_{\ell\ell'}$ is

$$M_{\ell\ell'} \equiv \frac{2\ell'+1}{2} \underbrace{\int_{-1}^{x_c} P_\ell(x) P_{\ell'}(x)\,\mathrm{d}x}_{\equiv I_{\ell\ell'}}\,, \tag{7.12}$$

with $x_c \equiv \cos\theta_c$. We label the complementary subhorizon signal as $(S_\ell^{\mathcal{B}})^\dagger \equiv C_\ell^{\mathcal{B}} - S_\ell^{\mathcal{B}}$. The mode-coupling integrals $I_{\ell\ell'}$ can be calculated analytically. The off-diagonal terms are given by

$$I_{\ell\ell'} = \frac{(\ell-\ell')x_c P_\ell P_{\ell'} + \ell' P_\ell P_{\ell'-1} - \ell P_{\ell-1} P_{\ell'}}{\ell(\ell+1) - \ell'(\ell'+1)}\,, \tag{7.13}$$

where the Legendre polynomials are evaluated at x_c, while the diagonal terms are determined by the recursion relation

$$I_{\ell\ell} = \frac{2\ell-1}{2\ell+1} I_{\ell-1,\ell-1} + \frac{2\ell-1}{2\ell+1}\frac{\ell+1}{\ell} I_{\ell+1,\ell-1} - \frac{\ell-1}{\ell} I_{\ell,\ell-2}\,. \tag{7.14}$$

We can think of the kernel (7.12) as an operator projecting the power spectrum onto its superhorizon subspace. In Fig. 7.2, we show the superhorizon part of the power spectrum predicted by inflation. We see that the features of the real space correlation function above θ_c are encoded in the oscillations of the power spectrum, where the frequency of the oscillations corresponds to the horizon size at recombination.

	θ_b [′]	$f_{\rm sky}$ [%]	$\Delta_{P,\rm eff}$ [μK′]	$s_{P,\rm eff}$ [μK′]
BICEP2	29	2.4	5.2	33.6
Keck Array	29	2.4	2.2	14.2
PolarBeaR-2	4	20	10.7	23.9
Simons Array	3	20	6.3	14.1
SPTPol	1	6	4.4	17.8
LiteBIRD	16	70	1.8	2.2
COrE	1	70	1.8	2.2

Table 7.1: Instrumental specifications for current and upcoming CMB polarization experiments [190–196].

7.2 Noise and leakage

Next, we describe the sources of noise and leakage that we will take into account in our analysis.

Instrumental noise. We represent instrumental noise by an uncorrelated Gaussian random field. Assuming white noise in the Stokes parameters, the noise power spectrum for B-modes can be expressed as [188]

$$N_\ell^B = \Delta_P^2 e^{\ell(\ell+1)/\ell_b^2}\,, \tag{7.15}$$

where Δ_P is the noise level of polarization sensitive detectors. The exponential factor in (7.15) represents the effect of deconvolving the Gaussian beam effect from the signal, with $\ell_b \equiv \sqrt{8\ln 2}/\theta_b$ and θ_b the full width at half maximum of the beam. The noise level is determined by

$$\Delta_P^2 = \frac{2{\rm NET}^2\Omega_{\rm sky}}{N_{\rm det}t_{\rm obs}Y} \equiv s_P^2 f_{\rm sky}\,, \tag{7.16}$$

where NET is the noise equivalent temperature of detectors, $N_{\rm det}$ denotes the number of detectors, $t_{\rm obs}$ is the time of observation, Y characterizes the detector yield, and $\Omega_{\rm sky} = 4\pi f_{\rm sky}$ is the observed sky area.[2] We will find it useful to consider the effective sensitivity of a full-sky experiment, s_P, and rescale it by the observed sky fraction, $f_{\rm sky}$. In an experiment with multiple frequency channels, a heuristic measure of the effective noise level is

$$\Delta_{P,\rm eff}^2 = \left[\sum_i \frac{1}{\Delta_{P,i}^2}\right]^{-1}\,, \tag{7.17}$$

where $\Delta_{P,i}$ denotes the noise level of channel i. The instrumental specifications and the effective noise levels for a selection of current and upcoming B-mode experiments are listed in Table 7.1 and in more detail in Appendix E.

[2]The current generation of experiments achieves NET $= 350\mu{\rm K}\sqrt{s}$ and $Y = 0.25$, with $N_{\rm det} \sim \mathcal{O}(10^3)$ [189]. In [137] ground-based experiments have been classified by the number of detectors as Stage-II, Stage-III and Stage-IV for $N_{\rm det} \sim 10^3$, 10^4 and 10^5, respectively.

As can be seen from (7.3), a measurement of $\mathcal{B}$-modes effectively involves taking derivatives of both the signal and the noise. This has the benefit that the observables become local quantities, but at the same time the noise spectrum for $\mathcal{B}$-modes acquires a factor of $n_\ell^2 \sim \ell^4$ relative to the noise for a B-mode measurement. White noise spectra for the Stokes parameters then translate into a blue spectrum for $\mathcal{B}$-modes, $N_\ell^{\mathcal{B}} \sim \ell^4$, implying a large contribution from small-scale noise. Because of the drastic difference in the properties of the noise, it is important to analyze the detectability of $\mathcal{B}$-modes separately, adopting a different strategy from measurements of B-modes if necessary.

In order to compensate for the blue noise spectrum, we will apply a low-pass filtering to both the signal and the noise:[3]

$$\mathcal{C}_\ell^{\mathcal{B}} \equiv f_\ell C_\ell^{\mathcal{B}}\,, \quad \mathcal{N}_\ell^{\mathcal{B}} \equiv f_\ell N_\ell^{\mathcal{B}}\,, \tag{7.18}$$

where f_ℓ denotes a filtering function. In real space, the procedure (7.18) corresponds to a convolution with a certain window function

$$f(\theta,\theta') = \sum_\ell \frac{2\ell+1}{2} f_\ell P_\ell(\cos\theta) P_\ell(\cos\theta')\,. \tag{7.19}$$

Depending on the experimental strategy, different window functions may be more suitable. For our purposes, there are several conditions that the filtering function f_ℓ needs to satisfy: (i) it needs to be sufficiently smooth to avoid the Gibbs phenomenon, (ii) it should decay early enough to suppress the small-scale noise efficiently, and (iii) it should retain the shape of the power spectrum up to $\ell \sim 100$ in order not to cause any distortion of the superhorizon features. A simple choice which satisfies the above requirements is a Gaussian filtering function:

$$f_\ell = e^{-\ell(\ell+1)/\ell_s^2}\,, \tag{7.20}$$

where ℓ_s defines the smoothing scale. To satisfy the second and third conditions, we choose $100 < \ell_s < \ell_b$, in which case the first condition is automatically satisfied.

Leakage. The filtering of the $\mathcal{B}$-mode spectrum is a necessary evil. An inevitable consequence of the filtering process is a transfer of part of the subhorizon signal to superhorizon scales (and vice versa). For lack of a better term, we will call this contamination *leakage*. Since the spurious modes due to leakage can confuse the detection of the true superhorizon signal, it will be important to treat them carefully in our analysis.

In Fig. 7.3, we show the filtered subhorizon and superhorizon $\mathcal{B}$-mode correlation functions. As we can see, there is a non-negligible amount of leakage around $\theta \sim 2^\circ$. On the other hand, the positive peak of the superhorizon signal at $\theta \sim 3^\circ$ is relatively clean and still serves as an unambiguous test of the inflationary superhorizon spectrum. When we want to make sure that we do not suffer from a large amount of leakage, we therefore focus on correlations with $\theta \gtrsim \theta_0 \equiv 2.6^\circ$.

[3] Calligraphic font will from now on denote filtered quantities.

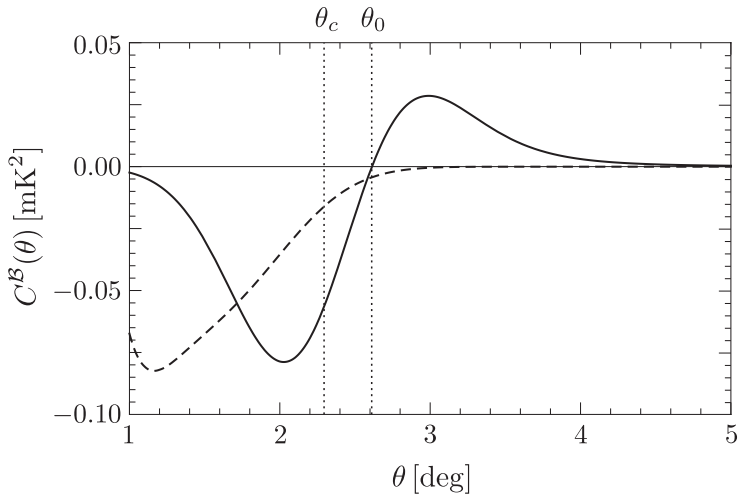


Figure 7.3: Local $\mathcal{B}$-mode correlation function for $r = 0.13$ using the Gaussian filter (7.20) with $\ell_s = 200$. The solid and dashed lines correspond to the superhorizon and subhorizon signals, respectively.

Moreover, at fixed θ, the leakage can be reduced by working with larger values of ℓ_s. However, making ℓ_s too large will reduce the signal-to-noise of the signal we wish to measure. In §7.5, we will discuss the optimal balance between minimal leakage and maximal signal-to-noise.

7.3 Foregrounds and lensing

Our ability to detect the primordial B-mode signal depends crucially on how well we can separate the signal from foreground contamination. The two major sources of foregrounds in the microwave range are polarized emissions from synchrotron and thermal dust. Their distinct frequency dependences, in principle, allow them to be distinguished from the primary CMB signal.

Synchrotron. Synchrotron radiation arises from the acceleration of relativistic cosmic-ray electrons in the magnetic field of the Galaxy. This is the dominant contribution to the polarized foreground emission below 70 GHz.

If the electrons have a power law distribution of energies, $N(E) \propto E^{-p}$, then the antenna temperature[4] of the signal is predicted to have a power-law dependence on frequency, $T(\nu) \propto \nu^{\beta_s}$, with $\beta_s = -\frac{1}{2}(p+3)$. This simple ansatz for the frequency spectrum fits observations rather well with $\beta_s \simeq -2.9$ [197]. The variation of the spectral index across the sky is of order 10%. The angular spectrum of the synchrotron emission is found to obey an approximate power law, $F_\ell^{s,B} \propto \ell^{\alpha_s}$, with $\alpha_s \simeq -2.6$ [198].

Combining the above facts, we are led to the following ansatz for the synchrotron B-mode

[4]Antenna temperature units are defined in reference to the Rayleigh-Jeans law, whereas thermodynamic temperature units are defined as the blackbody temperature obeying Planck's law. We calibrate quantities in thermodynamic temperature units, so that the primary CMB spectrum is frequency-independent.

power spectrum in thermodynamic temperature units [199]:

$$F_\ell^{s,B}(\nu) = \mathcal{A}_s \left(\frac{\ell}{\ell_0}\right)^{\alpha_s} h^s(\nu,\nu_0)\,, \tag{7.21}$$

where $\mathcal{A}_s$ is the amplitude of synchrotron emission defined at a reference frequency ν_0 and a reference scale ℓ_0. The function

$$h^s(\nu,\nu') \equiv \left(\frac{\nu}{\nu'}\right)^{2\beta_s} \left(\frac{f(\nu)}{f(\nu')}\right)^2 \tag{7.22}$$

encapsulates the spectral dependence, where the factor $f(\nu)$ accounts for the conversion from antenna temperature to thermodynamic temperature [200]:

$$f(\nu) \equiv \frac{(e^x-1)^2}{x^2 e^x}\,, \quad x \equiv \frac{h\nu}{kT_{\rm cmb}} \approx \frac{\nu}{56.8\,{\rm GHz}}\,, \tag{7.23}$$

where $T_{\rm cmb} = 2.725\,{\rm K}$ is the CMB blackbody temperature [33].

Thermal dust. Thermal emission from interstellar dust grains aligned with the Galactic magnetic field produces the dominant polarized foreground above 70 GHz.

The frequency dependence of the dust intensity takes the form of a modified blackbody, $I_\nu \propto \nu^{\beta_d} B_\nu(T_d)$, where the Planck spectrum $B_\nu(T_d)$ is determined by the observed dust temperature, $T_d \simeq 19.7\,{\rm K}$ [201]. The mean spectral index is found to be $\beta_d \simeq 1.5$ at microwave frequencies [197], with a variation of about 1% across the sky (much less than the variation of the synchrotron spectral index). The angular spectrum again satisfies a power law, $F_\ell^{d,B} \propto \ell^{\alpha_d}$, with $\alpha_d \simeq -2.3$ [202]. The dust B-mode power spectrum can therefore be modelled as [199]

$$F_\ell^{d,B}(\nu) = \mathcal{A}_d \left(\frac{\ell}{\ell_0}\right)^{\alpha_d} h^d(\nu,\nu_0)\,, \tag{7.24}$$

where $\mathcal{A}_d$ is the amplitude of the polarized dust emission defined at a reference frequency ν_0 and a reference scale ℓ_0. The spectral function for dust is[5]

$$h^d(\nu,\nu') \equiv \left(\frac{\nu}{\nu'}\right)^{2\beta_d} \left(\frac{B_\nu(T_d)}{B_{\nu'}(T_d)}\frac{g(\nu)}{g(\nu')}\right)^2\,, \tag{7.25}$$

where $g(\nu)$ is the conversion factor from intensity to thermodynamic temperature units [200],

$$g(\nu) \equiv \frac{f(\nu)}{\nu^2}\,. \tag{7.26}$$

The amplitude in (7.24) can be written as $\mathcal{A}_d = p^2 I_d$, where I_d is the unpolarized dust intensity and p is the polarization fraction. Both p and I_d can vary significantly across the sky.[6] The

[5] Eq. (7.25) corrects a typo in [203, 204].

[6] The latest Planck measurements suggest that the mean polarization fractions over most parts of the sky

precise amount of foreground contamination, therefore, depends on the region of the sky under consideration. In §7.5, we will consider a few different choices for the amplitude of the dust polarization, and allow for the relatively large uncertainties that still exist.

Foreground residuals. Multi-frequency observations allow some degree of foreground cleaning based on the distinct frequency dependence of the foregrounds. Detailed algorithms for foreground cleaning are discussed in [207]. Following [199, 204], we will assume that the foregrounds can be subtracted by the template cleaning method (e.g. [208, 209]), and simply parameterize the foreground residuals by rescaling the foreground amplitudes by two scale-independent factors, $\epsilon_x \in [0,1]$, with $x = \{s,d\}$ denoting synchrotron and dust, respectively. We propagate the noise of the template map into the foreground residuals. After cleaning, the residual foreground spectrum, then, is [203, 204]

$$R_\ell^B \equiv \sum_x \left[\epsilon_x F_\ell^{x,B} + \mathsf{N}_\ell^{x,B} h^x(\nu, \nu^x_{\rm ref})\right], \tag{7.27}$$

where $\nu^x_{\rm ref}$ is the reference frequency used as the template and $\mathsf{N}_\ell^{x,B}$ is the noise level of the template map for x.

We treat the foreground residuals as additional sources of uncorrelated noise (see Appendix E for a discussion). For an experiment with multiple frequency channels, we seek to find a linear combination of the maps with weightings chosen in such a way to minimize the variance of the power spectrum [200]. In Appendix E, we derive the optimal weighting scheme and show that the effective noise of the combined map is [199]

$$N^B_{{\rm eff},\ell} = \left[\sum_i \frac{1}{N^B_{i,\ell} + R^B_{i,\ell}}\right]^{-1}, \tag{7.28}$$

where the subscript i denotes the value at frequency ν_i. Appendix E also explains that any correlations between the foreground residuals at different frequencies tend to reduce the effective noise level, so working with (7.28) is a conservative choice.

Lensing. Even in the absence of primordial B-modes, a curl component of CMB polarization is generated by the lensing of primordial E-modes [210, 211]. This effect has to be considered an additional source of noise for the signal we are trying to measure.

On large angular scales, the lensing B-modes act like white noise with an effective amplitude of $4.4\,\mu$K′. In the low-noise regime ($\lesssim 5\,\mu$K′), the lensing effect provides a significant limitation to a measurement of the primordial signal, especially for low values of r. Since lensing does not induce any spectral distortions to the primary CMB, multi-frequency observations do not help to distinguish between these two signals. However, several methods have been proposed to reduce

(including highly polarized regions) fall in the range of 3 to 14% [205]. The dust intensity is constrained by the Finkbeiner-Davis-Schlegel (FDS) dust map [206].

the lensing noise statistically [212, 213] (see [214] for a comprehensive discussion). The most promising delensing procedure involves reconstructing the lensing potential from measurements of small-scale CMB polarization, which is subsequently used to remove the lensing contribution to the large-scale B-mode signal. This requires CMB experiments with high sensitivity and resolution (small beam size). Details of this approach to delensing can be found e.g. in [215].

In the absence of sky cuts, foregrounds, and instrumental systematics, a detection of the primordial tensor amplitude down to $r \sim 10^{-6}$, in principle, is achievable [216]. Nevertheless, experimental limitations and the presence of foregrounds practically limit an accurate quantification of the residual lensing, resulting in a possible bias in the estimator of the lensing potential. To avoid these practical uncertainties, we assume that the lensing estimator is unbiased, or that any significant biases are known and can be eliminated. Thus, the residual lensing contributes only to the variance, and does not bias the signal. The issue of potential lensing bias is the subject of many investigations in the literature, e.g. [217–223], but is beyond the scope of the present work.

We consider delensing in a heuristic way by multiplying the amplitude of the lensing B-modes L_ℓ^B by a scale-independent delensing fraction,

$$L_\ell^B \;\to\; \epsilon_L L_\ell^B \,, \tag{7.29}$$

where $\epsilon_L \in [0,1]$, and treat it as an additional noise. On large scales, both the residual spectrum and the original spectrum are approximately white noise (see e.g. Fig. 1 in [224]). Therefore, the ansatz (7.29) is a sufficiently good approximation to more sophisticated expressions for the lensing residuals found in Appendix A of [215]. The residual lensing power spectrum is then incorporated into the effective noise as

$$N^B_{\mathrm{eff},\ell} = \left[\sum_i \frac{1}{N^B_{i,\ell} + R^B_{i,\ell}}\right]^{-1} + \epsilon_L L_\ell^B \,. \tag{7.30}$$

Further justification for this formula is given in Appendix E.

7.4 Methodology

We now describe our method for quantifying the detectability of the superhorizon $\mathcal{B}$-mode signal. We first construct an estimator of the signal and then use it to define the signal-to-noise ratio of the measurement. We will explain that leakage introduces a bias in the estimator and describe a simple debiasing procedure. The methodology in this section and the next will be formulated mostly in real space, but see Appendix E for an equivalent treatment in harmonic space.

Superhorizon estimator. We would like to define an estimator of the superhorizon signal (7.10), given an estimator $\widehat{\mathcal{C}}_\ell$ for the total $\mathcal{B}$-mode power spectrum after filtering, which

satisfies $\langle\widehat{\mathcal{C}}_\ell\rangle = \mathcal{C}_\ell^{\mathcal{B}} \equiv f_\ell C_\ell^{\mathcal{B}}$. The associated covariance matrix is[7]

$$\mathscr{C}[\widehat{\mathcal{C}}_\ell, \widehat{\mathcal{C}}_{\ell'}] = \frac{2}{(2\ell+1) f_{\rm sky}} \left(\mathcal{C}_\ell^{\mathcal{B}} + \mathcal{N}_{{\rm eff},\ell}^{\mathcal{B}}\right)^2 \delta_{\ell\ell'} \,. \tag{7.31}$$

Selecting the total signal in the angular interval $\Theta \equiv [\theta_{\rm min}, \theta_{\rm max}]$, with $\theta_{\rm min} \geq \theta_c$, defines an estimator of the superhorizon signal[8]

$$\widehat{\mathcal{S}}(\theta; \theta_{\rm min}) \equiv \sum_\ell \frac{2\ell+1}{4\pi} \widehat{\mathcal{C}}_\ell P_\ell(\cos\theta) \Pi(\theta) \,, \tag{7.32}$$

where

$$\Pi(\theta) \equiv \begin{cases} 1 & \theta \in [\theta_{\rm min}, \theta_{\rm max}] \\ \\ 0 & \text{otherwise} \end{cases} . \tag{7.33}$$

For now, we will keep $\theta_{\rm min}$ general. The precise definition of $\theta_{\rm max}$ is not important, but will be limited by the maximum angular extent of a partial sky observation. The covariance of the estimator is given by

$$\mathscr{C}[\widehat{\mathcal{S}}(\theta), \widehat{\mathcal{S}}(\theta')] = \sum_{\ell\ell'} \frac{2\ell+1}{4\pi} \frac{2\ell'+1}{4\pi} \mathscr{C}[\widehat{\mathcal{C}}_\ell, \widehat{\mathcal{C}}_{\ell'}] P_\ell(\cos\theta) P_{\ell'}(\cos\theta') \Pi(\theta) \Pi(\theta') \,. \tag{7.34}$$

We emphasize that the estimator (7.32) is *biased*, since the total signal contains spurious contributions from the filtered subhorizon modes (see §7.2). We will quantify this bias and define a debiased version of the estimator below.

Signal-to-noise. To define the signal-to-noise of the measurement, we discretize (7.32) and (7.34), and split the signal into N uniformly spaced angular bins $\Theta_b \equiv \{\theta_{(b)} \pm \frac{1}{2}\Delta\theta\}$, for $b = 1, \ldots, N$. A natural sampling interval is $\Delta\theta \simeq 180^\circ/\ell_\star$, where $\ell_\star$ is the multipole moment at which the covariance matrix (7.34) converges.[9] The average signal assigned to each bin is

$$\widehat{\mathcal{S}}_b \equiv \frac{1}{Z_b} \int_{\Theta_b} {\rm d}\theta \sin\theta \; \widehat{\mathcal{S}}(\theta) \,, \tag{7.35}$$

[7] Lensing induces a non-Gaussian contribution to the covariance matrix whose explicit expression can be found in [225]. We have checked that the degradation caused by the non-Gaussian lensing covariance is much smaller than the systematic uncertainties due to the leakage.

[8] In Appendix E, we define the harmonic space equivalent of the estimator (7.32).

[9] The convergence of (7.34) at $\ell_\star$ means that we effectively take into account $\ell_\star$ independent modes of $\mathcal{C}_\ell^{\mathcal{B}}$, in which case the rank of the matrix (7.31) is $\ell_\star$. Since the transformation from harmonic space to real space is linear, the rank of the corresponding covariance matrix in real space is also $\ell_\star$. By restricting to a proper subinterval, $\Theta \equiv [\theta_{\rm min}, \theta_{\rm max}]$, we effectively reduce the rank by a factor of $\sim 180^\circ/(\theta_{\rm max} - \theta_{\rm min})$. Thus, a natural sampling interval is $\Delta\theta = 180^\circ/\ell_\star$. (In practice, the optimal $\Delta\theta$ is slightly larger, since the signal decays before it reaches $\ell_\star$ and we also include the non-Gaussian part of the covariance.) Errors at different angular separations are strongly correlated within the interval Θ, and oversampling will result in an ill-behaved covariance matrix.

where $Z_b \equiv \int_{\Theta_b} \mathrm{d}\theta \, \sin\theta$ is a normalization factor. The binned covariance matrix is given by

$$\mathscr{C}_{bb'} \equiv \frac{1}{Z_b Z_{b'}} \int_{\Theta_b} \int_{\Theta_{b'}} \mathrm{d}\theta \, \mathrm{d}\theta' \sin\theta \sin\theta' \, \mathscr{C}[\widehat{\mathcal{S}}(\theta), \widehat{\mathcal{S}}(\theta')] \,, \tag{7.36}$$

and the signal-to-noise ratio is defined as

$$(\mathrm{S/N})^2 = \sum_{bb'} \widehat{\mathcal{S}}_b \mathscr{C}^{-1}_{bb'} \widehat{\mathcal{S}}_{b'} \,, \tag{7.37}$$

where $\mathscr{C}^{-1}_{bb'}$ is the inverse of (7.36). In §7.5, we will evaluate (7.37) for various experimental configurations.

Leakage and debiasing. Since the total signal $\mathcal{S}(\theta)$ contains spurious modes from the leakage of the filtered subhorizon modes, $\widehat{\mathcal{S}}(\theta)$ is a biased estimator of the true superhorizon signal (7.10).[10] We will quantify this bias by comparing the signal-to-noise of the expected total signal with that of the spurious subhorizon modes.

Let us write the estimator (7.32) as $\widehat{\mathcal{S}}(\theta) = \widetilde{\mathcal{S}}(\theta) + \mathcal{S}^\dagger(\theta)$, where $\widetilde{\mathcal{S}}(\theta)$ denotes the unbiased estimator (i.e. the estimator of the pure superhorizon component) and $\mathcal{S}^\dagger(\theta)$ is the subhorizon signal. The total signal-to-noise (7.37) can then be written as

$$\begin{aligned}(\mathrm{S/N})^2 &= \sum_{bb'} \left(\widetilde{\mathcal{S}}_b \mathscr{C}^{-1}_{bb'} \widetilde{\mathcal{S}}_{b'} + 2 \widetilde{\mathcal{S}}_b \mathscr{C}^{-1}_{bb'} \mathcal{S}^\dagger_{b'} + \mathcal{S}^\dagger_b \mathscr{C}^{-1}_{bb'} \mathcal{S}^\dagger_{b'} \right) \\ &\equiv (\mathrm{S/N})^2_+ + (\mathrm{S/N})^2_\times + (\mathrm{S/N})^2_- \,, \end{aligned} \tag{7.38}$$

where $(\mathrm{S/N})_+$ and $(\mathrm{S/N})_-$ denote the parts coming from the true superhorizon modes and the subhorizon leakage, respectively, while $(\mathrm{S/N})_\times$ stands for their cross-correlation. We will use

$$\delta \equiv \frac{(\mathrm{S/N})_-}{\mathrm{S/N}} \tag{7.39}$$

as a diagnostic tool for quantifying the amount of leakage and, hence, the bias in the estimator (7.32). For small values of δ, we know that the expected signal is dominated by the true superhorizon modes. We will consider optimizing the analysis (e.g. by adjusting ℓ_s and $\theta_{\rm min}$), so that we get the maximum signal-to-noise while keeping the leakage fraction (7.39) small. We typically take an acceptable leakage fraction to be $\delta \leq 0.1$.

Alternatively, we can correct for the bias of the estimator (7.32) through a simple debiasing procedure. Subtracting the expected ensemble average of the spurious subhorizon mode from the estimator (7.32) leads to an unbiased estimator of the pure superhorizon signal:

$$\widetilde{\mathcal{S}}(\theta) \equiv \widehat{\mathcal{S}}(\theta) - \mathcal{S}^\dagger(\theta) \,. \tag{7.40}$$

[10] Another type of bias arises from the E-B mixing in partial sky observations. This bias is well-understood and can be treated by substituting the pseudo-C_ℓ estimators considered in [226] for $\widehat{C}^{\mathcal{B}}_\ell$.

In this case, we can treat the subhorizon signal as an extra source of noise. Applying this debiasing procedure, we may improve the signal-to-noise by allowing a smaller smoothing scale ℓ_s and/or a larger angular interval Θ.

7.5 Signal-to-noise forecasts

Finally, we are ready to investigate the detectability of the superhorizon $\mathcal{B}$-mode signal for current and future experiments. The signal-to-noise will, of course, depend on the tensor-to-scalar ratio of the primordial fluctuations. We will consider both a fiducial value of $r = 0.13$ (which corresponds to the canonical value of $m^2\phi^2$ chaotic inflation [20] and is chosen as a representative of the typical values predicted in large-field models of inflation), as well as the wider range $r = [0.001, 0.2]$. When interpreting the presented forecasts, one should keep in mind that the value $r = 0.13$ is disfavored at more than 2σ by the current constraint $r < 0.07$ (95% CL) [43].

All CMB spectra are computed with CAMB [227] using the best-fit parameters of the ΛCDM model from the Planck 2013 results [228]: $h = 0.67$, $\Omega_{b,0}h^2 = 0.022$, $\Omega_{c,0}h^2 = 0.12$, $\tau = 0.093$, $A_s = 2.2 \times 10^{-9}$, and $n_s = 0.96$.[11] For simplicity, the primordial tensor spectrum is taken to be scale-invariant, $n_t = 0$.

Preliminaries. We will use the estimators (7.32) and (7.40) defined on the interval $\Theta = [\theta_{\rm min}, \theta_{\rm max}]$, with $\Delta\theta = 0.30°$. For simplicity, we will fix $\theta_{\rm max} = 6.0°$ throughout. For $\theta_{\rm min}$, we will consider two different choices:

(I) For the biased estimator (7.32), we compute the signal-to-noise on an interval with $\theta_{\rm min} = 2.6°$, where the leakage from subhorizon modes is guaranteed to be small and constrained by causality.

(II) For the debiased estimator (7.40), we compute the signal-to-noise on an extended interval with $\theta_{\rm min} = 1.0°$, which is where the filtered pure superhorizon signal[12] starts to become appreciable (cf. Fig. 7.3).

The estimator (I) is clearly more conservative, but also rejects a significant fraction of the inflationary superhorizon signal. The estimator (II), on the other hand, includes all superhorizon modes, but is less immune to spurious subhorizon contamination due to leakage. Although the known bias due to the inflationary subhorizon modes has been corrected for in the estimator (7.40), a signal on the interval $[1°, 2°]$ from non-inflationary sources is strictly speaking not forbidden by causality. To perform a true causality test of inflationary tensor modes, we therefore aim to detect the signal with the estimator (I). Nevertheless, we will also show results for the estimator (II)

[11] Notice that these (possibly with the exception of the optical depth, τ) are not substantially different from the parameters from the Planck 2015 results [21] shown in Table. 2.4.

[12] Below we will show that the Gaussian (7.20) with $\ell_s = 200$ is a conservative filter function. We will take this as our fiducial choice of filtering, but also investigate the possibility of optimizing the smoothing scheme in particular examples.

which quantifies the signal-to-noise of the total superhorizon signal from inflation. In that case, the caveat that we just stated should be kept in mind.

We will consider two sets of foreground models:

- Ground-based experiments (§7.5.1) can target small, but exceptionally clean, patches of the sky, and lower estimates for the foreground amplitudes are therefore appropriate.
- Space-based all-sky experiments (§7.5.2) can't use the cleanest patches only, so we will use higher foreground levels in those cases.

Our precise choices for the foreground amplitudes will depend on the experiment under consideration and will be presented in the following sections.

7.5.1 Ground-based experiments

We first consider the capabilities of ground-based experiments, as illustrated by a few representative examples.

Keck Array. The BICEP2 experiment has recently been upgraded to the Keck Array [229]. The Keck Array, unlike BICEP2, has multiple frequency channels, and the combination of its 95, 150, and 220 GHz detectors yields an effective noise of $\Delta_{P,\mathrm{eff}} = 2.2\,\mu\mathrm{K}'$ ($s_{P,\mathrm{eff}} = 14.2\,\mu\mathrm{K}'$). In the near future, the BICEP3 experiment [230] will start to observe the same part of the sky, with higher sensitivity at 95 GHz. In combination with the Keck Array, the effective noise will then reduce to $\Delta_{P,\mathrm{eff}} = 1.4\,\mu\mathrm{K}'$ ($s_{P,\mathrm{eff}} = 9\,\mu\mathrm{K}'$). In the following, we will refer to this combination of the Keck Array and BICEP3 simply as the 'Keck Array'.

Like for BICEP2, observations are made in the "Southern Hole" ($f_{\mathrm{sky}} = 0.024$), a region where both galactic and extragalactic foreground emissions are expected to be very low. For the foreground amplitudes in the Southern Hole, we will use the estimates given in [231]:

$$\mathcal{A}_s = \xi_s \times (1.5 \times 10^{-7}\,\mu\mathrm{K}^2)\,, \tag{7.41}$$

$$\mathcal{A}_d = \xi_d \times (1.8 \times 10^{-6}\,\mu\mathrm{K}^2)\,, \tag{7.42}$$

where these amplitudes are measured at $\nu_0 = 100\,\mathrm{GHz}$ and $\ell_0 = 100$. The parameters ξ_s and ξ_d allow for our uncertainties concerning the synchrotron and dust amplitudes in the Southern Hole. We will use $\xi_s \in [0.67, 1.33]$ and $\xi_d \in [0.33, 1.67]$ which corresponds to the 1σ uncertainties in [231].

Using the 220 GHz map of the Keck Array as a template, internal foreground removal of polarized dust emission at lower frequencies will be possible to some extent. This requires the spectral index of the dust signal to be well-constrained, which will be the case if external information from Planck is folded in. Our uncertainty in the level of foreground residuals that can ultimately be achieved will be characterized by the parameters ϵ_s and ϵ_d in (7.27).

The large beam size of the Keck Array ($\theta_b \sim 30'$) means that internal delensing will not be possible. Yet, a joint analysis with a higher resolution experiment observing the same part of the sky may allow some modest amount of delensing. SPTPol [232] is indeed also observing in the Southern Hole, but its current sensitivity is not at a level that would make delensing a realistic possibility. In the following, we will therefore assume Keck Array observations without any delensing as the default, i.e. $\epsilon_L = 1$ in (7.29), but also give results invoking a small amount of delensing, $\epsilon_L = \{0.5, 0.3\}$, as might become possible with an upgrade of SPTPol.

Simons Array. The Simons Array [233] is a planned successor of the PolarBeaR experiment [192, 195]. Located in the Atacama desert in Chile, it will provide high-resolution observations of a relatively large fraction of the sky ($f_{\rm sky} = 0.2$). The frequency bands of the Simons Array are the same as those of the Keck Array: 95, 150, 220 GHz. The effective noise level is $\Delta_{P,\rm eff} = 6.3\,\mu{\rm K}'$ ($s_{P,\rm eff} = 14.1\,\mu{\rm K}'$).

In the absence of detailed information about the polarized emission in the region observed by the Simons Array, we will use the same foreground levels (7.41) and (7.42) as for the Keck Array, with the same, relatively large, uncertainties. Its small beam size ($\theta_b = 2.7'$ at 220 GHz) allows the Simons Array to serve as a useful probe to the gravitational lensing of the CMB on small angular scales, and internal delensing will be possible to some degree. We will thus show results for $\epsilon_L = \{0.5, 0.3\}$.

Results. In Fig. 7.4, we present results for the signal-to-noise achievable by the Keck Array and the Simons Array for the fiducial value $r = 0.13$ as a function of the level of foreground cleaning ϵ_d. Shown are various levels of the delensing fraction $\epsilon_L = \{1, 0.5, 0.3\}$. We see that a 3σ detection will marginally be possible with the Simons Array if both delensing and foreground cleaning can be achieved to a relatively high standard. On the other hand, a detection with the Keck Array does not look feasible.

The above results were derived using our canonical choice of filtering: the Gaussian filter (7.20) with $\ell_s = 200$. Slight improvements in the signal-to-noise are possible by optimizing the smoothing scheme. Figure 7.5 shows the dependence of the signal-to-noise and the leakage fraction on the smoothing parameter ℓ_s for $r = 0.13$ and $\epsilon_L = \{\epsilon_s, \epsilon_d\} = 0.5$. We see that the signal-to-noise initially increases with ℓ_s, reaches a maximum at $\ell_s \simeq 120$, and then decreases as more small-scale noise is allowed for higher ℓ_s. At the maximum, S/N = 2.2 and 3.8 for the Keck Array and the Simons Array, respectively. The leakage fraction at the maximum is $\delta = 0.11$.

For optimal results, we pick the smoothing scale in such a way that it maximizes the signal-to-noise while keeping $\delta < 0.1$ for all values of r that yield S/N > 3. The optimal smoothing scale for both experiments is then $\ell_s = 150$, giving a 15% increase in the signal-to-noise (see Fig. 7.5).[13]

[13] We have also tested other forms of filtering functions. For example, using a tanh-filter, we were able to achieve a 10 to 20% improvement on the overall signal-to-noise with similar degrees of leakage for various parameters and values of r. This is because the tanh-filter is characterized by two smoothing parameters (the cut-off scale and the width), and this extra degree of freedom allows us to control the filtering process more precisely, giving us more

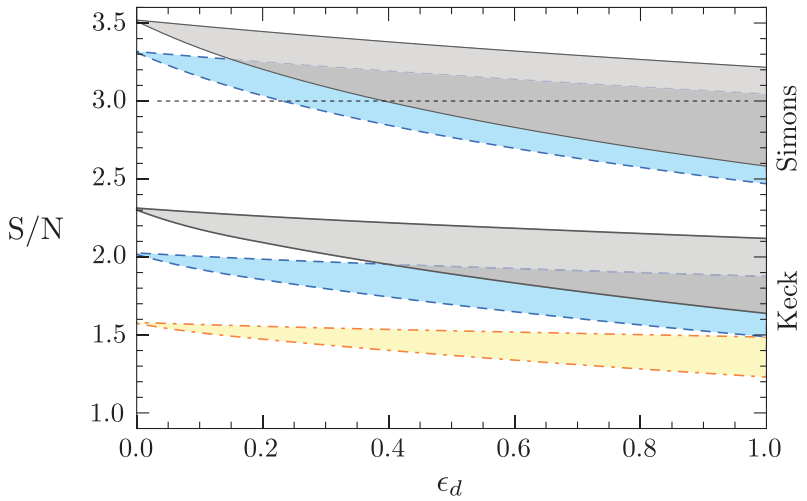


Figure 7.4: Signal-to-noise on the interval $[2.6°, 6.0°]$ for $r = 0.13$ as a function of ϵ_d. The plot shows experiments with Keck Array (bottom) and Simons Array (top) specifications for three different delensing fractions: $\epsilon_L = 1.0$ (yellow, dot-dashed), $\epsilon_L = 0.5$ (dashed, blue), and $\epsilon_L = 0.3$ (solid, gray). The bands correspond to the uncertainty in the foreground amplitudes, $\xi_s = [0.67, 1.33]$ and $\xi_d = [0.33, 1.67]$.

With this optimization, a more than 3σ detection becomes possible with the Simons Array even for only modest amounts of cleaning, $\epsilon_L = \{\epsilon_s, \epsilon_d\} = 0.5$. To achieve a similar level of significance with the Keck Array, we still require a high level of cleaning, $\epsilon_L = \{\epsilon_s, \epsilon_d\} = 0.1$.

One may argue that we have been too conservative by choosing $\theta_{\rm min} = 2.6°$ as our criterion for the superhorizon signal. In particular, as can be seen from Fig. 7.3, a large part of the inflationary superhorizon signal is not captured by this definition. In order to quantify the size of the total signal, we therefore also consider the extended interval with $\theta_{\rm min} = 1°$. We use the debiased estimator so that the known leakage of inflationary subhorizon modes is corrected for. Figure 7.6 shows the signal-to-noise on the interval $[1.0°, 6.0°]$ as a function of r without optimization of the filtering. We see that a 3σ detection will be possible if $r \gtrsim 0.1$ and 0.04 for the Keck Array and the Simons Array, respectively, assuming a modest amount of delensing and foreground removal of 50%. With the optimization described above, we get S/N > 3 if $r \gtrsim 0.05$ and 0.025 for the Keck Array and the Simons Array, respectively. While this detection wouldn't constitute a perfect causality test, it would still be a strong indication for inflationary superhorizon tensors. Moreover, at sufficiently high S/N it will be possible to measure the shape of the signal in Fig. 7.3, which would further strengthen this interpretation.

optimized results. However, for simplicity of presentation, all the results in the chapter were produced with the Gaussian filter (7.20).

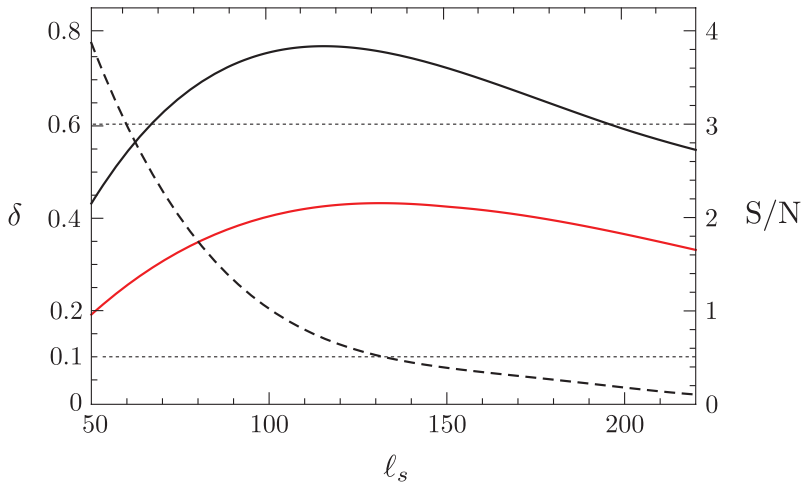


Figure 7.5: Signal-to-noise (solid) and leakage fraction (dashed) for $r = 0.13$ as a function of ℓ_s for experiments with Keck Array (red) and Simons Array (black) specifications. Only a single curve is shown for δ because the curves for the Keck Array and the Simons Array are almost identical. The plot assumes $\epsilon_L = \{\epsilon_s, \epsilon_d\} = 0.5$. Decreasing the smoothing scale from $\ell_s = 200$ to $\ell_s = 150$ increases the signal-to-noise by about 15%. The leakage fraction δ is less than 10% as long as $\ell_s \gtrsim 140$.

7.5.2 Space-based experiments

To perform a true causality test, the $\mathcal{B}$-mode signal has to be measured above $\theta_{\rm min} = 2.6^\circ$. We have seen that, for $r > 0.1$, this is (marginally) possible with ground-based experiments. For $r < 0.1$, on the other hand, a future satellite mission will be required. For purposes of illustration, we now examine the LiteBIRD [196] and COrE [193] proposals.[14]

All-sky surveys don't have the luxury of observing only the cleanest patches of the sky, so we need to adjust our estimates for the expected foreground levels accordingly. The level of polarized synchrotron emission is constrained by the WMAP polarization measurements between 23 and 94 GHz [198]. Those results imply

$$\mathcal{A}_s \;\simeq\; 5.8 \times 10^{-7}\,\mu\text{K}^2\,, \tag{7.43}$$

which is comparable to the 95% upper limit of the synchrotron amplitude determined by DASI [235]. For polarized dust emission, we take the template used by the Planck collaboration in [197, 202] which, for $f_{\rm sky} = 0.7$, gives

$$\mathcal{A}_d \;\simeq\; 5.5 \times 10^{-5}\,\mu\text{K}^2\,. \tag{7.44}$$

This choice is consistent with the FDS model [206] with an average polarization fraction of about 7%. Both of the above amplitudes are defined with respect to $\nu_0 = 100\,\text{GHz}$ and $\ell_0 = 100$.

LiteBIRD. LiteBIRD [196] is a next-generation full-sky satellite experiment, optimized to probe

[14] For our analysis, we will use the experimental configurations of the COrE mission proposed in 2011 [193]. The experimental speculations of the recent CORE proposal can be found in [234].

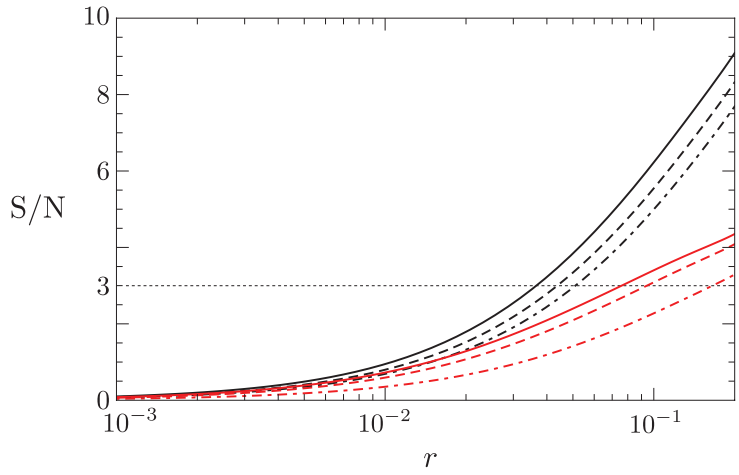


Figure 7.6: Signal-to-noise on the extended interval $[1.0°, 6.0°]$ as a function of r for experiments with Keck Array (red) and Simons Array (black) specifications. The foreground amplitudes have been fixed to the mean values in (7.41) and (7.42). The different curves correspond to $\epsilon_L = \{\epsilon_s, \epsilon_d\} = 0.1$ (solid), $\epsilon_L = 0.5$, $\{\epsilon_s, \epsilon_d\} = 0.1$ (dashed), and $\epsilon_L = \{\epsilon_s, \epsilon_d\} = 0.5$ (dot-dashed).

large-scale B-mode polarization. It is equipped with six frequency bands in the range from 60 to 280 GHz. This frequency coverage is wide enough to perform a high level of foreground removal of both synchrotron and dust [236]. We will therefore consider relatively small values of ϵ_s and ϵ_d, namely 0.1 (realistic) and 0.01 (optimistic). The large beams of the LiteBIRD experiment mean that delensing will only be possible in a joint analysis with external data sets [237]. We will assume that this will be possible only to a modest degree, $\epsilon_L \geq 0.5$.

COrE. COrE [193] is a proposed space mission which is anticipated to deliver a full-sky CMB polarization map with a sensitivity 10 to 30 times better than its predecessor Planck. With 15 frequency bands between 45 and 795 GHz, COrE will allow a very high degree of foreground cleaning, so we will consider $\{\epsilon_s, \epsilon_d\} = 0.1$ (pessimistic) and 0.01 (realistic). The small beams of COrE also mean that a significant amount of internal delensing can be achieved, so we take a delensing fraction of $\epsilon_L = 0.1$ as a realistic assumption [193].

Results. Figure 7.7 displays the signal-to-noise for LiteBIRD and COrE as a function of r. We see that a 3σ detection will be possible if $r > 0.04$ (0.01) with $\{\epsilon_s, \epsilon_d\} = 0.1$, and $r > 0.02$ (0.007) with $\{\epsilon_s, \epsilon_d\} = 0.01$, for LiteBIRD (COrE). Depending on the actual delensing level attained by these experiments, the detection bounds stated above may shift slightly. In any case, incorporating the optimization scheme described earlier, the signal-to-noise can be improved by about 20%. Thus, both LiteBIRD and COrE are capable of detecting the superhorizon $\mathcal{B}$-mode signal for $r \gtrsim 0.01$, in most realistic scenarios. For $0.001 < r < 0.01$, a statistically significant detection will only be possible if the extended interval $[1.0°, 6.0°]$ is used.

Summary. The conclusions of this section are summarized in Fig. 7.8, which shows the signal-to-noise on the interval $[2.6°, 6.0°]$ for $r = 0.13$ as a function of the sky fraction $f_{\rm sky}$ and the

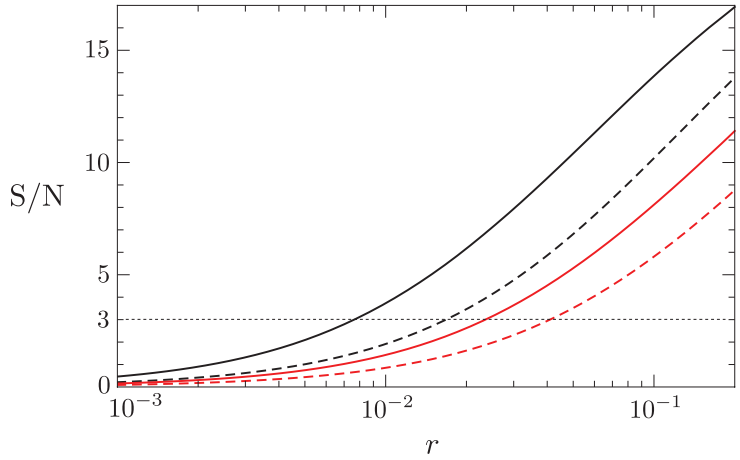


Figure 7.7: Signal-to-noise on the interval $[2.6°, 6.0°]$ as a function of r for experiments with COrE (black) and LiteBIRD (red) specifications. The solid lines correspond to $\{\epsilon_s, \epsilon_d\} = 0.01$, while the dashed lines assume $\{\epsilon_s, \epsilon_d\} = 0.1$. The delensing fractions have been fixed to $\epsilon_L = 0.5$ and $\epsilon_L = 0.1$ for LiteBIRD and COrE, respectively.

effective instrumental sensitivity $s_{P,\mathrm{eff}}$. This time the residual foreground amplitudes have been fixed to $\mathcal{A}_s = 5.8 \times 10^{-9}\,\mu\mathrm{K}^2$ and $\mathcal{A}_d = 5.5 \times 10^{-7}\,\mu\mathrm{K}^2$ at $\nu_0 = 100\,\mathrm{GHz}$, $\ell_0 = 100$. As we can see, for experiments with high instrumental sensitivity, $s_{P,\mathrm{eff}} \lesssim 20\,\mu\mathrm{K}'$, sky coverage is the main factor determining whether the signal is detectable. This is because $\mathrm{S/N} \propto \sqrt{f_{\mathrm{sky}}}$ in the cosmic variance limit, whereas $\mathrm{S/N} \propto 1/s^2_{P,\mathrm{eff}}$ for experiments dominated by instrumental noise. Hence, full-sky satellite missions have the best prospects for measuring the superhorizon $\mathcal{B}$-mode signal, though ground-based experiments such as the Simons Array can be feasible, if $r \gtrsim 0.1$.

7.6 Conclusions

The significance of a detection of primordial B-modes cannot be overstated [42, 51, 199, 238]. However, even if the signal is established to be of primordial origin, we still wish to determine whether it was generated by vacuum fluctuations during inflation or has an alternative, post-inflationary origin.

In this chapter, we have revisited the proposal of [177] for using the superhorizon part of the B-mode spectrum in real space as a model-insensitive diagnostic of inflationary gravitational waves. We found that the causality test for B-modes in its original form is not unambiguous, since we must deal with the issue of the mixing between subhorizon and superhorizon modes that is induced by the finite resolution of the experiment and the smoothing of the raw data. We have quantified this effect and shown how future experiments have to be designed in order to maximize the signal-to-noise of the superhorizon signal while rejecting unwanted contaminations from spurious subhorizon modes.

We have found that future ground-based experiments are capable of detecting the superhorizon $\mathcal{B}$-mode signal at more than 3σ significance, if the tensor-to-scalar ratio is as large as what can

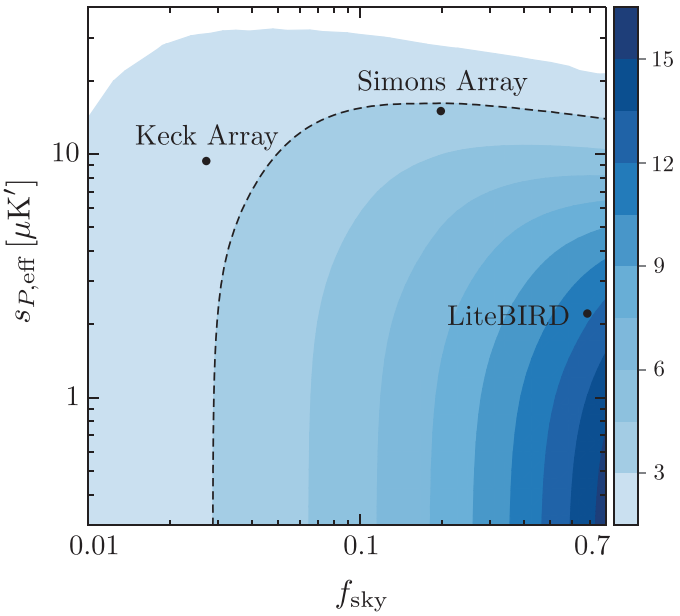


Figure 7.8: Signal-to-noise on the interval $[2.6°, 6.0°]$ for $r = 0.13$ as a function of f_{sky} and $s_{P,\mathrm{eff}}$. The plot was created using the optimized Gaussian filter with $\ell_s = 150$ and assumes 50% delensing. Moreover, it assumes a fixed beam size characteristic of Keck Array, Simons Array and LiteBIRD. Not shown therefore is the result for COrE which has smaller beams and therefore increased sensitivity to the signal (see Fig. 7.7). The dashed line indicates the 3σ detection bound.

be allowed by the current constraint [239], i.e. if $r \gtrsim 0.1$. If the value of r is significantly smaller, then the measurement will require a full-sky survey. We have found that a 3σ detection is possible with LiteBIRD and COrE as long as $r \gtrsim 0.01$, and if 90% foreground cleaning and more than 50% delensing can be achieved.

8. Discussion

Observations of the CMB anisotropies can be traced back to the moment of horizon crossing during inflation. These observations therefore probe energies of order the inflationary Hubble scale. One of the key challenges in cosmology is to relate these measurements to the unknown UV dynamics of inflation. In this thesis, we have approached the problem using the framework of effective field theory, which has allowed us to describe the inflationary dynamics in a model-insensitive way. We have explored tests of the physics of inflation using two cosmological observables: non-Gaussianity and tensor modes. In this final chapter, we will summarize the main results obtained in this thesis and make some remarks on their implications for future research.

8.1 Summary of the main results

Our strategy to understand the nature of inflation in this thesis involved two concepts: causality constraints[1] and phenomenology of inflationary models motivated by string theory.[2] As we will recapitulate in the following, the former puts non-trivial constraints on cosmological observables, whereas the latter can provide important hints for the UV completion of inflation.

▷ *Causality constraints.* The analyticity of response functions and scattering amplitudes implies powerful relations between low-energy observables and the underlying short-distance dynamics. In Chapter 4, we derived similar connections in inflation and obtained a new consistency condition in single-field inflation that constrains the size and sign of the four-point function in terms of the amplitude of the three-point function. A violation of this consistency condition would point towards less conventional models of inflation, and open a new window for model-building in inflation. More drastically, a violation could signal the breakdown of some basic properties of the UV completion of the EFT of inflation, such as causality, unitarity and Lorentz invariance. Hence, testing our consistency condition will provide useful information about the physics of inflation.

[1] Note that we considered two different implications of causality. In Chapter 4, we utilized causality in the UV theory of inflation to place non-trivial constraints on the parameters of the EFT of inflation. In Chapter 7, we used the causal structure of the spacetime to demand that distances that are separated by more than the horizon are out of causal contact.

[2] In Chapters 5 and 6, we considered models without a large hierarchy between the string scale and the Hubble scale. Inflationary models of this kind are hard to construct, since they would likely involve strongly-coupled dynamics for the background. On the other hand, these are the models that can lead to the most dramatic imprints of stringy effects. In this thesis, we have opted for a bottom-up approach and focused on the characteristic signatures of these effects on cosmological observables.

▷ *Particle spectroscopy.* The presence of extra fields is generically expected in realistic inflationary scenarios embedded in a concrete UV theory. For example, models of inflation in string theory typically involve a plentitude of moduli fields and Kaluza-Klein modes. In Chapter 5, we discussed the imprints of higher-spin particles—the low-energy manifestations of string excitations—on the non-Gaussianity of inflationary perturbations. Using the framework of the EFT of inflation, we classified the couplings of these particles to the Goldstone boson of broken time translations and the graviton. We found that these particles lead to distinct oscillations and an angular dependence in the correlation functions arising from the masses and the spins of the particles, respectively. We showed that it is possible to generate observable non-Gaussianity within the regime of validity of the effective theory, as long as the masses of the particles are close to the Hubble scale and their interactions break the approximate conformal symmetry of the inflationary background. Observing signatures of these higher-spin particles would be fascinating, since it would be indicative of inflation originating from string theory.

▷ *Higher-curvature effects.* Single-field slow-roll models of inflation predict a specific relation between the amplitude and scale-dependence of the tensor power spectrum, $r = -8\,n_t$. In Chapter 6, we showed that the coupling of the inflaton field to higher-curvature tensors can lead to a violation of this consistency condition. The correction to the tensor tilt from higher-curvature terms scales parametrically different ($\propto\sqrt{\varepsilon}\,H^2/M^2$) from the contribution from Einstein gravity ($\propto\varepsilon$) and can be the dominant effect if H/M is not too small. Moreover, the correction can have either sign, meaning that the tensor spectrum can have a blue tilt without necessarily violating the null energy condition. Together with the imprints of higher-spin fields on non-Gaussianity, observing higher-curvature corrections to the tensor spectrum would constitute an interesting way to probe stringy effects during inflation.

▷ *Superhorizon test for tensors.* To search for signatures of new physics in the tensor power spectrum, it would be helpful to first firmly establish the inflationary origin of B-modes. In the absence of inflation, correlations that are separated by more than the horizon scale at recombination, $\theta_c \approx 2.3^\circ$, are strictly forbidden by causality, while such superhorizon correlations are a characteristic of inflationary perturbations. In Chapter 7, we considered the prospects for measuring the inflationary superhorizon signature in future observations. This involves measuring the precise shape of the tensor power spectrum. We explained that the finite resolution of an experiment and the filtering of the raw data induces a transfer of spurious subhorizon power to superhorizon scales, and described ways to correct for it. We also provided a detailed treatment of possible sources of noise in the measurement, and presented forecasts for the detectability of the signal with future CMB polarization experiments. We consider this as a powerful model-independent way to test for the inflationary origin of tensor modes.

8.2 Outlook

The measurement of the CMB blackbody spectrum by COBE [11] marked the beginning of the era of precision cosmology. Since then, the CMB temperature power spectrum has been measured

at an unprecedented level of accuracy [13], saturating the cosmic variance limit over a vast range of scales. Nevertheless, the information about the primordial density perturbations that can be extracted from the two-point function of temperature fluctuations is limited, being characterized by just two numbers: the amplitude, A_s, and the scale-dependence, n_s. Although in a remarkable agreement with the basic predictions of inflation, the measurement of these two parameters alone is insufficient to unravel the mysteries of the microscopic mechanism responsible for inflation, making the need to detect other cosmological observables essential. In particular, primordial non-Gaussianity and tensor fluctuations can serve as powerful discriminators of models of inflation, and thereby have the potential to significantly enrich our understanding about the physics of inflation.

To date, observations are consistent with (nearly) Gaussian primordial density perturbations [40], as expected in single-field slow-roll inflation [25, 41]. Yet, many theoretically well-motivated models beyond this minimal paradigm naturally lead to higher levels of non-Gaussianity (see e.g. [240] for a review). Figure 8.1 is a schematic illustration of current and future constraints on (scale-invariant) primordial non-Gaussianities. We see that the perturbatively interesting regime spans about seven orders of magnitude in $f_{\rm NL}$. Of this regime, three orders of magnitude have been ruled out by current CMB observations, leaving a window of opportunity of about four orders of magnitude. Accessing these low levels of non-Gaussianity will be challenging. Even optimistic projections for future CMB observations won't reduce the constraints by more than an order of magnitude. Digging deeper will require new cosmological probes, such as observations of the large-scale structure (LSS) of the universe [241] and the tomography of the 21cm transition of neutral hydrogen gas [242]. The existing forecasts (e.g. [243]) suggest that one may hope to eventually verify the single-field slow-roll paradigm with futuristic 21cm surveys (and a lot of optimism!).

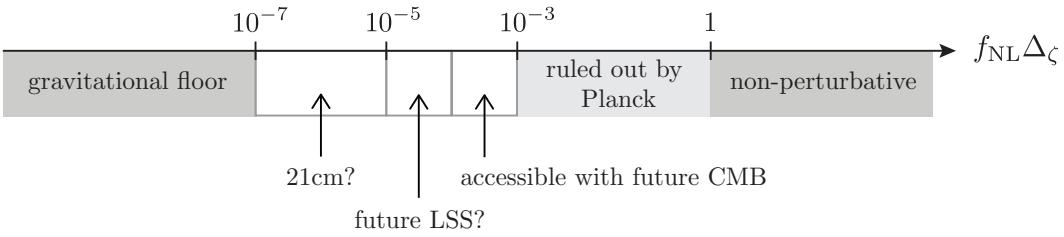


Figure 8.1: Schematic illustration of current and future constraints on (scale-invariant) primordial non-Gaussianity. The "gravitational floor" denotes the minimal level of non-Gaussianity created by purely gravitational interactions during inflation [25, 41].

On another observational front, large experimental efforts are currently underway to detect primordial gravitational waves (see [244] for a recent review). At present, the best constraint on r comes from the joint analysis of BICEP2/Keck Array and Planck [43], with an upper bound

of $r < 0.07$. Our knowledge on primordial tensors will be significantly improved by the next generation of B-mode experiments that have the sensitivity to reach $r = \mathcal{O}(10^{-3})$ [44, 45]: a natural threshold for a large class of inflationary models. A detection of primordial tensors would tell us two pivotal facts about inflation: that it occurred near the grand unification scale of 10^{16} GeV and that the inflaton underwent a super-Planckian excursion. This incredible feat would have a significant impact on our understanding of the microscopic origin of inflation and foster deeper connections between high-energy physics and cosmology.

With the advent of highly sensitive cosmological experiments, it is timely to ask how to put diverse models of inflation to the observational test. The landscape of inflationary scenarios is vast, which makes testing for individual models difficult. Moreover, the fact that the explicit construction of these models require the background description to be weakly coupled for theoretical control means that it does not entirely cover the perturbatively allowed regime for perturbations. As we described in this thesis, the methods of effective field theory offer a model-independent framework to describe interactions during inflation. In particular, we focused on imposing model-insensitive constraints on cosmological observables and finding distinctive UV signatures of inflation. To efficiently search for new physics in the upcoming observational data, a further exploration in this direction would be illuminating. To this end, we will present a list of possible avenues for future research, which stem from the work described in this thesis.

▷ *Veneziano correlator.* In analogy to scattering amplitudes being the tool for analyzing particle interactions in colliders, cosmological correlation functions can be used to probe the particle content during inflation. In Chapter 5, we studied the impact of individual higher-spin particles on cosmological correlation functions. However, the string spectrum in fact contains an infinite tower of higher-spin excitations of ever growing mass and spin, which restricts their scattering amplitude to take a unique form, the Veneziano amplitude. Similarly, one may expect the cosmological correlators to take a unique form when incorporating the infinite tower of intermediate higher-spin states [102, 245]. It would be interesting to see whether a characteristic trait of the string spectrum shows up in the squeezed limit of the bispectrum. Additionally, many aspects of this derivation may also shed light on the intriguing conjecture that string theory is the unique theory of massive higher-spin particles [246].

▷ *(Partially) massless higher-spin fields.* Another interesting generalization of our work would involve investigating the phenomenology of partially massless higher-spin fields during inflation [114] (see also [247]). These intriguing particles have no flat space analogs and possess extra symmetries that constrain their interactions. Unlike ordinary massive particles, their production rate is not exponentially suppressed and they may survive until late times, suggesting that they can have a more dramatic impact on cosmological observables. Moreover, it would be interesting to study the dynamics of purely massless higher-spin fields, whose interacting structure in de Sitter space is described by Vasiliev theory [248–251]. Coupling these fields to the inflaton may lead to a rich phenomenology and interesting applications to inflationary consistency relations [140].

▷ *Cosmological bootstrap.* Correlation functions in fact contain a much richer structure than scattering amplitudes,[3] so a natural extension of the work described in Chapter 4 is to directly analyze the consistency conditions on the former. In recent years, there has been great progress in using the conformal bootstrap to constrain conformal field theories (see e.g. [253] for a review). Since in an inflationary spacetime the de Sitter isometries act as conformal symmetries on the spatial boundary at late times, it would be interesting to use similar techniques to constrain inflationary observables. One of the technical challenges for cosmological applications is the necessity for a momentum-space formulation, which is customary in cosmological analyses as it makes spatial homogeneity manifest. Although the conformal three-point function in momentum space has been derived before [141, 254–257], the corresponding formula for the four-point function is currently lacking. The challenge would be to derive this and study the analogs of conformal blocks in the collapsed limit, for which the internal momentum is taken to be soft. This symmetry-based approach can offer a powerful model-independent way to constrain cosmological correlators, even in the non-perturbative regime.

▷ *Superhorizon test for NG.* With upcoming experiments, we will soon enter an observational regime where primordial density perturbations would no longer be the dominant source of non-Gaussianity in the CMB anisotropies. For example, second-order effects at recombination contribute to local non-Gaussianity at order $f_{\rm NL} = \mathcal{O}(1)$ (see e.g. [258–260]). In the presence of such a contamination (along with foregrounds and other systematics), it would be nice to have a clean way of testing for the inflationary origin of subdominant non-Gaussian signals. As explained in Chapter 5, inflation gives rise to superhorizon correlations at recombination that are otherwise forbidden by causality. The superhorizon part of higher-point functions can thus be used as a model-independent diagnostic of the inflationary nature of non-Gaussian correlations.

The importance of unveiling the precise microphysical mechanism of inflation cannot be overstated. Pursuing this goal requires both advancing our understanding of the theoretical foundations of inflationary cosmology and developing precise observational tests. In this thesis, we have made modest steps towards deciphering the intimate, yet subtle, link between high-energy physics and inflationary cosmology. We believe that the results presented in this thesis will help to find optimal observational strategies for extracting subtle signatures of the microscopic theory of inflation.

[3] This is due to the following non-trivial property that connects between correlation functions and scattering amplitudes: after analytic continuation of the momenta, the residue of the leading pole at zero total momentum reproduces the high-energy limit of the flat-space scattering amplitude [252].

A. Goldstone scattering

In this appendix, we provide supplementary material to Chapter 4. We first discuss the analytic properties of $\pi\pi \to \pi\pi$ scattering in §A.1. In §A.2, we present details of the analysis of the weakly coupled example of §4.3.1. Specifically, we will show how the sum rule (4.31) is realized in this particular example and demonstrate explicitly that it satisfies our positivity bound. In §A.3, we compute the low-energy amplitude of $\pi\pi \to \pi\pi$ scattering in the EFT of inflation at leading order in the derivative expansion.

A.1 Analyticity and pology

Without loss of generality, the non-forward scattering amplitude $\mathcal{M}$ may be written as a function of the following variables:

$$\begin{aligned} \omega_{12} &\equiv \omega_1 + \omega_2\,, \qquad & \mathbf{k}_{12} &\equiv \mathbf{k}_1 + \mathbf{k}_2\,, \\ \omega_{13} &\equiv \omega_1 - \omega_3\,, \qquad & \mathbf{k}_{13} &\equiv \mathbf{k}_1 - \mathbf{k}_3\,, \\ \omega_{14} &\equiv \omega_1 - \omega_4\,, \qquad & \mathbf{k}_{14} &\equiv \mathbf{k}_1 - \mathbf{k}_4\,. \end{aligned} \tag{A.1}$$

In the UV, i.e. for $\omega_a \gg \rho$, we expect the amplitude to become a function of the standard Mandelstam variables (s, t, u). Moreover, in the low-energy theory, some of the contributions to $\mathcal{M}$ may simplify to expressions in terms of the re-defined Mandelstam variables $(\tilde{s}, \tilde{t}, \tilde{u})$ associated with the re-scaled momenta, $\tilde{p}_a \equiv (\omega_a, c_s(\omega_a)\mathbf{k}_a)$. These contributions come from the terms in the effective action that mimic relativistic interactions after the rescaling of the spatial coordinates, e.g. $(\tilde{\partial}\pi)^4$. The general expression for $\mathcal{M}$ in the non-relativistic regime may (and will) contain additional Lorentz symmetry breaking combinations.

At low energies, the scattering amplitude computed in the EFT description must, of course, match the one computed in the full theory. Analyzing this matching in general may be cumbersome. However, for forward scattering in the center-of-mass frame ($\mathbf{k}_{ab} = 0$, $\omega_{13} = \omega_{14} = 0$, $\omega_{12} \equiv 2\omega$) some simplifications occur. In particular, the amplitude $\mathcal{A}_{\rm cm}$ can be expressed in terms of the square of the center-of-mass energy, $\omega_{12}^2 = 4\omega^2$, which in this frame is equal to both s and $\tilde{s}$. For notational simplicity, we will write the forward scattering amplitude in the center-of-mass frame as $\mathcal{A}_{\rm cm}(s) \equiv \mathcal{A}_{\rm cm}(4\omega^2)$. In the main text, we dropped the subscript 'cm', but here we keep it explicit in order to highlight expressions which are only valid in a fixed frame. The

distinction becomes important when studying the implications of crossing symmetry, since these are better described in a frame-independent manner and dropping the 'cm' subscript could lead to confusion.

The standard properties of the relativistic formalism (cf. §4.1.1) apply to the full amplitude in the UV. This means that any singularities in $\mathcal{A}_{\rm cm}(s)$ off the real axis, if present, would have to come from the non-relativistic low-energy behavior of the amplitude. On the one hand, for positive real s, the argument that restricts the non-analytic behavior to a minimum (to be consistent with unitarity and the optical theorem) remains unchanged. Moreover, for $s < -\rho^2$, crossing symmetry relates the amplitudes in the s- and u-channels, where similar considerations apply. On the other hand, for $-\rho^2 < s < 0$, we will demonstrate that crossing symmetry does not simply relate $\mathcal{A}_{\rm cm}(s)$ to $\mathcal{A}_{\rm cm}(-s)$, as in the relativistic case. However, except for some rather peculiar behavior, which we will discuss later, singularities for unphysical values of s will be associated with physical poles and/or branch cuts for physical values of s (albeit not directly symmetric points). We therefore do expect the Mandelstam hypothesis of maximal analyticity to hold, and any non-analytic behavior to be restricted to the real s-axis. It remains to be analyzed whether these singularities along the negative real axis, especially in the region $-\rho^2 < s < 0$, could jeopardize positivity of the sum rule discussed in §4.1.2. As we shall see, crossing symmetry still plays a major role in determining the location of the non-analytic behavior.

In quantum field theory, crossing symmetry follows from the properties of the Green's functions and the LSZ reduction formula [261]. Put simply, field operators may create an incoming particle or an outgoing anti-particle out of the vacuum. For a relativistic theory with identical scalar particles, it is easy to use this property to connect regions of the scattering amplitude when (s, t, u) are exchanged. For non-relativistic theories, the LSZ formula still applies at low energies, but the relation between the different channels becomes more subtle. In particular, for energies near the cutoff Λ and below the UV scale ρ, extra poles or cuts may develop. The computation in terms of field operators implies that the crossing symmetry between the s- and u-channels relates the scattering amplitude under the exchange $\omega_2 \leftrightarrow -\omega_4$ and $\mathbf{k}_2 \leftrightarrow -\mathbf{k}_4$. At forward scattering, this transformation implies

$$\tilde{s} \equiv \omega_{12}^2 - c_s^2 \mathbf{k}_{12}^2 = 4\omega^2 \quad \leftrightarrow \quad \tilde{u} \equiv \omega_{14}^2 - c_s^2 \mathbf{k}_{14}^2 = -4\omega^2 = -\tilde{s}\,, \tag{A.2}$$

where we have evaluated the expressions on-shell. The part of the amplitude that is only a function of $\tilde{s}$ (in a generic frame) is therefore an even function of $\tilde{s}$. However, in principle the scattering amplitude $\mathcal{M}$ also has contributions that do not transform as easily. For instance, $\omega_{12} \leftrightarrow \omega_{14}$ under the crossing symmetry, but ω_{14} vanishes in the center-of-mass frame, while $\tilde{s} = \omega_{12}^2$ prior to the crossing transformation. Terms that vanish in the center-of-mass frame, e.g. those proportional to ω_{14}, play a vital role in making crossing symmetry manifest. For this reason, it is useful to distinguish functional dependence on $\tilde{s}$ from explicit functions of ω_{12}.

To illustrate these considerations, let us study the exchange of a heavy state in the s-channel,

away from forward scattering and in a generic frame. Using standard 'pology' arguments [261], we expect the amplitude to take the following form

$$\mathcal{M}_s \subset \frac{\mathcal{Z}(\omega_{ab}, \mathbf{k}_{ab}\cdot\mathbf{k}_{cd})}{\omega_{12}^2 - c_r^2\mathbf{k}_{12}^2 - M^2 + i\epsilon}\,, \tag{A.3}$$

where M is the energy of the intermediate state, $c_r \equiv c_r(\omega_{12})$ is its speed of propagation, and $\mathcal{Z}$ is some unknown function of the quantities defined in (A.1). This expression must be symmetric with respect to permutations of the momenta that leave the s-channel fixed: i.e. $\{1 \leftrightarrow 2\}$ and $\{3 \leftrightarrow 4\}$. It is useful to write the amplitude in terms of variables that make this invariance manifest (after using the on-shell conditions), namely[1]

$$\mathcal{Z}(\omega_{ab}, \mathbf{k}_{ab}\cdot\mathbf{k}_{cd}) \equiv \mathcal{Z}(\tilde{s}, \omega_{12}^2, \omega_{13}^2+\omega_{14}^2, \mathbf{k}_{13}\cdot\mathbf{k}_{14}, \omega_{13}\omega_{14})\,, \tag{A.4}$$

where we have chosen to express $\mathbf{k}_{12}^2$ in terms of $\tilde{s}$ and ω_{12}. We can then use crossing symmetry to determine the location of the pole in the u-channel, which we denote by $\mathcal{M}_u$. Putting both contributions together, we find

$$\begin{aligned}\mathcal{M} &= \mathcal{M}_s + \mathcal{M}_u \\ &\subset \frac{\mathcal{Z}(\tilde{s}, \omega_{12}^2, \omega_{13}^2+\omega_{14}^2, \mathbf{k}_{13}\cdot\mathbf{k}_{14}, \omega_{13}\omega_{14})}{\omega_{12}^2 - c_r^2\mathbf{k}_{12}^2 - M^2 + i\epsilon} + \frac{\mathcal{Z}(\tilde{u}, \omega_{14}^2, \omega_{13}^2+\omega_{12}^2, \mathbf{k}_{13}\cdot\mathbf{k}_{12}, \omega_{13}\omega_{12})}{\omega_{14}^2 - c_r^2\mathbf{k}_{14}^2 - M^2 + i\epsilon}\,.\end{aligned} \tag{A.5}$$

Taking the forward limit, $\mathbf{k}_{13} \to 0$ — but still in a generic frame — we get

$$\mathcal{A} \subset \frac{\mathcal{Z}(\tilde{s}, \omega_{12}^2, \omega_{14}^2)}{\omega_{12}^2 - c_r^2\mathbf{k}_{12}^2 - M^2 + i\epsilon} + \frac{\mathcal{Z}(-\tilde{s}, \omega_{14}^2, \omega_{12}^2)}{\omega_{14}^2 - c_r^2\mathbf{k}_{14}^2 - M^2 + i\epsilon}\,, \tag{A.6}$$

where $\mathcal{Z}(x,y,z) \equiv \mathcal{Z}(x,y,z,0,0)$. At high energies, $\omega \gg \rho$, Lorentz invariance is restored and we expect crossing symmetry to act in the familiar way. To see this, we note that $c_s(\omega), c_r(\omega) \to 1$ and $\tilde{s} \to s$ in the UV. Moreover, the amplitude will be dominated by a relativistically invariant function,

$$\mathcal{Z}(\tilde{s}, \ldots) \xrightarrow{\omega\gg\rho} \mathcal{Z}(s, \ldots) = \mathcal{Z}^{\text{UV}}(s)\big(1 + \mathcal{O}(\rho/\omega)\big)\,, \tag{A.7}$$

where $\mathcal{Z}^{\text{UV}}(s) = \mathcal{Z}^{\text{UV}}(-s)$, as required by crossing symmetry when the theory becomes relativistic. As expected, the expression in (A.6) therefore becomes symmetric under $s \to -s$ (and $\epsilon \to -\epsilon$). This is also manifest in the center-of-mass frame, where we have

$$\mathcal{A}_{\text{cm}}(s) \subset \frac{\mathcal{Z}^{\text{UV}}_{\text{cm}}(s)}{s - M^2 + i\epsilon} + \frac{\mathcal{Z}^{\text{UV}}_{\text{cm}}(-s)}{-s - M^2 + i\epsilon}\,, \qquad \text{for } s \gg \rho^2\,. \tag{A.8}$$

At low energies, on the other hand, crossing symmetry does not guaranteed that $\mathcal{A}_{\text{cm}}(s)$ is an

[1] To avoid a proliferation of different names we abuse notation and denote both functions in (A.4) by $\mathcal{Z}(\cdots)$.

even function of s. Instead, we have

$$\mathcal{A}_{\rm cm}(s) \subset \frac{\mathcal{Z}_{\rm cm}(s,s,0)}{s-M^2+i\epsilon}+\frac{\mathcal{Z}_{\rm cm}(-s,0,s)}{-c_r^2c_s^{-2}\,s-M^2+i\epsilon}\,, \qquad \text{for } s<\rho^2\,, \tag{A.9}$$

and the two terms are not necessarily related by reflection. First of all, when $c_r \neq c_s$, the location and residue of the pole on the negative axis is not the symmetric counterpart of the one on the positive axis. (This is seen explicitly in the perturbative example discussed in §4.3.1 and §A.2) Furthermore, while the optical theorem forces the residue of the s-channel pole in (A.9) to be positive, this does not imply positivity of the residue of the u-channel pole. Unitarity alone is not sufficient to guarantee positivity because the function $\mathcal{Z}_{\rm cm}(x,y,z)$ is evaluated for different arguments in the s- and u-channels. We may then worry that the residue from the negative s-axis may be negative, and dominate over the positive contribution from the s-channel. Fortunately, in many circumstances we find that $\mathcal{Z}$ is invariant under permutations of ω_a, such that

$$\mathcal{Z} \to \mathcal{Z}(\tilde{s},\omega_1\omega_2\omega_3\omega_4)\,. \tag{A.10}$$

For example, this property arises when time derivatives act on the external legs. In the center-of-mass frame, this means that[2]

$$\mathcal{Z}(\tilde{s},\omega_1\omega_2\omega_3\omega_4) \;\to\; \mathcal{Z}_{\rm cm}(s,\omega^4)=\mathcal{Z}_{\rm cm}(s,s^2)\,, \tag{A.11}$$

which extends the original form of the crossing symmetry to all energies. The residues on the positive and negative s-axes are therefore related, and both constrained to be positive by the optical theorem. This is also manifestly true in the example of §4.3.1 and a large class of weakly coupled extensions.

The above reasoning takes into account poles and branch cuts that originate in the s-channel. There is, however, a final subtlety to be discussed. Since Lorentz invariance is broken, interactions can in principle have an unequal number of time and space derivatives. For example, a quartic interaction with three time derivatives and two spatial derivatives can be consistent with the symmetries of the EFT. This could in principle produce contributions in the low-energy limit of the form

$$\mathcal{A} \propto \omega_{12}^5 = (\omega_{12}^2)^{5/2} \to \mathcal{A}_{\rm cm} \propto s^{5/2}\,. \tag{A.12}$$

To be consistent with unitarity, we must choose these cuts to run along the negative axis. Notice, however, that in a generic frame crossing symmetry maps $\omega_{12} \to \omega_{14} \to 0$. Therefore these type of singularities do not have an s-channel counterpart. Furthermore, while potentially dangerous, these terms are always subdominant in perturbation theory, since they involve higher powers of s. Moreover, they must be absent in the UV theory, which is dominated by relativistic interactions.

[2] In a perturbative setting, attempts to put an unequal number of time derivatives on each leg fail to produce singularities from the u-channel in the center-of-mass frame. After summing over permutations, the u-channel amplitude becomes a function of ω_{14}, which vanishes when $\omega_1=\omega_4=\omega$.

Hence, at lowest order in s, these rather peculiar singularities do not present a problem for our positivity argument.

A.2 Positivity in the $\pi\sigma$-model

For convenience, we recall the Lagrangian for the $\pi\sigma$-model:

$$\mathcal{L} = -\frac{1}{2}(\partial\bar{\pi})^2 - \frac{1}{2}(\partial\sigma)^2 - \rho\sigma\dot{\bar{\pi}} - \frac{\sigma(\partial\bar{\pi})^2}{2M} - \frac{1}{2}m^2\sigma^2 - \frac{1}{3!}\mu\sigma^3\,. \tag{A.13}$$

In the flat space limit, the linearized equations of motion are given by

$$(\omega^2 - k^2)\pi + i\rho\omega\sigma = 0\,, \tag{A.14}$$

$$(\omega^2 - k^2 - m^2)\sigma - i\rho\omega\bar{\pi} = 0\,, \tag{A.15}$$

so that the propagator for $\boldsymbol{\phi} \equiv (\bar{\pi}\ \sigma)$ is

$$\langle T(\boldsymbol{\phi}_p \boldsymbol{\phi}^T_{-p})\rangle = \frac{i}{(\omega^2 - k^2)(\omega^2 - k^2 - m^2) - \omega^2\rho^2 + i\epsilon} \begin{pmatrix} \omega^2 - k^2 - m^2 & -i\rho\omega \\ i\rho\omega & \omega^2 - k^2 \end{pmatrix}. \tag{A.16}$$

The poles of the propagator are associated with the non-trivial solutions for $\bar{\pi}$ and σ, which satisfy

$$\omega^2_\pm = k^2 + \tfrac{1}{2}(\rho^2 + m^2) \pm \sqrt{\rho^2 k^2 + \tfrac{1}{4}(\rho^2 + m^2)^2}\,. \tag{A.17}$$

The mixing of $\bar{\pi}$ and σ presents an additional complication because at low energies neither $\bar{\pi}$ nor σ creates an energy eigenstate. To correct for this, we will compute the S-matrix elements using the LSZ formula [68]:

$$S = \left(\prod_{a=1}^{4} \lim_{\omega_a \to E_a} \frac{\omega_a^2 - E_a^2}{Z(\omega_a)}\right) \langle T(\pi_{p_1}\pi_{p_2}\pi_{p_3}\pi_{p_4})\rangle\,, \tag{A.18}$$

where E_a is the energy of the gapless state. The function $Z(\omega)$ is the relative normalization between $\bar{\pi}$ and the canonically normalized energy eigenmode,

$$\langle T(\bar{\pi}_p \bar{\pi}_{-p})\rangle = \frac{Z^2(\omega)}{\omega^2 - E^2(k)}\,, \tag{A.19}$$

which in this particular case is given by

$$Z(\omega_a = E_a(k_a)) = \left(\frac{m^2 - \rho^2 + \sqrt{4k^2\rho^2 + (m^2+\rho^2)^2}}{2\sqrt{4k^2\rho^2 + (m^2+\rho^2)^2}}\right)^{1/2}. \tag{A.20}$$

$\boldsymbol{\mu = 0}$. We first consider the special case $\mu = 0$. The forward scattering amplitude gets contributions from exchange diagrams that include all the matrix elements in the propagator (A.16).

There are three classes of these diagrams: σ-exchange, π-exchange and $\pi\sigma$-exchange. At low energies, and for $c_s \ll 1$, the σ-exchange contribution dominates the amplitude. In the forward limit, we then find the following amplitude in the center-of-mass frame

$$\mathcal{A} = -\frac{Z^4(k^2)}{M^2}\left\{(\omega^2+k^2)^2\left[\frac{1}{4\omega^2-m^2-\rho^2}-\frac{1}{4k^2+m^2}\right]-(\omega^2-k^2)^2\frac{1}{m^2}\right\}, \tag{A.21}$$

where the last term is from the t-channel exchange. In the limit $\omega \to 0$, this amplitude indeed matches the result of the EFT computation. We see that the amplitude has poles at $4\omega^2 = m^2+\rho^2 \simeq \rho^2$ and $4k^2 = -m^2$ (or $4\omega^2 \simeq -\frac{3}{4}c_s^4\rho^2$); cf. Fig. 4.3. For small c_s, the pole on the negative axis is located much closer to the origin than that on the positive axis.

We wish to see how the sum rule (4.31) works for the amplitude (A.21). It is easy to see that the residue of the pole on the negative axis dominates: the pole on the positive axis is suppressed by a factor of c_s^2, while the relativistic regime, $M > \omega \gg \rho$, only contributes $\ln(M/\rho)/M^4$. Using that the imaginary part associated with the pole on the negative axis is

$$\mathrm{Im}[\mathcal{A}(s<0)] = \frac{Z^4(k^2)}{M^2}(\omega^2+k^2)^2\,\pi\,\delta(-4k^2-m^2), \tag{A.22}$$

the sum rule can be written as

$$\frac{1}{2}\mathcal{A}''(s\to 0) = \frac{1}{\pi}\int_{-\infty}^{0}\frac{\mathrm{d}s}{s^3}\,\mathrm{Im}[\mathcal{A}(s)] = \int_0^\infty \mathrm{d}q\,\frac{s'(q)}{s^3(q)}\frac{Z^4(q)}{M^2}\frac{q^2}{16}\,\delta(q-m^2), \tag{A.23}$$

where $q=-4k^2$ and

$$s(q) \equiv -q+2(m^2+\rho^2)-2\sqrt{-q\rho^2+(m^2+\rho^2)^2}, \tag{A.24}$$

$$s'(q) \equiv \frac{ds}{dq} = -1+\frac{\rho^2}{\sqrt{-q\rho^2+(m^2+\rho^2)^2}}. \tag{A.25}$$

At leading order in $c_s = m/\rho \ll 1$, we have

$$s(q=m^2) = -\frac{3}{4}c_s^2, \quad s'(q=m^2) = -\frac{3}{2}c_s^2, \quad Z^2(q=m^2) = \frac{3}{4}c_s^2. \tag{A.26}$$

Substituting this into (A.23), we find

$$\frac{1}{2}\mathcal{A}''(s\to 0) = \frac{1}{8m^2M^2} = \frac{1}{8|\dot{\bar{\phi}}|^2c_s^2}, \tag{A.27}$$

where we have used $m = c_s\rho$ and $\rho = |\dot{\phi}_0|/M$. The left-hand side of (A.27) can also be computed directly in the EFT for the canonically normalized field $\pi_c = c_s\pi/|\dot{\phi}_0|$ (after integrating out σ). In the limit $c_s \ll 1$, eq. (4.23) becomes

$$\frac{1}{2}\mathcal{A}''(s\to 0) = c_s^3\times\frac{1}{8}\frac{1}{\Lambda^4} = \frac{1}{8|\dot{\bar{\phi}}|^2c_s^2}, \tag{A.28}$$

where we used $\Lambda = f_\pi c_s$ and $f_\pi^4 = |\dot{\bar{\phi}}|^2 c_s$.[3] We thus find exact agreement, at leading order in $c_s \ll 1$, with the single pole contribution to the dispersion relation.

$\mu \neq 0$. Finally, we compute the forward scattering amplitude for $\mu \neq 0$. We will assume that μ is sufficiently large that we can neglect all other cubic terms. This example generates large c_3 and c_4 in the EFT. We wish to determine whether the derived EFT parameters satisfy our positivity constraint. A similar computation to the one above gives the $\mathcal{O}(\mu^2)$ contribution to the forward amplitude

$$\mathcal{A}_{\mu^2} = -\frac{\mu^2}{Z^4(\omega)}\left(\frac{\omega\rho}{\sqrt{4p^2\rho^2+(m^2+\rho^2)^2}}\right)^4\left[\frac{1}{4\omega^2-m^2-\rho^2}-\frac{1}{4p^2+m^2}-\frac{1}{m^2}\right]. \tag{A.29}$$

The analytic properties of this amplitude are similar to the previous case with poles located at $4\omega^2 = m^2+\rho^2 \simeq \rho^2$ and $4k^2 = -m^2$. In the limit $\omega \to 0$, we get

$$\mathcal{A}_{\mu^2} \to \frac{1}{8}\frac{\mu^2}{m^6}\, s^2\,, \tag{A.30}$$

which is manifestly positive.

Although the analysis of this section was a non-trivial check of our positivity constraint, the underlying reason for the positivity was already anticipated in §A.1. Specifically, the low-energy amplitude was UV completed through the exchange of a single heavy state. As a result, the coefficient function, Z, must scale as s^2 in order to match the low-energy scaling of the EFT.[4] It is clear that this scaling arises from a single derivative acting on each external leg and therefore Z is manifestly crossing symmetric. As a result, the residues of the u- and s-channel poles must have the same sign, and therefore the forward amplitude must be positive. One can check that this conclusion cannot be altered by changing the form of the interactions or of the mixing term. We conclude that positivity of the sum rule is a generic feature of weakly coupled UV completions of the EFT of inflation.

A.3 Computation of the amplitude

At tree level, we have two types of diagrams: (i) exchange diagrams involving the combination of two cubic vertices and (ii) contact diagrams involving quartic vertices. We will treat these two scattering processes in turn.

[3] In §4.2, we rescaled the coordinates by $\tilde{x}^i = x^i/c_s$. This rescaling changes the normalization of the amplitude. We have corrected for this difference by rescaling the result of §4.2 by a factor of c_s^3.

[4] One may also have $Z \propto s$ in such a way that the leading contributions to the s- and u-channels cancel in the limit $s \to 0$, leaving $\mathcal{A}(s) \propto s^2$, as required. Such a cancelation will only occur when the sign of the u-channel term is consistent with a positive contribution to our sum rule and therefore does not present a loophole to this argument.

Exchange diagrams. The Lagrangian at cubic order is

$$\tilde{\mathcal{L}}_3 = \frac{1}{\Lambda^2}\left[\alpha_1\,\dot{\pi}_c^3 - \alpha_2\,\dot{\pi}_c(\tilde{\partial}\pi_c)^2\right], \tag{A.31}$$

where the parameters α_i are defined in (4.16). For each exchange diagram, we get factors of $\frac{1}{2}i^2$ from the two vertices, $i^3(-i)^3$ from the six momenta, and i from the propagator, leading to an overall factor of $-\frac{1}{2}i$. The two interactions in (A.31) lead to three different types of exchange diagrams:

- $\boldsymbol{\dot{\pi}^3 \times \dot{\pi}^3}$. We first consider the diagram involving two factors of the interaction $\dot{\pi}^3$. The internal contraction for this diagram only involves time derivatives, which implies that only the s-channel is non-vanishing in the center-of-mass frame (using $\omega_{13} = \omega_{14} = 0$). There are $3^2 = 9$ ways of choosing this internal contraction and $4 \times 2 = 8$ diagrams for the s-channel; hence the symmetry factor in this case is 72. The vertices give a factor of α_1^2/Λ^4, and we get

$$i\mathcal{M}_{\dot{\pi}^3} = -\frac{1}{2}i \cdot 72 \cdot \frac{\alpha_1^2}{\Lambda^4} \cdot \Big[\omega^2(2\omega)\Big]\,\frac{1}{s}\,\Big[\omega^2(2\omega)\Big] = -\frac{9i}{4}\,\alpha_1^2\,\frac{s^2}{\Lambda^4}\,, \tag{A.32}$$

  where the final equality holds in the center-of-mass frame.

- $\boldsymbol{\dot{\pi}(\partial\pi)^2 \times \dot{\pi}(\partial\pi)^2}$. The computation of the diagram involving two factors of $\dot{\pi}(\partial\pi)^2$ is slightly more involved. Now there are three possible internal contractions:

  - $\dot{\pi}\dot{\pi}$. This internal contraction consists of time derivatives only, so only the s-channel survives. Since there is only one way of choosing the internal contraction, the symmetry factor is 8 and we get

$$i\mathcal{M}_{\dot{\pi}(\partial\pi)^2,a} = -\frac{1}{2}i \cdot 8 \cdot \frac{(-\alpha_2)^2}{\Lambda^4}\Big[(\tilde{p}_1\cdot\tilde{p}_2)2\omega\Big]\,\frac{1}{s}\,\Big[(\tilde{p}_3\cdot\tilde{p}_4)2\omega\Big] = -i\,\alpha_2^2\,\frac{s^2}{\Lambda^4}\,. \tag{A.33}$$

  - $\dot{\pi}\partial\pi$. Again, only the s-channel contributes, but now there are 4 possible ways of choosing the internal contraction, giving a symmetry factor of $8 \times 4 = 32$. The amplitude is

$$i\mathcal{M}_{\dot{\pi}(\partial\pi)^2,b} = -\frac{1}{2}i \cdot 32 \cdot \frac{(-\alpha_2)^2}{\Lambda^4}\Big[\omega\,\tilde{p}_1\cdot(\tilde{p}_1+\tilde{p}_2)\Big]\,\frac{1}{s}\,\Big[2\omega(\tilde{p}_3\cdot\tilde{p}_4)\Big] = -2i\,\alpha_2^2\,\frac{s^2}{\Lambda^4}\,. \tag{A.34}$$

  - $\partial\pi\partial\pi$. Since there are no time derivatives appearing in the internal contraction this time, naively we would expect that both the t- and u-channels would contribute. However, it turns out that both vanish in this case too. To see this, note that the scattering amplitude in the t-channel contains terms such as $(\tilde{p}_1-\tilde{p}_3)\cdot(\tilde{p}_1+\tilde{p}_3) = 0$ (and similarly for the u-channel), giving zero amplitude.[5] Noting that the symmetry factor for the

[5]The absence of low-energy poles is a genuine feature for all tree level exchange diagrams in the EFT of

s-channel is again 32, we find

$$i\mathcal{M}_{\dot\pi(\partial\pi)^2,c} = -\frac{1}{2}i\cdot 32\cdot\frac{(-\alpha_2)^2}{\Lambda^4}\Big[\omega\,\tilde p_1\cdot(\tilde p_1+\tilde p_2)\Big]\,\frac{1}{s}\,\Big[(\tilde p_3+\tilde p_4)\cdot\tilde p_3\,\omega\Big] = -i\,\alpha_2^2\,\frac{s^2}{\Lambda^4}\,. \quad \text{(A.35)}$$

- $\boldsymbol{\dot\pi^3\times\dot\pi(\partial\pi)^2}$. Finally, we consider the exchange diagram involving both interactions, $\dot\pi^3$ and $\dot\pi(\partial\pi)^2$. There are two such cross-terms, each with amplitude proportional to $\alpha_1\alpha_2/\Lambda^4$. We have two types of internal contractions:
  - $\dot\pi\dot\pi$. There are three ways of obtaining this internal contraction, giving the symmetry factor of $3\times 8 = 24$ for the s-channel. We therefore have

$$i\mathcal{M}_{\dot\pi^3\times\dot\pi(\partial\pi)^2,a} = -i\frac{1}{2}\cdot 2\cdot 24\cdot\frac{-\alpha_1\alpha_2}{\Lambda^4}\Big[(\tilde p_1\cdot\tilde p_2)2\omega\Big]\,\frac{1}{s}\,\Big[2\omega\omega^2\Big] = -3i\,\alpha_1\alpha_2\,\frac{s^2}{\Lambda^4}\,. \quad \text{(A.36)}$$

  - $\dot\pi\partial\pi$. The number of terms with this internal contraction is $3\times 2 = 6$, so the symmetry factor is $6\times 8 = 48$. We get

$$i\mathcal{M}_{\dot\pi^3\times\dot\pi(\partial\pi)^2,b} = -i\frac{1}{2}\cdot 2\cdot 48\cdot\frac{-\alpha_1\alpha_2}{\Lambda^4}\Big[\omega\tilde p_1\cdot(\tilde p_1+\tilde p_2)\Big]\,\frac{1}{s}\,\Big[2\omega\omega^2\Big] = -3i\,\alpha_1\alpha_2\,\frac{s^2}{\Lambda^4}\,. \quad \text{(A.37)}$$

Contact diagrams. The Lagrangian at quartic order is

$$\tilde{\mathcal{L}}_4 = \frac{1}{\Lambda^4}\Big[\beta_1\,\dot\pi_c^4 - \beta_2\,\dot\pi_c^2(\tilde\partial\pi_c)^2 + \beta_3\,(\tilde\partial\pi_c)^4\Big]\,, \quad \text{(A.38)}$$

where the parameters β_i are defined in (4.17). For each contact diagram, we get an overall factor of $i(-i)^2i^2 = i$. The three interactions in (A.38) lead to the following amplitudes:

- $\boldsymbol{\dot\pi^4}$. In the center-of-mass frame, this quartic interaction has equal contributions from s-, t- and u-channels, and comes with a symmetry factor of 24, giving

$$i\mathcal{M}_{\dot\pi^4} = i\cdot 24\cdot\frac{\beta_1}{\Lambda^4}\omega^4 = \frac{3i}{2}\,\beta_1\,\frac{s^2}{\Lambda^4}\,. \quad \text{(A.39)}$$

- $\boldsymbol{\dot\pi^2(\partial\pi)^2}$. For this interaction each channel comes with a symmetry factor of 8, and we get

$$\begin{aligned} i\mathcal{M}_{\dot\pi^2(\partial\pi)^2} &= i\cdot 8\cdot\frac{-\beta_2}{\Lambda^4}\Big[\omega^2(\tilde p_1\cdot\tilde p_2)+\omega^2(\tilde p_1\cdot\tilde p_3)+\omega^2(\tilde p_1\cdot\tilde p_4)\Big] \\ &= i\,\beta_2\,\frac{s(s-\tilde t-\tilde u)}{\Lambda^4} = 2i\,\beta_2\,\frac{s^2}{\Lambda^4}\,, \end{aligned} \quad \text{(A.40)}$$

inflation, so that the forward scattering limit is well-defined in spite of π being massless. To see this, first note that any internal contraction involving time derivative operators will vanish in the t-channel, and moreover those involving box operators will bring up factors of t, cancelling with the poles in the denominator. The remaining contractions then involve terms of the form $\partial_{\mu_1\cdots\mu_n}\pi\,\partial_{\nu_1\cdots\nu_m}\pi$. However, since these indices must be contracted with external legs, they will again induce factors of t, either cancelling within themselves due to antisymmetry to give zero contribution (as in our case) or with the poles to yield non-zero but pole-free amplitudes. Similar arguments hold for the absence of s- and u-channel poles.

where we used the relation $s + \tilde{t} + \tilde{u} = 0$ to represent the result in terms of s only.

- $\boldsymbol{(\partial\pi)^4}$. The symmetry factor for this interaction is again 8 for each channel, giving

$$\begin{aligned} i\mathcal{M}_{(\partial\pi)^4} &= i \cdot 8 \cdot \frac{\beta_3}{\Lambda^4}\Big[(\tilde{p}_1 \cdot \tilde{p}_2)(\tilde{p}_3 \cdot \tilde{p}_4) + (\tilde{p}_1 \cdot \tilde{p}_3)(\tilde{p}_2 \cdot \tilde{p}_4) + (\tilde{p}_1 \cdot \tilde{p}_4)(\tilde{p}_2 \cdot \tilde{p}_3)\Big] \\ &= 2i\,\beta_3\,\frac{(s^2 + \tilde{t}^2 + \tilde{u}^2)}{\Lambda^4}\,. \end{aligned} \tag{A.41}$$

Notice that this is the only amplitude with a non-trivial angular dependence.

Total amplitude. Adding the above results, gives the total amplitude

$$\mathcal{M}(s,\tilde{t}) \;=\; \left(-\frac{9}{4}\alpha_1^2 - 4\alpha_2^2 - 6\alpha_1\alpha_2 + \frac{3}{2}\beta_1 + 2\beta_2\right)\frac{s^2}{\Lambda^4} + 2\beta_3\,\frac{(s^2 + \tilde{t}^2 + \tilde{u}^2)}{\Lambda^4}\,. \tag{A.42}$$

In the forward limit, $\tilde{t} \to 0$, we find

$$\mathcal{A}(s) \;=\; \left(-\frac{9}{4}\alpha_1^2 - 4\alpha_2^2 - 6\alpha_1\alpha_2 + \frac{3}{2}\beta_1 + 2\beta_2 + 4\beta_3\right)\frac{s^2}{\Lambda^4}\,. \tag{A.43}$$

B. Particles in de Sitter space

In this appendix, we will derive various mathematical results that have been used in Chapter 5. In §B.1, we obtain the mode functions for massive spinning fields in de Sitter space by solving their equations of motion. We then derive the formula for their two-point function in §B.2.

Preliminaries. We will work with the components of the spinning field $\sigma_{\mu_1\cdots\mu_s}$ projected onto spatial slices, i.e. $\sigma_{i_1\cdots i_n\eta\cdots\eta}$. We will find it convenient to write these as

$$\sigma_{i_1\cdots i_n\eta\cdots\eta} = \sum_\lambda \sigma^\lambda_{n,s}\varepsilon^\lambda_{i_1\cdots i_n}\,, \tag{B.1}$$

where $\varepsilon^\lambda_{i_1\cdots i_n}$ is a suitably normalized polarization tensor (see insert below). The sub/superscripts on the mode functions $\sigma^\lambda_{n,s}$ label three "quantum numbers": s is the spin (or the rank) of the spacetime tensor field, n is its "spatial" spin, and λ is the helicity component of the spatial spin.

Polarization tensors.—In this insert, we will derive explicit expressions for the polarization tensors of arbitrary spin and helicity. The longitudinal polarization tensors are functions of $\hat{\mathbf{k}}$, while the transverse polarization tensors in addition depend on two polarization directions $\hat{\boldsymbol{\varepsilon}}^\pm$, with $\hat{\mathbf{k}}\cdot\hat{\boldsymbol{\varepsilon}}^\pm = 0$. Since $\hat{\boldsymbol{\varepsilon}}^+$ and $\hat{\boldsymbol{\varepsilon}}^-$ are related to each other by the reality condition $\hat{\boldsymbol{\varepsilon}}^+ = (\hat{\boldsymbol{\varepsilon}}^-)^*$, let us denote one of them by $\hat{\boldsymbol{\varepsilon}}$. The polarization tensors of helicity λ satisfy the following conditions:

i) symmetric: $\varepsilon^\lambda_{i_1\cdots i_s} = \varepsilon^\lambda_{(i_1\cdots i_s)}$.

ii) traceless: $\varepsilon^\lambda_{iii_3\cdots i_s} = 0$.

iii) transverse: $\hat{k}_{i_1}\cdots\hat{k}_{i_n}\varepsilon^\lambda_{i_1\cdots i_s} = 0$, when $n > s-\lambda$.

The last condition implies that the polarization tensor is of the form

$$\varepsilon^\lambda_{i_1\cdots i_s}(\hat{\mathbf{k}},\hat{\boldsymbol{\varepsilon}}) = \varepsilon^\lambda_{(i_1\cdots i_\lambda}(\hat{\boldsymbol{\varepsilon}})\, f_{i_{\lambda+1}\cdots i_s)}(\hat{\mathbf{k}})\,, \tag{B.2}$$

where $\hat{k}_{i_1}\varepsilon^\lambda_{i_1\cdots i_\lambda}(\hat{\boldsymbol{\varepsilon}}) = 0$ and $f_{i_1\cdots i_{s-\lambda}}$ is some tensor. Let us contract with vectors $\mathbf{q}$ and define

$$F^\lambda_s(x,y,z) \equiv q_{i_1}\cdots q_{i_s}\varepsilon^\lambda_{i_1\cdots i_s}(\hat{\mathbf{k}},\hat{\boldsymbol{\varepsilon}})\,, \tag{B.3}$$

where we have defined $x \equiv q^2$, $y \equiv \mathbf{q}\cdot\hat{\mathbf{k}}$, and $z \equiv q_{i_1}\cdots q_{i_\lambda}\varepsilon^\lambda_{i_1\cdots i_\lambda}$. The function F^λ_s is a homogeneous polynomial in $\mathbf{q}$, so that

$$2x\,F^\lambda_{s,x} + y\,F^\lambda_{s,y} + \lambda z\,F^\lambda_{s,z} = s\,F^\lambda_s\,. \tag{B.4}$$

The transverse and traceless conditions translate into

$$z\,F^{\lambda}_{s,z} = F^{\lambda}_{s}\,, \tag{B.5}$$

$$4x\,F^{\lambda}_{s,xx} + 4y\,F^{\lambda}_{s,xy} + 4\lambda z\,F^{\lambda}_{s,xz} + 2d\,F^{\lambda}_{s,x} + F^{\lambda}_{s,yy} = 0\,, \tag{B.6}$$

where d is the number of spatial dimensions. Taking derivatives of (B.4) and (B.5), and substituting into (B.6), we get

$$(x-y^2)F^{\lambda}_{s,yy} - (2\lambda+d-1)y\,F^{\lambda}_{s,y} + (s-\lambda)(s+\lambda+d-2)F^{\lambda}_{s} = 0\,. \tag{B.7}$$

Without loss of generality, we now set $x = q^2 \equiv 1$. The solution to (B.5) and (B.7) is

$$F^{\lambda}_{s}(y,z) \propto z\hat{P}^{\beta_\lambda}_{\beta_s}(y)\,, \tag{B.8}$$

where $\hat{P}^{\beta_\lambda}_{\beta_s}$ is part of the associated Legendre polynomial $P^{\beta_\lambda}_{\beta_s}$ of degree $\beta_s \equiv \frac{1}{2}(2s+d-3)$ and order $\beta_\lambda \equiv \frac{1}{2}(2\lambda+d-3)$, defined by $P^{\beta_\lambda}_{\beta_s}(y) = (1-y^2)^{\beta_\lambda/2}\hat{P}^{\beta_\lambda}_{\beta_s}$. We will set $P^{\beta_\lambda}_{s} \equiv P^{|\beta_\lambda|}_{s}$ and distinguish the opposite helicities only by the phase. For $d=3$, this reduces to

$$F^{\lambda}_{s}(y,z) \propto z\hat{P}^{\lambda}_{s}(y)\,. \tag{B.9}$$

This result includes longitudinal polarization tensors for $\lambda = 0$ and $z = 1$. It is straightforward to obtain explicit expressions for the polarization tensors by stripping off the contractions with $\mathbf{q}$ in (B.9) and symmetrizing the indices:

$$\varepsilon^{\lambda}_{i_1\cdots i_s}(\hat{\mathbf{k}},\hat{\boldsymbol{\varepsilon}}) = \frac{1}{(2\lambda-1)!!}\sum_{n=0}^{s-\lambda} B_n\,\varepsilon^{\lambda}_{(i_1\cdots i_\lambda}(\hat{\boldsymbol{\varepsilon}})\,\hat{k}_{i_{\lambda+1}}\cdots\hat{k}_{i_{\lambda+n}}\delta_{i_{\lambda+n+1}\cdots i_s)}\,, \tag{B.10}$$

where

$$B_n \equiv \frac{2^s}{n!(s-n-\lambda)!}\frac{\Gamma[\frac{1}{2}(n+\lambda+1+s)]}{\Gamma[\frac{1}{2}(n+\lambda+1-s)]}\,, \quad \delta_{i_1\cdots i_n} \equiv \begin{cases} \delta_{i_1 i_2}\cdots\delta_{i_{n-1}i_n} & n \text{ even} \\ 0 & n \text{ odd} \end{cases}\,. \tag{B.11}$$

The self-contraction of the polarization tensors can be written as

$$\varepsilon^{\lambda}_{i_1\cdots i_s}\varepsilon^{\lambda *}_{i_1\cdots i_s} = \frac{(2s-1)!!(s+\lambda)!}{2^\lambda[(2\lambda-1)!!]^2 s!(s-\lambda)!}\varepsilon^{\lambda}_{i_1\cdots i_\lambda}\varepsilon^{\lambda *}_{i_1\cdots i_\lambda}\,. \tag{B.12}$$

When choosing the orthogonal direction to be in, say, the z-direction, there will be in total of 2^s non-zero components for the polarized tensor $\varepsilon^{s}_{i_1\cdots i_s}$, which are ± 1 or $\pm i$ up to a phase. This means that $\varepsilon^{s}_{i_1\cdots i_s}\varepsilon^{s *}_{i_1\cdots i_s} = 2^s$ with some overall normalization which we set to unity for convenience.

B.1 Mode functions

In this section, we will derive the de Sitter mode functions for fields with spin. We will explicitly derive the mode functions for fields with spins 1 and 2, and present the results for arbitrary spin at the end.

Spin-1. The equation of motion of a massive spin-1 field σ_μ is

$$(\Box - m_1^2)\sigma_\mu = 0\,, \tag{B.13}$$

with $\nabla^\mu\sigma_\mu = 0$ and $m_1^2 = m^2+3H^2$. The equation of motion expressed in terms of the components σ_η and σ_i are

$$\sigma_\eta'' - \left(\partial_j^2 - \frac{m^2/H^2-2}{\eta^2}\right)\sigma_\eta = \frac{2}{\eta}\partial_i\sigma_i\,, \tag{B.14}$$

$$\sigma_i'' - \left(\partial_j^2 - \frac{m^2/H^2}{\eta^2}\right)\sigma_i = \frac{2}{\eta}\partial_i\sigma_\eta\,, \tag{B.15}$$

where a prime denotes a derivative with respect to conformal time, and the transverse condition gives

$$\sigma_\eta' - \frac{2}{\eta}\sigma_\eta = \partial_i\sigma_i\,. \tag{B.16}$$

In order to decouple the equations (B.14) and (B.15), we expand the field σ_μ into its different helicity components,

$$\sigma_\mu = \sum_{\lambda=-1}^{1}\sigma_\mu^{(\lambda)}\,, \tag{B.17}$$

where

$$\sigma_\eta^{(0)} = \sigma_{0,1}^0\,, \qquad \sigma_\eta^{(\pm1)} = 0\,, \tag{B.18}$$

$$\sigma_i^{(0)} = \sigma_{1,1}^0\varepsilon_i^0\,, \qquad \sigma_i^{(\pm1)} = \sigma_{1,1}^{\pm1}\varepsilon_i^{\pm1}\,. \tag{B.19}$$

We demand that the polarization vectors $\varepsilon_i^\lambda(\hat{\mathbf{k}})$ satisfy

$$\hat{k}_i\varepsilon_i^0 = 1\,, \quad \hat{k}_i\varepsilon_i^{\pm1} = 0\,, \quad \varepsilon_i^{\pm1} = \varepsilon_i^{\mp1*}\,, \quad \varepsilon_i^{\pm1}\varepsilon_i^{\pm1*} = 2\,. \tag{B.20}$$

The choice of the normalization (B.20) uniquely fixes the longitudinal polarization vector to be $\varepsilon_i^0(\hat{\mathbf{k}}) = \hat{k}_i$, and the transverse polarization vectors are fixed up to a phase. For momentum along the z-direction, they can be chosen to be $\varepsilon_i^{\pm1}(\hat{\mathbf{z}}) = (1, \pm i, 0)$.

In terms of the mode functions defined in (B.18) and (B.19), the equations (B.14) and (B.15) decouple

$${\sigma_{0,1}^0}'' - \frac{2}{\eta}{\sigma_{0,1}^0}' + \left(k^2 + \frac{m^2/H^2+2}{\eta^2}\right)\sigma_{0,1}^0 = 0\,, \tag{B.21}$$

$${\sigma_{1,1}^0}'' - \frac{k^2\eta^2}{k^2\eta^2+m^2/H^2}\frac{2}{\eta}{\sigma_{1,1}^0}' + \left(k^2 + \frac{m^2/H^2}{\eta^2}\right)\sigma_{1,1}^0 = 0\,, \tag{B.22}$$

$$\sigma_{1,1}^{\pm1\prime\prime} + \left(k^2 + \frac{m^2/H^2}{\eta^2}\right)\sigma_{1,1}^{\pm1} = 0\,, \tag{B.23}$$

and the transverse condition (B.16) becomes

$$\sigma^0_{1,1} = -\frac{i}{k}\left(\sigma^{0\,\prime}_{0,1} - \frac{2}{\eta}\sigma^0_{0,1}\right). \tag{B.24}$$

The solutions to these equations with the Bunch-Davies initial condition are

$$\sigma^0_{0,1} = \mathcal{A}_1 N_1 (-k\eta)^{3/2} H_{i\mu_1}\,, \tag{B.25}$$

$$\sigma^0_{1,1} = \frac{i}{2}\mathcal{A}_1 N_1 (-k\eta)^{1/2}\Big[k\eta\big(H_{i\mu_1+1} - H_{i\mu_1-1}\big) - H_{i\mu_1}\Big]\,, \tag{B.26}$$

$$\sigma^{\pm1}_{1,1} = \mathcal{A}_1 Z_1^{\pm1} (-k\eta)^{1/2} H_{i\mu_1}\,, \tag{B.27}$$

where $\mathcal{A}_1 \equiv e^{i\pi/4}e^{-\pi\mu_1/2}$ and $Z_1^{\pm1}$ denotes the normalization constant for the helicity-±1 mode of the spin-1 field. We have also suppressed the argument $-k\eta$ of the Hankel functions $H_{i\mu_1} \equiv H^{(1)}_{i\mu_1}$ for brevity.

A few comments are in order. First, note that for $m = 0$ equation (B.23) for the transverse mode becomes the flat space wave equation, whose solutions are simply plane waves. This is because the action of a massless spin-1 field is conformally invariant, so the mode in de Sitter space behaves as if it were in flat space. On the other hand, we do not see this behavior for the longitudinal mode. In particular, the longitudinal mode blows up relative to the transverse mode as we go to the infinite past $\eta \to -\infty$. We can understand this as follows. The mass term $m^2/H^2\eta^2$ in the action (5.1) is time dependent, so the spin-1 field is effectively massless in the infinite past, in which case the longitudinal mode turns into a pure gauge mode.

We still need to determine the normalization constants N_1 and $Z_1^{\pm1}$. This is done by imposing orthonormality of mode functions under the inner product

$$\left\langle \sigma^{(\lambda)}_\mu(\mathbf{k},\eta)e^{i\mathbf{k}\cdot\mathbf{x}}, \sigma^{(\lambda')}_\nu(\mathbf{k}',\eta)e^{i\mathbf{k}'\cdot\mathbf{x}}\right\rangle = \delta_{\lambda\lambda'}\delta(\mathbf{k}-\mathbf{k}')\,. \tag{B.28}$$

This orthonormality condition guarantees that we get the standard equal-time commutation relation upon canonical quantization. We have

$$\begin{aligned}\left\langle \sigma^{(0)}_\mu(\mathbf{k},\eta)e^{i\mathbf{k}\cdot\mathbf{x}}, \sigma^{(0)}_\nu(\mathbf{k}',\eta)e^{i\mathbf{k}'\cdot\mathbf{x}}\right\rangle &= -i\eta^{\mu\nu}\int \mathrm{d}^3x\left[\sigma^{(0)}_\mu\sigma^{(0)*\prime}_\nu - \sigma^{(0)\prime}_\mu\sigma^{(0)*}_\nu\right]e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} \\ &= -i\left[-\mathcal{W}(\sigma^0_{0,1},\sigma^{0*}_{0,1}) + \mathcal{W}(\sigma^0_{1,1},\sigma^{0*}_{1,1})\right]\delta(\mathbf{k}-\mathbf{k}')\,,\end{aligned} \tag{B.29}$$

where $\mathcal{W}$ denotes the Wronskian. Substituting (B.25) and (B.27), we obtain

$$\mathcal{W}(\sigma^0_{0,1},\sigma^{0*}_{0,1}) = \frac{4ik^3\eta^2}{\pi} \times N_1^2\,, \tag{B.30}$$

$$\mathcal{W}(\sigma^0_{1,1},\sigma^{0*}_{1,1}) = \frac{4ik(k^2\eta^2 + 1/4 + \mu_1^2)}{\pi} \times N_1^2\,. \tag{B.31}$$

Note that the time dependences in (B.30) and (B.31) cancel in (B.29). Imposing (B.28), we then

get

$$N_1 = \sqrt{\frac{\pi}{2}}\frac{1}{\sqrt{2k}}\frac{1}{(1/4+\mu_1^2)^{1/2}} = \sqrt{\frac{\pi}{2}}\frac{1}{\sqrt{2k}}\frac{H}{m}\,. \tag{B.32}$$

The normalization for the transverse mode can be determined in a similar way. We get

$$Z_1^{\pm 1} = \sqrt{\frac{\pi}{2}}\frac{1}{\sqrt{2k}}\,. \tag{B.33}$$

Notice that the normalization for the longitudinal mode blows up when $m = 0$, which, again, does not signal any pathologies, since the longitudinal mode becomes a pure gauge mode in this limit.

Spin-2. The equations of motion and the constraints satisfied by a massive spin-2 field $\sigma_{\mu\nu}$ are

$$(\Box - m^2 - 2H^2)\sigma_{\mu\nu} = 0\ , \quad \nabla^\mu\sigma_{\mu\nu} = 0\ , \quad \tilde{\sigma} \equiv \sigma^\mu{}_\mu = 0\ . \tag{B.34}$$

In terms of components, these are

$$\sigma''_{\eta\eta} + \frac{2}{\eta}\sigma'_{\eta\eta} - \left(\partial_k^2 - \frac{m^2/H^2-6}{\eta^2}\right)\sigma_{\eta\eta} = \frac{4}{\eta}\partial_i\sigma_{i\eta} + \frac{2}{\eta^2}\sigma_{ii}\,, \tag{B.35}$$

$$\sigma''_{i\eta} + \frac{2}{\eta}\sigma'_{i\eta} - \left(\partial_k^2 - \frac{m^2/H^2-6}{\eta^2}\right)\sigma_{i\eta} = \frac{2}{\eta}\partial_i\sigma_{\eta\eta} + \frac{2}{\eta}\partial_j\sigma_{ij}\,, \tag{B.36}$$

$$\sigma''_{ij} + \frac{2}{\eta}\sigma'_{ij} - \left(\partial_k^2 - \frac{m^2/H^2-2}{\eta^2}\right)\sigma_{ij} = \frac{4}{\eta}\partial_{(i}\sigma_{j)\eta} + \frac{2}{\eta^2}\sigma_{\eta\eta}\delta_{ij}\,, \tag{B.37}$$

and

$$\sigma'_{\eta\eta} - \partial_i\sigma_{i\eta} - \frac{1}{\eta}\sigma_{\eta\eta} - \frac{1}{\eta}\sigma_{ii} = 0\,, \tag{B.38}$$

$$\sigma'_{i\eta} - \partial_j\sigma_{ij} - \frac{2}{\eta}\sigma_{i\eta} = 0\,, \tag{B.39}$$

$$\sigma_{\eta\eta} - \sigma_{ii} = 0\,. \tag{B.40}$$

As before, we expand the Fourier modes into helicity eigenstates

$$\sigma_{\mu\nu} = \sum_{\lambda=-2}^{2}\sigma_{\mu\nu}^{(\lambda)}\,. \tag{B.41}$$

Let us denote the traceless part of the spatial tensor by $\hat{\sigma}_{ij}$, so that $\sigma_{ij} = \hat{\sigma}_{ij} + \frac{1}{3}\sigma_{\eta\eta}\delta_{ij}$, and decompose the mode functions into different helicities:

$$\sigma_{\eta\eta}^{(0)} = \sigma_{0,2}^0\,, \qquad \sigma_{\eta\eta}^{(\pm1)} = 0\,, \qquad \sigma_{\eta\eta}^{(\pm2)} = 0\,, \tag{B.42}$$

$$\sigma_{i\eta}^{(0)} = \sigma_{1,2}^0\varepsilon_i^0\,, \quad \sigma_{i\eta}^{(\pm1)} = \sigma_{1,2}^{\pm1}\varepsilon_i^{\pm1}\,, \quad \sigma_{i\eta}^{(\pm2)} = 0\,, \tag{B.43}$$

$$\hat{\sigma}_{ij}^{(0)} = \sigma_{2,2}^0\varepsilon_{ij}^0\ , \quad \hat{\sigma}_{ij}^{(\pm1)} = \sigma_{2,2}^{\pm1}\varepsilon_{ij}^{\pm1}\,, \quad \hat{\sigma}_{ij}^{(\pm2)} = \sigma_{2,2}^{\pm2}\varepsilon_{ij}^{\pm2}\,. \tag{B.44}$$

Demanding that the polarization tensors satisfy

$$\hat{k}_i \varepsilon^0_{ij} = \varepsilon^0_j\,, \quad \hat{k}_i \varepsilon^{\pm 1}_{ij} = \frac{3}{2}\varepsilon^{\pm 1}_j\,, \quad k_i \varepsilon^{\pm 2}_{ij} = 0\,, \quad \varepsilon^{\pm 2}_{ij} = \varepsilon^{\mp 2 *}_{ij}\,, \quad \varepsilon^{\pm 2}_{ij}\varepsilon^{\pm 2 *}_{ij} = 4\,, \tag{B.45}$$

leads to

$$\varepsilon^0_{ij} = \frac{3}{2}\left(\hat{k}_i\hat{k}_j - \frac{1}{3}\delta_{ij}\right), \quad \varepsilon^{\pm 1}_{ij} = \frac{3}{2}\left(\hat{k}_i\varepsilon^{\pm 1}_j + \hat{k}_j\varepsilon^{\pm 1}_i\right), \tag{B.46}$$

and fixes $\varepsilon^{\pm 2}_{ij}$ up to a phase. For $\hat{\mathbf{k}}$ along the z-direction, this can be chosen to be

$$\varepsilon^{\pm 2}_{ij}(\hat{\mathbf{z}}) = \begin{pmatrix} 1 & \pm i & 0 \\ \pm i & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{B.47}$$

The equations satisfied by the different helicity modes are

$$\sigma^{0}_{0,2}{}'' - \frac{2}{\eta}\sigma^{0}_{0,2}{}' + \left(k^2 + \frac{m^2/H^2}{\eta^2}\right)\sigma^0_{0,2} = 0\,, \tag{B.48}$$

$$\sigma^{\pm 1}_{1,2}{}'' + \left(k^2 + \frac{m^2/H^2 - 2}{\eta^2}\right)\sigma^{\pm 1}_{1,2} = 0\,, \tag{B.49}$$

$$\sigma^{\pm 2}_{2,2}{}'' + \frac{2}{\eta}\sigma^{\pm 2}_{2,2}{}' + \left(k^2 + \frac{m^2/H^2 - 2}{\eta^2}\right)\sigma^{\pm 2}_{2,2} = 0\,, \tag{B.50}$$

subject to the transverse conditions

$$\sigma^0_{1,2} = -\frac{i}{k}\left(\sigma^{0}_{0,2}{}' - \frac{2}{\eta}\sigma^0_{0,2}\right), \quad \sigma^0_{2,2} = -\frac{i}{k}\left(\sigma^{0}_{1,2}{}' - \frac{2}{\eta}\sigma^0_{1,2}\right) - \frac{1}{3}\sigma^0_{0,2}\,, \tag{B.51}$$

$$\sigma^{\pm 1}_{2,2} = -\frac{i}{k}\left(\sigma^{\pm 1\prime}_{1,2} - \frac{2}{\eta}\sigma^{\pm 1}_{1,2}\right). \tag{B.52}$$

The solutions with Bunch-Davies initial conditions are

$$\sigma^0_{0,2} = \mathcal{A}_2\, N_2 (-k\eta)^{3/2} H_{i\mu_2}\,, \tag{B.53}$$

$$\sigma^0_{1,2} = \frac{i}{2}\mathcal{A}_2\, N_2 (-k\eta)^{1/2}\Big[k\eta\big(H_{i\mu_2+1} - H_{i\mu_2-1}\big) - H_{i\mu_2}\Big]\,, \tag{B.54}$$

$$\sigma^0_{2,2} = \frac{1}{12}\mathcal{A}_2\, N_2 (-k\eta)^{-1/2}\Big[6k\eta\big((2+i\mu_2)H_{i\mu_2+1} - (2-i\mu)H_{i\mu_2-1}\big) - (9 - 8k^2\eta^2)H_{i\mu_2}\Big]\,, \tag{B.55}$$

for the longitudinal modes, and

$$\sigma^{\pm 1}_{1,2} = \mathcal{A}_2\, Z_2^{\pm 1} (-k\eta)^{1/2} H_{i\mu_2}\,, \tag{B.56}$$

$$\sigma^{\pm 1}_{2,2} = \frac{i}{2}\mathcal{A}_2\, Z_2^{\pm 1} (-k\eta)^{-1/2}\Big[k\eta\big(H_{i\mu_2+1} - H_{i\mu_2-1}\big) - 3H_{i\mu_2}\Big]\,, \tag{B.57}$$

$$\sigma^{\pm 2}_{2,2} = \mathcal{A}_2\, Z_2^{\pm 2} (-k\eta)^{-1/2} H_{i\mu_2}\,, \tag{B.58}$$

for the higher-helicity modes.

To fix the normalization, we again impose orthonormality of the mode functions

$$\left\langle \sigma^{(\lambda)}_{\mu\alpha}(\mathbf{k},\eta)e^{i\mathbf{k}\cdot\mathbf{x}}, \sigma^{(\lambda')}_{\nu\beta}(\mathbf{k}',\eta)e^{i\mathbf{k}'\cdot\mathbf{x}} \right\rangle = \delta_{\lambda\lambda'}\delta(\mathbf{k}-\mathbf{k}')\,. \tag{B.59}$$

We have

$$\begin{aligned}\left\langle \sigma^{(0)}_{\mu\alpha}(\mathbf{k},\eta)e^{i\mathbf{k}\cdot\mathbf{x}}, \sigma^{(0)}_{\nu\beta}(\mathbf{k}',\eta)e^{i\mathbf{k}'\cdot\mathbf{x}} \right\rangle &= -\frac{i}{a^2}\eta^{\mu\nu}\eta^{\alpha\beta}\int \mathrm{d}^3x\, \left[\sigma^{(0)}_{\mu\alpha}\sigma^{(0)*\prime}_{\nu\beta} - \sigma^{(0)\prime}_{\mu\alpha}\sigma^{(0)*}_{\nu\beta}\right] e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} \\ &= -\frac{i}{a^2}\left[\frac{4}{3}\mathcal{W}(\sigma^0_{0,2},\sigma^{0*}_{0,2}) - 2\mathcal{W}(\sigma^0_{1,2},\sigma^{0*}_{1,2}) + \frac{3}{2}\mathcal{W}(\sigma^0_{2,2},\sigma^{0*}_{2,2})\right]\delta(\mathbf{k}-\mathbf{k}')\,,\end{aligned} \tag{B.60}$$

where

$$\mathcal{W}(\sigma^0_{0,2},\sigma^{0*}_{0,2}) = \frac{4ik^3\eta^2}{\pi}\times N_2^2\,, \tag{B.61}$$

$$\mathcal{W}(\sigma^0_{1,2},\sigma^{0*}_{1,2}) = \frac{4ik(k^2\eta^2+1/4+\mu_2^2)}{\pi}\times N_2^2\,, \tag{B.62}$$

$$\mathcal{W}(\sigma^0_{2,2},\sigma^{0*}_{2,2}) = \frac{i[32k^4\eta^4+96k^2\eta^2(1/4+\mu_2^2)+72(1/4+\mu_2^2)(9/4+\mu_2^2)]}{18\pi k\eta^2}\times N_2^2\,. \tag{B.63}$$

The condition (B.60) then sets the normalization constant to be

$$N_2 = \sqrt{\frac{\pi}{3}}\frac{1}{\sqrt{2k}}\frac{k}{H}\frac{1}{\left[(1/4+\mu_2^2)(9/4+\mu_2^2)\right]^{1/2}}\,. \tag{B.64}$$

We see that this diverges at $m^2=0$ and $m^2=2H^2$. This is again to be expected. For $m=0$, the action gains gauge invariance, in which case only the helicity-±2 modes are physical. For $m^2=2H^2$, the field becomes partially massless, and the number of propagating degrees of freedom becomes four. In both cases, the longitudinal mode becomes a pure gauge mode. Finally, determining the normalizations of the transverse modes in an analogous way, we get

$$Z_2^{\pm1} = \frac{\sqrt{\pi}}{3}\frac{1}{\sqrt{2k}}\frac{k}{H}\frac{1}{(9/4+\mu_2^2)^{1/2}}\,,\quad Z_2^{\pm2} = \sqrt{\frac{\pi}{2}}\frac{1}{\sqrt{2k}}\frac{k}{H}\,. \tag{B.65}$$

In the massless limit, $Z_2^{\pm1}$ diverges and only $Z_2^{\pm2}$ remains finite.

Spin-s. For spins higher than two, we need to solve the on-shell equations (5.7). In order to decouple these equations, we expand the field $\sigma_{\mu_1\cdots\mu_s}$ into its different helicity components

$$\sigma_{\mu_1\cdots\mu_s} = \sum_{\lambda=-s}^{s}\sigma^{(\lambda)}_{\mu_1\cdots\mu_s}\,. \tag{B.66}$$

A mode of helicity λ and n polarization directions can be written as

$$\sigma^{(\lambda)}_{i_1\cdots i_n\eta\cdots\eta} = \sigma^{\lambda}_{n,s}\varepsilon^{\lambda}_{i_1\cdots i_n}\,, \tag{B.67}$$

where $\sigma^{\lambda}_{n,s} = 0$ for $n < |\lambda|$. The helicity-λ mode function with $n = |\lambda|$ number of polarization directions satisfies

$$\sigma^{\lambda}_{|\lambda|,s}{}'' - \frac{2(1-\lambda)}{\eta}\sigma^{\lambda}_{|\lambda|,s}{}' + \left(k^2 + \frac{m^2/H^2 - (s+\lambda-2)(s-\lambda+1)}{\eta^2}\right)\sigma^{\lambda}_{|\lambda|,s} = 0\,, \tag{B.68}$$

whose solution is given by

$$\sigma^{\lambda}_{|\lambda|,s} = \mathcal{A}_s\, Z^{\lambda}_s (-k\eta)^{3/2-\lambda} H_{i\mu_s}\,. \tag{B.69}$$

The other mode functions can then be obtained iteratively using the following recursion relation:

$$\sigma^{\lambda}_{n+1,s} = -\frac{i}{k}\left(\sigma^{\lambda}_{n,s}{}' - \frac{2}{\eta}\sigma^{\lambda}_{n,s}\right) - \sum_{m=|\lambda|}^{n} B_{m,n+1}\,\sigma^{\lambda}_{m,s}\,, \tag{B.70}$$

where

$$B_{m,n} \equiv \frac{2^n\, n!}{m!(n-m)!(2n-1)!!}\frac{\Gamma[\frac{1}{2}(1+m+n)]}{\Gamma[\frac{1}{2}(1+m-n)]}\,. \tag{B.71}$$

Having obtained the formula that enables us to compute the mode functions of arbitrary spin and helicity, let us now fix their normalization constants. In order to do so, we first define an inner product between two mode functions. Note that if $f_{\mu_1\cdots\mu_s}$ and $h_{\nu_1\cdots\mu_s}$ are two solutions to (5.7), then the current

$$J_\mu \equiv f^{\nu_1\cdots\nu_s}\nabla_\mu h^*_{\nu_1\cdots\nu_s} - h^*_{\nu_1\cdots\nu_s}\nabla_\mu f^{\nu_1\cdots\nu_s}\,, \tag{B.72}$$

is conserved, $\nabla^\mu J_\mu = 0$. This means that we can define an inner product of two solutions

$$\langle f_{\mu_1\cdots\mu_s}, h_{\nu_1\cdots\nu_s}\rangle \equiv -ig^{\mu_1\nu_1}\cdots g^{\mu_s\nu_s}\int \mathrm{d}\Sigma\, n^\lambda\sqrt{\hat{g}}\left[f_{\mu_1\cdots\mu_s}\nabla_\lambda h^*_{\nu_1\cdots\nu_s} - h^*_{\nu_1\cdots\nu_s}\nabla_\lambda f_{\mu_1\cdots\mu_s}\right], \tag{B.73}$$

where Σ denotes a spacelike hypersurface, $\hat{g}$ is the determinant of the spatial metric, and n^μ is the timelike unit vector orthogonal to Σ. The conservation of the current (B.72) implies that the inner product is time independent. For the FRW metric, the above inner product reduces to

$$\langle f_{\mu_1\cdots\mu_s}, h_{\nu_1\cdots\nu_s}\rangle = -\frac{i}{a^{2(s-1)}}\eta^{\mu_1\nu_1}\cdots\eta^{\mu_s\nu_s}\int \mathrm{d}^3x\left[f_{\mu_1\cdots\mu_s}h^{*\prime}_{\nu_1\cdots\nu_s} - f'_{\mu_1\cdots\mu_s}h^*_{\nu_1\cdots\nu_s}\right]. \tag{B.74}$$

The normalization in (B.69) is then determined by imposing orthonormality under the inner product (B.74):

$$\left\langle \sigma^{(\lambda)}_{\mu_1\cdots\mu_s}(\mathbf{k},\eta)e^{i\mathbf{k}\cdot\mathbf{x}}, \sigma^{(\lambda')}_{\nu_1\cdots\nu_s}(\mathbf{k}',\eta)e^{i\mathbf{k}'\cdot\mathbf{x}}\right\rangle = \delta_{\lambda\lambda'}\delta(\mathbf{k}-\mathbf{k}')\,. \tag{B.75}$$

Since the inner product is time independent, it does not matter which time slice we choose to evaluate the integral on. We will therefore evaluate the integral on the future boundary by taking

the limit $\eta \to 0$. From (B.70), we note that $\sigma^{\lambda}_{n_1,s}$ is subleading compared to $\sigma^{\lambda}_{n_2,s}$ in the limit $\eta \to 0$ for all $n_1 < n_2$, so we simply need to compute the Wronskian of the mode with the highest number of polarization directions, $\sigma^{\lambda}_{s,s}$. If we had kept all the Wronskians, then the subleading time-dependent terms would cancel. Note also that the trace terms in (B.70) become subleading in the limit $\eta \to 0$, so we will drop these terms. Since (B.74) is a constant, the leading term in the Wronskian must scale as $\eta^{2(1-s)}$ to cancel off the factor $a^{2(1-s)}$. In the insert below, we will show that the orthonormality condition fixes the normalization constant to be

$$(Z^{\lambda}_s)^2 = \frac{1}{k}\left(\frac{k}{H}\right)^{2s-2}(\mathcal{Z}^{\lambda}_s)^2\,, \tag{B.76}$$

$$(\mathcal{Z}^{\lambda}_s)^2 \equiv \frac{\pi}{4}\frac{[(2\lambda-1)!!]^2 s!(s-\lambda)!}{(2s-1)!!(s+\lambda)!}\frac{\Gamma(\frac{1}{2}+\lambda+i\mu_s)\Gamma(\frac{1}{2}+\lambda-i\mu_s)}{\Gamma(\frac{1}{2}+s+i\mu_s)\Gamma(\frac{1}{2}+s-i\mu_s)}\,. \tag{B.77}$$

Note that the normalization constant has poles at $\mu_s^2 = \{-(n+\frac{1}{2})^2\}^s_{n=\lambda}$, at which the spinning field becomes (partially) massless and some of the helicity modes become unphysical. For convenience, we will denote the normalization of the longitudinal mode by $N_s \equiv Z^0_s$ ($\mathcal{N}_s \equiv \mathcal{Z}^0_s$).

Derivation of (B.76).—First, note that the n-th mode function can be cast in the form

$$\sigma^{\lambda}_{n,s} = \mathcal{A}_s Z^{\lambda}_s(-k\eta)^{3/2-n}\Big[(x_n+iy_n)H_{i\mu_s}+(w_n+iz_n)k\eta H_{i\mu_s+1}\Big], \tag{B.78}$$

by use of the recursion relation $H_{i\mu_s+1}(x)+H_{i\mu_s-1}(x) = (2i\mu_s/x)H_{i\mu_s}(x)$. The coefficients x_n, y_n, w_n, and z_n can in general depend on time, but are constant in the limit $\eta \to 0$. The Wronskian is

$$\mathcal{W}\big[\sigma^{\lambda}_{n,s},\sigma^{\lambda *}_{n,s}\big] = \frac{4ik(Z^{\lambda}_s)^2}{\pi(k\eta)^{2(n-1)}}\Big[X_n - 2\mu_s Y_n(\coth\pi\mu_s - 1)\Big]\,, \tag{B.79}$$

where

$$X_n \equiv x_n^2+y_n^2\,, \quad Y_n \equiv x_n z_n - y_n w_n + (w_n^2+z_n^2)\mu_s\,. \tag{B.80}$$

Let us show that in fact $Y_n = 0$ for any n-th order mode function. We do this by induction. First, it is trivial to check that this is satisfied by the mode (B.69). Now, assume that $Y_n = 0$ is satisfied at some n-th order. Using the recursion relation (B.70), and taking the limit $\eta \to 0$, we get

$$\sigma^{\lambda}_{n+1,s} = \mathcal{A}(\mu_s)Z^{\lambda}_s(-k\eta)^{1/2-n}\Big[(x_{n+1}+iy_{n+1})H_{i\mu_s}+(w_{n+1}+iz_{n+1})k\eta H_{i\mu_s+1}\Big], \tag{B.81}$$

where

$$\begin{aligned} 2x_{n+1} &= -2\mu_s x_n + (2n+1)y_n\,, & 2y_{n+1} &= -(2n+1)x_n - 2\mu_s y_n\,,\\ 2w_{n+1} &= -2y_n + 2\mu_s w_n + (2n+1)z_n\,, & 2z_{n+1} &= 2x_n - (2n+1)w_n + 2\mu_s z_n\,. \end{aligned} \tag{B.82}$$

These coefficients then give

$$X_{n+1} = \big[(n+\tfrac{1}{2})^2+\mu_s^2\big]X_n\,, \quad Y_{n+1} = \big[(n+\tfrac{1}{2})^2+\mu_s^2\big]Y_n\,. \tag{B.83}$$

Hence, $Y_{n+1}=0$. Since n was arbitrary, we conclude that $Y_n=0$ for all n. Next, we show that the Wronskian of the n-th longitudinal mode function has the form

$$\mathcal{W}\big[\sigma^{\lambda}_{n,s},\sigma^{\lambda*}_{n,s}\big]=\frac{4ik(Z^{\lambda}_s)^2}{\pi(k\eta)^{2(n-1)}}\frac{\Gamma(\frac{1}{2}+n+i\mu_s)\Gamma(\frac{1}{2}+n-i\mu_s)}{\Gamma(\frac{1}{2}+\lambda+i\mu_s)\Gamma(\frac{1}{2}+\lambda-i\mu_s)}\,. \tag{B.84}$$

The Wronskian of the mode function (B.69) is

$$\mathcal{W}\big[\sigma^{\lambda}_{\lambda,s},\sigma^{\lambda*}_{\lambda,s}\big]=\frac{4ik(Z^{\lambda}_s)^2}{\pi(k\eta)^{2(\lambda-1)}}\,, \tag{B.85}$$

and hence satisfies (B.84). Assuming that (B.84) is true at n-th order and using (B.83), we get

$$\begin{aligned}\mathcal{W}\big[\sigma^{\lambda}_{n+1,s},\sigma^{\lambda*}_{n+1,s}\big]&=\frac{4ik(Z^{\lambda}_s)^2}{\pi(k\eta)^{2n}}X_{n+1}=\frac{\big[(n+\frac{1}{2})^2+\mu_s^2\big]}{(k\eta)^2}\frac{4ik(Z^{\lambda}_s)^2}{\pi(k\eta)^{2(n-1)}}X_n\\&=\frac{\big[(n+\frac{1}{2})^2+\mu_s^2\big]}{(k\eta)^2}\mathcal{W}\big[\sigma^{\lambda}_{n,s},\sigma^{\lambda*}_{n,s}\big]\\&=\frac{4ik(Z^{\lambda}_s)^2}{\pi(k\eta)^{2n}}\frac{\Gamma(\frac{3}{2}+n+i\mu_s)\Gamma(\frac{3}{2}+n-i\mu_s)}{\Gamma(\frac{1}{2}+\lambda+i\mu_s)\Gamma(\frac{1}{2}+\lambda-i\mu_s)}\,,\end{aligned} \tag{B.86}$$

where in the last line we have use the fact that

$$\frac{\Gamma(\frac{3}{2}+n+i\mu_s)\Gamma(\frac{3}{2}+n-i\mu_s)}{\Gamma(\frac{1}{2}+n+i\mu_s)\Gamma(\frac{1}{2}+n-i\mu_s)}=(n+\tfrac{1}{2})^2+\mu_s^2\,. \tag{B.87}$$

Thus, we have proven (B.84). Finally, the inner product (B.74) is given by

$$\begin{aligned}&\Big\langle\sigma^{(\lambda)}_{\mu_1\cdots\mu_s}(\mathbf{k},\eta)e^{i\mathbf{k}\cdot\mathbf{x}},\sigma^{(\lambda)}_{\nu_1\cdots\nu_s}(\mathbf{k}',\eta)e^{i\mathbf{k}'\cdot\mathbf{x}}\Big\rangle\\&\quad=-\frac{i}{a^{2(s-1)}}\eta^{\mu_1\nu_1}\cdots\eta^{\mu_s\nu_s}\int\mathrm{d}^3x\,\Big[\sigma^{(\lambda)}_{\mu_1\cdots\mu_s}\sigma^{(\lambda)*\prime}_{\nu_1\cdots\nu_s}-\sigma^{(\lambda)\prime}_{\mu_1\cdots\mu_s}\sigma^{(\lambda)*}_{\nu_1\cdots\nu_s}\Big]\,e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}}\\&\quad=-i(-H\eta)^{2(s-1)}\mathcal{W}\big[\sigma^{\lambda}_{s,s},\sigma^{\lambda*}_{s,s}\big]\varepsilon^{\lambda}_{i_1\cdots i_s}\varepsilon^{\lambda*}_{i_1\cdots i_s}\delta(\mathbf{k}-\mathbf{k}')\\&\quad=\frac{4k(Z^{\lambda}_s)^2}{\pi}\left(\frac{H}{k}\right)^{2(s-1)}\frac{\Gamma(\frac{1}{2}+s+i\mu_s)\Gamma(\frac{1}{2}+s-i\mu_s)}{\Gamma(\frac{1}{2}+\lambda+i\mu_s)\Gamma(\frac{1}{2}+\lambda-i\mu_s)}\varepsilon^{\lambda}_{i_1\cdots i_s}\varepsilon^{\lambda*}_{i_1\cdots i_s}\delta(\mathbf{k}-\mathbf{k}')\,.\end{aligned} \tag{B.88}$$

Note that our final normalization depends on the normalization of the polarization tensors. This does not affect correlation functions, however, as we show in the next section. Plugging (B.12) into (B.88) and imposing (B.75), we obtain (B.76).

B.2 Two-point function

In this section, we will compute the two-point functions of spinning fields. For this purpose, it will be convenient to contract free indices of the spinning fields with auxiliary vectors. In other words, we will compute

$$\big\langle(n\cdot\sigma)^2\big\rangle'_s\equiv\big\langle(n_{i_1}\cdots n_{i_s}\sigma_{i_1\cdots i_s}(\eta))\,(\tilde{n}_{j_1}\cdots\tilde{n}_{j_s}\sigma_{j_1\cdots j_s}(\eta'))\big\rangle'\,, \tag{B.89}$$

where $\mathbf{n} \equiv (\cos\alpha, \sin\alpha, i)$ and $\tilde{\mathbf{n}} \equiv (\cos\beta, \sin\beta, -i)$ are null vectors. For generic η and η', the two-point function is

$$\big\langle (n\cdot\sigma)^2 \big\rangle'_s = \sum_{\lambda=-s}^{s} e^{i\lambda\chi} \left[\frac{(2s-1)!!}{(2\lambda-1)!!(s-\lambda)!}\right]^2 \sigma^{\lambda}_{s,s}(-k\eta)\sigma^{\lambda *}_{s,s}(-k\eta')\,, \tag{B.90}$$

where $\chi \equiv \alpha - \beta$. In the late-time limit (or the long-wavelength limit), the two-point function simplifies considerably. We get

$$\lim_{\eta,\eta'\to 0} \big\langle (n\cdot\sigma)^2 \big\rangle'_s = \frac{(H^2\eta\eta')^{3/2-s}}{4\pi H} \sum_{\lambda=-s}^{s} e^{i\lambda\chi} \left[\mathcal{C}(\mu_s,\lambda,s)\,\Gamma(-i\mu_s)^2 \left(\frac{k^2\eta\eta'}{4}\right)^{i\mu_s} + c.c.\right], \tag{B.91}$$

where

$$\mathcal{C}(\mu_s,\lambda,s) \equiv \frac{(2s-1)!!\,s!}{(s-\lambda)!(s+\lambda)!} \frac{\Gamma(\frac{1}{2}+s-i\mu_s)\Gamma(\frac{1}{2}+\lambda+i\mu_s)}{\Gamma(\frac{1}{2}+s+i\mu_s)\Gamma(\frac{1}{2}+\lambda-i\mu_s)}\,. \tag{B.92}$$

This late-time expectation value matches the two-point function of a spin-s field of a conformal field theory living on the future boundary, which have been computed in [102].

Derivation of (B.91).—The two-point function (B.89) can be written as

$$\big\langle (n\cdot\sigma)^2 \big\rangle'_s = \sum_{\lambda=-s}^{s} (n_{i_1}\cdots n_{i_s}\varepsilon^{\lambda}_{i_1\cdots i_s})(\tilde{n}_{j_1}\cdots \tilde{n}_{j_s}\varepsilon^{\lambda *}_{j_1\cdots j_s})\,\sigma^{\lambda}_{s,s}\sigma^{\lambda *}_{s,s}\,. \tag{B.93}$$

Let us compute $\sigma^{\lambda}_{s,s}\sigma^{\lambda *}_{s,s}$ in the late-time limit. First, recall that we can cast the mode function in the form

$$\sigma^{\lambda}_{n,s} = \mathcal{A}_s\, Z^{\lambda}_s (-k\eta)^{3/2-n}\Big[(x_n+iy_n)H_{i\mu_s} + (w_n+iz_n)k\eta H_{i\mu_s+1}\Big]\,. \tag{B.94}$$

Taking the asymptotic limits of the Hankel functions, we get

$$\sigma^{\lambda}_{n,s}\sigma^{\lambda *}_{n,s}\Big|_{\eta,\eta'\to 0} = (Z^{\lambda}_s)^2 \frac{(k^2\eta\eta')^{3/2-n}}{\pi^2} \left[W_n\,\Gamma(-i\mu_s)^2\left(\frac{k^2\eta\eta'}{4}\right)^{i\mu_s} + c.c.\right] + \text{local terms}\,, \tag{B.95}$$

where

$$W_n \equiv x_n^2 + y_n^2 + 2\mu_s(x_n+iy_n)(iw_n+z_n)\,. \tag{B.96}$$

Using (B.82), we obtain the recursion relation

$$W_{n+1} = \big(n+\tfrac{1}{2}-i\mu_s\big)^2\, W_n\,. \tag{B.97}$$

Following similar arguments as in the previous section, it can then be shown that

$$W_n = \frac{\Gamma(\frac{1}{2}+s-i\mu_s)^2}{\Gamma(\frac{1}{2}+\lambda-i\mu_s)^2}\,. \tag{B.98}$$

Substituting (B.76), (B.95), and (B.98) into (B.93), we obtain

$$\langle (n\cdot\sigma)^2\rangle'_s = \frac{(H^2\eta\eta')^{3/2-s}}{4\pi H}\sum_{\lambda=-s}^{s} I_s^\lambda(\mathbf{n},\tilde{\mathbf{n}})\left[\mathcal{D}(s,\lambda,\mu_s)\Gamma(-i\mu_s)^2\left(\frac{k^2\eta\eta'}{4}\right)^{i\mu_s}+c.c.\right], \tag{B.99}$$

where we have dropped the local terms and defined

$$I_s^\lambda(\mathbf{n},\tilde{\mathbf{n}}) \equiv \frac{(n_{i_1}\cdots n_{i_s}\varepsilon^\lambda_{i_1\cdots i_s})(\tilde{n}_{i_1}\cdots\tilde{n}_{i_s}\varepsilon^{\lambda *}_{i_1\cdots i_s})}{\varepsilon^\lambda_{i_1\cdots i_s}\varepsilon^{\lambda *}_{i_1\cdots i_s}}, \tag{B.100}$$

$$\mathcal{D}(s,\lambda,\mu_s) \equiv \frac{\Gamma(\frac{1}{2}+s-i\mu_s)\Gamma(\frac{1}{2}+\lambda+i\mu_s)}{\Gamma(\frac{1}{2}+s+i\mu_s)\Gamma(\frac{1}{2}+\lambda-i\mu_s)}. \tag{B.101}$$

To obtain an expression for I_s^λ, let us first recall that the structure of the polarization tensors are given by the (associated) Legendre polynomials. Contracting with null vectors, only the term with the leading power in k survives (with no Kronecker delta's), whose coefficient is $(2s-1)!!/[(2\lambda-1)!!(s-\lambda)!]$. This means that

$$(n_{i_1}\cdots n_{i_s}\varepsilon^\lambda_{i_1\cdots i_s})(\tilde{n}_{i_1}\cdots\tilde{n}_{i_s}\varepsilon^{\lambda *}_{i_1\cdots i_s}) = \left[\frac{(2s-1)!!}{(2\lambda-1)!!(s-\lambda)!}\right]^2 e^{is\chi}, \tag{B.102}$$

where we used the fact that we get one factor of $e^{i\alpha}$ for each contraction with a null vector, i.e. $n_{i_1}\cdots n_{i_s}\varepsilon^s_{i_1\cdots i_s} = e^{is\alpha}$. Combining (B.102) and (B.12), we get

$$I_s^\lambda(\mathbf{n},\tilde{\mathbf{n}}) = \frac{(2s-1)!!\,s!}{(s-\lambda)!(s+\lambda)!}e^{i\lambda\chi}. \tag{B.103}$$

Substituting this into (B.99), we obtain (B.91).

C. Details of in-in computations

In this appendix, we present details of the in-in computations of Chapter 5. In particular, we will give explicit expressions for the shape functions introduced in (5.61), (5.76), (5.79) and (5.84).

Preliminaries. The expectation value of an operator $\mathcal{Q}$ is computed by

$$\langle \mathcal{Q}(\eta) \rangle = \langle 0 | \left[\bar{\mathrm{T}} e^{i \int_{-\infty}^{\eta} \mathrm{d}\eta' \hat{H}_I(\eta')} \right] \mathcal{Q}(\eta) \left[\mathrm{T} e^{-i \int_{-\infty}^{\eta} \mathrm{d}\eta' \hat{H}_I(\eta')} \right] |0\rangle \ , \tag{C.1}$$

where $|0\rangle$ is the vacuum state of the free theory, T and $\bar{\mathrm{T}}$ denote time-ordering and anti-time-ordering, respectively, and $\hat{H}_I$ is the interaction Hamiltonian. To compute the quantum expectation values, we promote the fields π, γ, σ to operators and expand in Fourier space

$$\pi(\mathbf{k},\eta) = \pi_k(\eta) a(\mathbf{k}) + h.c. \,, \quad \gamma_{ij}(\mathbf{k},\eta) = \sum_{\lambda=\pm 2} \varepsilon_{ij}^{\lambda}(\mathbf{k}) \gamma_k^{\lambda}(\eta) b(\mathbf{k},\lambda) + h.c. \ , \tag{C.2}$$

$$\hat{\sigma}_{i_1 \cdots i_s}(\mathbf{k},\eta) = \sum_{\lambda=-s}^{s} \varepsilon_{i_1 \cdots i_s}^{\lambda}(\mathbf{k}) \sigma_{s,s}^{\lambda}(k,\eta) b_s(\mathbf{k},\lambda) + h.c. \,, \tag{C.3}$$

where the creation and annihilation operators obey the usual canonical commutation relations

$$[a(\mathbf{k}), a^{\dagger}(\mathbf{k}')] = (2\pi)^3 \delta(\mathbf{k}-\mathbf{k}') \ , \tag{C.4}$$

$$[b(\mathbf{k},\lambda), b^{\dagger}(\mathbf{k}',\lambda')] = [b_s(\mathbf{k},\lambda), b_s^{\dagger}(\mathbf{k}',\lambda')] = (2\pi)^3 \delta(\mathbf{k}-\mathbf{k}') \delta_{\lambda\lambda'} \ . \tag{C.5}$$

The mode functions for the Goldstone and the graviton are

$$\pi_k(\eta) = \frac{H}{f_\pi^2} \frac{i}{\sqrt{2k^3}} (1 + i c_\pi k\eta) e^{-i c_\pi k \eta} \,, \quad \gamma_k^{\lambda}(\eta) = \frac{H}{M_{\rm pl}} \frac{i}{\sqrt{2k^3}} (1 + ik\eta) e^{-ik\eta} \,. \tag{C.6}$$

The mode functions $\sigma_{s,s}^{\lambda}(k,\eta)$ were derived in Appendix B. It will be convenient to write the longitudinal and helicity-± 2 mode functions as

$$\sigma_{s,s}^{0}(-k\eta) = N_s (-k\eta)^{3/2-s} G_{i\mu_s}^{(s)}(-k\eta) \,, \quad \sigma_{s,s}^{\pm 2}(-k\eta) = Z_s^{\pm 2} (-k\eta)^{3/2-s} \tilde{G}_{i\mu_s}^{(s)}(-k\eta) \,, \tag{C.7}$$

where the functions $G_{i\mu_s}^{(s)} \equiv G_{i\mu_s}^{(s,\lambda=0,n=s)}$ and $\tilde{G}_{i\mu_s}^{(s)} \equiv G_{i\mu_s}^{(s,\lambda=\pm 2,n=s)}$ can be obtained recursively

using (B.70), or

$$G^{(s,\lambda,n+1)}_{i\mu_s}(x) = \frac{i}{2}\left[2x\partial_x G^{(s,\lambda,n)}_{i\mu_s}(x) + (1-2n)G^{(s,\lambda,n)}_{i\mu_s}(x)\right] - \sum_{m=\lambda}^{s} B_{m,n+1} G^{(s,\lambda,m)}_{i\mu_s}(x)\,, \tag{C.8}$$

given $G^{(s,\lambda,\lambda)}_{i\mu_s}(x) = \mathcal{A}_s H_{i\mu_s}(x)$. For $s = 1$ and 2, we get

$$G^{(1)}_{i\mu_1}(x) \equiv \frac{i}{2}\mathcal{A}_1\left[x\big(H_{i\mu_1-1}(x) - H_{i\mu_1+1}(x)\big) - H_{i\mu_1}(x)\right], \tag{C.9}$$

$$G^{(2)}_{i\mu_2}(x) \equiv \frac{1}{12}\mathcal{A}_2\left[6x\big[(2-i\mu_2)H_{i\mu_2-1}(x) - (2+i\mu_2)H_{i\mu_2+1}(x)\big] - (9-8x^2)H_{i\mu_2}(x)\right]. \tag{C.10}$$

C.1 Integral expressions

In §5.3, the results for the bispectra were defined in terms of a number of momentum-dependent functions. In the following, we give explicit integral expressions for these functions:

- For $s \geq 2$, the functions $\mathcal{I}^{(s)}$ in (5.61) are given by

$$\mathcal{I}^{(s)} \equiv \sum_{j=1}^{3} \frac{2\pi^3 \mathcal{N}_s^2}{k_1^{3/2}k_2^{7/2}k_3}\mathrm{Re}[\mathcal{I}_j^{(s)}]\,, \tag{C.11}$$

$$\mathcal{I}_1^{(s)} \equiv -\int_0^\infty \mathrm{d}x\, \tilde{\mathcal{T}}^{(s)*}_{i\mu_s}(c_\pi,k_1,k_2,k_3,x)\int_0^\infty \mathrm{d}y\, \tilde{\mathcal{F}}^{(s)}_{i\mu_s}(c_\pi,y)\,, \tag{C.12}$$

$$\mathcal{I}_2^{(s)} \equiv \int_0^\infty \mathrm{d}x\, \mathcal{T}^{(s)}_{i\mu_s}(c_\pi,k_1,k_2,k_3,x)\int_{\kappa_{12}x/c_\pi}^\infty \mathrm{d}y\, \tilde{\mathcal{F}}^{(s)}_{i\mu_s}(c_\pi,y)\,, \tag{C.13}$$

$$\mathcal{I}_3^{(s)} \equiv \int_0^\infty \mathrm{d}x\, \mathcal{F}^{(s)}_{i\mu_s}(c_\pi,x)\int_{c_\pi\kappa_{21}x}^\infty \mathrm{d}y\, \tilde{\mathcal{T}}^{(s)}_{i\mu_s}(c_\pi,k_1,k_2,k_3,y)\,, \tag{C.14}$$

where $\kappa_{ij} \equiv k_i/k_j$ and $\mathcal{N}_s \equiv \mathcal{Z}_s^0$ is the normalization constant defined in (B.77). The integrands are represented by the functions

$$\mathcal{F}^{(s)}_{i\mu_s}(c_\pi,x) \equiv x^{s-5/2}(1+ic_\pi x)G^{(s)}_{i\mu_s}(y)e^{-ic_\pi x}\,, \tag{C.15}$$

$$\tilde{\mathcal{F}}^{(s)}_{i\mu_s}(c_\pi,x) \equiv x^{s-5/2}(1+ic_\pi x)G^{(s)*}_{i\mu_s}(y)e^{-ic_\pi x}\,, \tag{C.16}$$

$$\mathcal{T}^{(s)}_{i\mu_s}(c_\pi,k_1,k_2,k_3,x) \equiv x^{s-1/2}\,(1+ix)G^{(s)}_{i\mu_s}(xk_1/c_\pi k_2)e^{-ix(1+k_3/k_2)}\,, \tag{C.17}$$

$$\tilde{\mathcal{T}}^{(s)}_{i\mu_s}(c_\pi,k_1,k_2,k_3,x) \equiv x^{s-1/2}\,(1+ix)G^{(s)*}_{i\mu_s}(xk_1/c_\pi k_2)e^{-ix(1+k_3/k_2)}\,, \tag{C.18}$$

where $G^{(s)}_{i\mu_s}$ was defined in (C.7). The integral $\int_0^\infty \mathrm{d}x\, \mathcal{F}^{(1)}_{i\mu_1}$ is in fact IR divergent. To avoid this issue, we integrate by parts and work with $\mathcal{F}^{(1)}_{i\mu_1} \to x^{1/2}H_{i\mu_1}(x)e^{ix}$ and $\tilde{\mathcal{F}}^{(1)}_{i\mu_1} \to x^{1/2}H_{i\mu_1}(x)e^{-ix}$.

- The functions $\mathcal{J}^{(s)}$ in (5.76) are given by

$$\mathcal{J}^{(s)} = \sum_{j=1}^{6} \frac{2\pi^3 \mathcal{N}_s^4}{k_1^3 k_2^{3/2}k_3^{3/2}}\mathrm{Im}[\mathcal{J}_j^{(s)}]\,, \tag{C.19}$$

$$\mathcal{J}_1^{(s)} \equiv -\int_0^\infty \mathrm{d}x\, \hat{\mathcal{G}}_{i\mu_s}^{(s)*}(c_\pi,k_1,k_2,k_3,x)\int_0^\infty \mathrm{d}y\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}(c_\pi,y)\int_{\kappa_{32}y}^\infty \mathrm{d}z\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}(c_\pi,z)\,, \tag{C.20}$$

$$\mathcal{J}_2^{(s)} \equiv -\int_0^\infty \mathrm{d}x\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)*}(c_\pi,x)\int_0^\infty \mathrm{d}y\, \tilde{\mathcal{G}}_{i\mu_s}^{(s)}(c_\pi,k_1,k_2,k_3,y)\int_{\kappa_{31}y/c_\pi}^\infty \mathrm{d}z\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}(c_\pi,z)\,, \tag{C.21}$$

$$\mathcal{J}_3^{(s)} \equiv -\int_0^\infty \mathrm{d}x\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)*}(c_\pi,x)\int_0^\infty \mathrm{d}y\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}(c_\pi,y)\int_{c_\pi\kappa_{12}y}^\infty \mathrm{d}z\, \hat{\mathcal{G}}_{i\mu_s}^{(s)}(c_\pi,k_1,k_2,k_3,z)\,, \tag{C.22}$$

$$\mathcal{J}_4^{(s)} \equiv \int_0^\infty \mathrm{d}x\, \mathcal{G}_{i\mu_s}^{(s)}(c_\pi,k_1,k_2,k_3,x)\int_{\kappa_{12}x/c_\pi}^\infty \mathrm{d}y\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}(c_\pi,y)\int_{\kappa_{31}y}^\infty \mathrm{d}z\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}(c_\pi,z)\,, \tag{C.23}$$

$$\mathcal{J}_5^{(s)} \equiv \int_0^\infty \mathrm{d}x\, \mathcal{F}_{i\mu_s}^{(s)}(c_\pi,x)\int_{c_\pi\kappa_{32}x}^\infty \mathrm{d}y\, \tilde{\mathcal{G}}_{i\mu_s}^{(s)}(c_\pi,k_1,k_2,k_3,y)\int_{\kappa_{12}y/c_\pi}^\infty \mathrm{d}z\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}(c_\pi,z)\,, \tag{C.24}$$

$$\mathcal{J}_6^{(s)} \equiv \int_0^\infty \mathrm{d}x\, \mathcal{F}_{i\mu_s}^{(s)}(c_\pi,x)\int_{\kappa_{21}x}^\infty \mathrm{d}y\, \mathcal{F}_{i\mu_s}^{(s)}(c_\pi,y)\int_{c_\pi\kappa_{32}y}^\infty \mathrm{d}z\, \hat{\mathcal{G}}_{i\mu_s}^{(s)}(c_\pi,k_1,k_2,k_3,z)\,, \tag{C.25}$$

where

$$\mathcal{G}_{i\mu_s}^{(s)}(c_\pi,k_1,k_2,k_3,x) \equiv x\, G_{i\mu_s}^{(s)}(xk_2/c_\pi k_1)G_{i\mu_s}^{(s)}(xk_3/c_\pi k_1)e^{-ix}\,, \tag{C.26}$$

$$\tilde{\mathcal{G}}_{i\mu_s}^{(s)}(c_\pi,k_1,k_2,k_3,x) \equiv x\, G_{i\mu_s}^{(s)*}(xk_2/c_\pi k_1)G_{i\mu_s}^{(s)}(xk_3/c_\pi k_1)e^{-ix}\,, \tag{C.27}$$

$$\hat{\mathcal{G}}_{i\mu_s}^{(s)}(c_\pi,k_1,k_2,k_3,x) \equiv x\, G_{i\mu_s}^{(s)*}(xk_2/c_\pi k_1)G_{i\mu_s}^{(s)*}(xk_3/c_\pi k_1)e^{-ix}\,. \tag{C.28}$$

- The functions $\mathcal{K}^{(s)}$ in (5.79) are given by

$$\mathcal{K}^{(s)} = \sum_{j=1}^{10}\frac{2\pi^3\mathcal{N}_s^6}{k_1^3k_2^{3/2}k_3^{3/2}}\mathrm{Re}[\mathcal{K}_j^{(s)}]\,, \tag{C.29}$$

$$\mathcal{K}_1^{(s)} \equiv -\int_0^\infty \mathrm{d}w\, \mathcal{H}_{i\mu_s}^{(s)}(k_1,k_2,k_3,w)\int_w^\infty \mathrm{d}x\, \mathcal{F}_{i\mu_s}^{(s)*}\int_0^\infty \mathrm{d}y\, \mathcal{F}_{i\mu_s}^{(s)*}\int_{\kappa_{32}y}^\infty \mathrm{d}z\, \mathcal{F}_{i\mu_s}^{(s)*}\,, \tag{C.30}$$

$$\mathcal{K}_2^{(s)} \equiv -\int_0^\infty \mathrm{d}w\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)*}\int_w^\infty \mathrm{d}x\, \tilde{\mathcal{H}}_{i\mu_s}^{(s)}(k_1,k_2,k_3,x)\int_0^\infty \mathrm{d}y\, \mathcal{F}_{i\mu_s}^{(s)*}\int_{\kappa_{32}y}^\infty \mathrm{d}z\, \mathcal{F}_{i\mu_s}^{(s)*}\,, \tag{C.31}$$

$$\mathcal{K}_3^{(s)} \equiv \int_0^\infty \mathrm{d}w\, \mathcal{H}_{i\mu_s}^{(s)}(k_1,k_2,k_3,w)\int_0^\infty \mathrm{d}x\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}\int_{\kappa_{21}x}^\infty \mathrm{d}y\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}\int_{\kappa_{32}y}^\infty \mathrm{d}z\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}\,, \tag{C.32}$$

$$\mathcal{K}_4^{(s)} \equiv \int_0^\infty \mathrm{d}w\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)*}\int_0^\infty \mathrm{d}x\, \tilde{\mathcal{H}}_{i\mu_s}^{(s)}(k_1,k_2,k_3,x)\int_{\kappa_{21}x}^\infty \mathrm{d}y\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}\int_{\kappa_{32}y}^\infty \mathrm{d}z\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}\,, \tag{C.33}$$

$$\mathcal{K}_5^{(s)} \equiv \int_0^\infty \mathrm{d}w\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)*}\int_0^\infty \mathrm{d}x\, \mathcal{F}_{i\mu_s}^{(s)}\int_{\kappa_{12}x}^\infty \mathrm{d}y\, \hat{\mathcal{H}}_{i\mu_s}^{(s)}(k_1,k_2,k_3,y)\int_{\kappa_{31}y}^\infty \mathrm{d}z\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}\,, \tag{C.34}$$

$$\mathcal{K}_6^{(s)} \equiv \int_0^\infty \mathrm{d}w\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)*}\int_0^\infty \mathrm{d}x\, \mathcal{F}_{i\mu_s}^{(s)}\int_{\kappa_{32}x}^\infty \mathrm{d}y\, \mathcal{F}_{i\mu_s}^{(s)}\int_{\kappa_{13}y}^\infty \mathrm{d}z\, \bar{\mathcal{H}}_{i\mu_s}^{(s)}(k_1,k_2,k_3,w)\,, \tag{C.35}$$

$$\mathcal{K}_7^{(s)} \equiv -\int_0^\infty \mathrm{d}w\, \mathcal{H}_{i\mu_s}^{(s)}(k_1,k_2,k_3,w)\int_w^\infty \mathrm{d}x\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}\int_{\kappa_{21}x}^\infty \mathrm{d}y\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}\int_{\kappa_{32}y}^\infty \mathrm{d}z\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}\,, \tag{C.36}$$

$$\mathcal{K}_8^{(s)} \equiv -\int_0^\infty \mathrm{d}w\, \mathcal{F}_{i\mu_s}^{(s)}\int_w^\infty \mathrm{d}x\, \tilde{\mathcal{H}}_{i\mu_s}^{(s)}(k_1,k_2,k_3,x)\int_{\kappa_{21}x}^\infty \mathrm{d}y\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}\int_{\kappa_{32}y}^\infty \mathrm{d}z\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}\,, \tag{C.37}$$

$$\mathcal{K}_9^{(s)} \equiv -\int_0^\infty \mathrm{d}w\, \mathcal{F}_{i\mu_s}^{(s)}\int_{\kappa_{21}w}^\infty \mathrm{d}x\, \mathcal{F}_{i\mu_s}^{(s)}\int_{\kappa_{12}x}^\infty \mathrm{d}y\, \hat{\mathcal{H}}_{i\mu_s}^{(s)}(k_1,k_2,k_3,y)\int_{\kappa_{31}y}^\infty \mathrm{d}z\, \tilde{\mathcal{F}}_{i\mu_s}^{(s)}\,, \tag{C.38}$$

$$\mathcal{K}_{10}^{(s)} \equiv -\int_0^\infty \mathrm{d}w\, \mathcal{F}_{i\mu_s}^{(s)} \int_{\kappa_{21}w}^\infty \mathrm{d}x\, \mathcal{F}_{i\mu_s}^{(s)} \int_{\kappa_{32}x}^\infty \mathrm{d}y\, \mathcal{F}_{i\mu_s}^{(s)} \int_{\kappa_{13}y}^\infty \mathrm{d}z\, \bar{\mathcal{H}}_{i\mu_s}^{(s)}(k_1,k_2,k_3,w)\,, \tag{C.39}$$

where we have suppressed some arguments and defined

$$\mathcal{H}_{i\mu_s}^{(s)}(k_1,k_2,k_3,x) \equiv x^{1/2} G_{i\mu_s}^{(s)}(x) G_{i\mu_s}^{(s)}(k_2x/k_1) G_{i\mu_s}^{(s)}(k_3x/k_1)\,, \tag{C.40}$$

$$\tilde{\mathcal{H}}_{i\mu_s}^{(s)}(k_1,k_2,k_3,x) \equiv x^{1/2} G_{i\mu_s}^{(s)*}(x) G_{i\mu_s}^{(s)}(k_2x/k_1) G_{i\mu_s}^{(s)}(k_3x/k_1)\,, \tag{C.41}$$

$$\hat{\mathcal{H}}_{i\mu_s}^{(s)}(k_1,k_2,k_3,x) \equiv x^{1/2} G_{i\mu_s}^{(s)*}(x) G_{i\mu_s}^{(s)*}(k_2x/k_1) G_{i\mu_s}^{(s)}(k_3x/k_1)\,, \tag{C.42}$$

$$\bar{\mathcal{H}}_{i\mu_s}^{(s)}(k_1,k_2,k_3,x) \equiv x^{1/2} G_{i\mu_s}^{(s)*}(x) G_{i\mu_s}^{(s)*}(k_2x/k_1) G_{i\mu_s}^{(s)*}(k_3x/k_1)\,. \tag{C.43}$$

- The functions $\mathcal{B}^{(s)}$ in (5.84) are given by

$$\mathcal{B}^{(s)} \equiv \sum_{i=1}^3 \frac{\pi^3 \tilde{\mathcal{N}}_s^2}{4k_1^{3/2}k_2^{7/2}k_3} \mathrm{Re}[\mathcal{B}_i^{(s)}]\,, \tag{C.44}$$

$$\mathcal{B}_1^{(s)} = -\int_0^\infty \mathrm{d}x\, \tilde{\mathcal{R}}_{i\mu_s}^{(s)*}(c_\pi,k_1,k_2,k_3,x) \int_0^\infty \mathrm{d}y\, y^{s-5/2} \tilde{G}_{i\mu_2}^{(s)*}(y) e^{-iy}\,, \tag{C.45}$$

$$\mathcal{B}_2^{(s)} = \int_0^\infty \mathrm{d}x\, \mathcal{R}_{i\mu_s}^{(s)}(c_\pi,k_1,k_2,k_3,x) \int_{\kappa_{13}x/c_\pi}^\infty \mathrm{d}y\, y^{s-5/2} \tilde{G}_{i\mu_2}^{(s)*}(y) e^{-iy}\,, \tag{C.46}$$

$$\mathcal{B}_3^{(s)} = \int_0^\infty \mathrm{d}x\, x^{s-5/2} \tilde{G}_{i\mu_2}^{(s)}(x) e^{-ix} \int_{c_\pi\kappa_{31}x}^\infty \mathrm{d}y\, \tilde{\mathcal{R}}_{i\mu_s}^{(s)}(c_\pi,k_1,k_2,k_3,x)\,. \tag{C.47}$$

where $\tilde{\mathcal{N}}_s \equiv \mathcal{Z}_s^{\pm 2}$ is the normalization constant defined in (B.77) and

$$\mathcal{R}_{i\mu_s}^{(s)}(c_\pi,k_1,k_2,k_3,x) \equiv x^{s-1/2}(1+ix) \tilde{G}_{i\mu_s}^{(s)}(xk_1/c_\pi k_2) e^{-ix(1+k_3/k_2)}\,, \tag{C.48}$$

$$\tilde{\mathcal{R}}_{i\mu_s}^{(s)}(c_\pi,k_1,k_2,k_3,x) \equiv x^{s-1/2}(1+ix) \tilde{G}_{i\mu_s}^{(s)*}(xk_1/c_\pi k_2) e^{-ix(1+k_3/k_2)}\,. \tag{C.49}$$

C.2 Soft limits

In this section, we will derive analytic formulas for the soft limits of the non-analytic parts of all correlation functions that we considered in this work.

$\langle\zeta\zeta\zeta\rangle$. We will focus on the squeezed limit of the scalar three-point function for the single-exchange diagram (cf. Fig. 5.6a), and consider even spins first. This leads to a non-analytic behavior if the quadratic mixing leg is taken to be soft. In the squeezed limit, $k_1 \ll k_2 \approx k_3$, this contribution is given by

$$\lim_{k_1\ll k_3} \frac{\langle \zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle'}{\Delta_\zeta^4} = \alpha_s \Delta_\zeta^{-1} \times P_s(\hat{\mathbf{k}}_1\cdot\hat{\mathbf{k}}_3) \times \mathcal{I}^{(s)}(\mu_s,c_\pi,k_1,k_3,k_3) + (\mathbf{k}_2 \leftrightarrow \mathbf{k}_3)\,, \tag{C.50}$$

where the functions $\mathcal{I}^{(s)}$ are given by

$$\mathcal{I}^{(s)} \equiv -\frac{(2\pi)^3 c_\pi^{s-3/2} H^{5-2s}}{8} \sum_{\pm\pm} (\pm i k_1^{s-3})(\pm i c_\pi^2 k_3^{s-4})\, \mathcal{I}_{\pm\pm}^{(s)}\,, \tag{C.51}$$

$$\mathcal{I}^{(s)}_{\pm\pm} \equiv \int_{-\infty}^{0} \frac{\mathrm{d}\eta}{a^{2s-3}}\,\eta(1\mp ic_\pi k_3\eta)e^{\pm 2ic_\pi k_3\eta}\int_{-\infty}^{0}\frac{\mathrm{d}\tilde{\eta}}{a^{2s-4}}\,(1\mp ic_\pi k_1\tilde{\eta})e^{\pm ic_\pi q\tilde{\eta}}\,G_{\pm\pm}(k_1,\eta,\tilde{\eta})\,. \tag{C.52}$$

In (C.52) we introduced the time-ordered Green's functions on the Schwinger-Keldysh contours

$$G_{++}(k,\eta,\tilde{\eta}) = G^{>}(k,\eta,\tilde{\eta})\,\Theta(\eta-\tilde{\eta}) + G^{<}(k,\eta,\tilde{\eta})\,\Theta(\tilde{\eta}-\eta)\,, \tag{C.53}$$

$$G_{+-}(k,\eta,\tilde{\eta}) = G^{<}(k,\eta,\tilde{\eta})\,, \tag{C.54}$$

$$G_{-+}(k,\eta,\tilde{\eta}) = G^{>}(k,\eta,\tilde{\eta})\,, \tag{C.55}$$

$$G_{--}(k,\eta,\tilde{\eta}) = G^{<}(k,\eta,\tilde{\eta})\,\Theta(\eta-\tilde{\eta}) + G^{>}(k,\eta,\tilde{\eta})\,\Theta(\tilde{\eta}-\eta)\,, \tag{C.56}$$

where

$$G^{>}(k,\eta,\tilde{\eta}) = \sigma^{0}_{s,s}(-k\eta)\,\sigma^{0*}_{s,s}(-k\tilde{\eta})\,,\quad G^{<}(k,\eta,\tilde{\eta}) = \sigma^{0*}_{s,s}(-k\eta)\,\sigma^{0}_{s,s}(-k\tilde{\eta})\,, \tag{C.57}$$

denote the Wightman functions of the longitudinal mode of a spin-s field, and $\pm$ indicates the (anti-)time-ordering along the integration contour. The non-local part of the Green's function is independent of the sign of the time difference, in which case the time-ordered Green's can be replaced with the non-time-ordered ones, $G_{+\pm} = G_{-\pm}$. The integrals thus factorize, and substituting for the σ mode functions, the integral (C.51) becomes

$$\mathcal{I}^{(s)} = N^{(s)}\sum_{\pm}(\pm k_1^{s-3}k_3^{s-4})\,\mathcal{P}^{(s)}_{\pm}(k_1,c_\pi k_3)\,\mathcal{Q}^{(s)*}(c_\pi,k_1) + c.c.\,, \tag{C.58}$$

where we used the fact that $\mathcal{I}^{(s)}_{+\pm} = \mathcal{I}^{(s)*}_{-\mp}$ and defined

$$N^{(s)} \equiv -\frac{s!\pi^5 c_\pi^{s+1/2}}{4(2s-1)!!}\frac{\operatorname{sech}\pi\mu_s}{\Gamma(\frac{1}{2}+s-i\mu_s)\Gamma(\frac{1}{2}+s+i\mu_s)}\,, \tag{C.59}$$

$$\mathcal{P}^{(s)}_{\pm}(k_1,c_\pi k_3) \equiv e^{-\pi\mu_s/2}\int_0^\infty \mathrm{d}x\,x^{s-1/2}(1\mp ic_\pi k_3 x)G^{(s)}_{i\mu_s}(k_1x)e^{\pm 2ic_\pi k_3 x}\,, \tag{C.60}$$

$$\mathcal{Q}^{(s)*}(c_\pi,k_1) \equiv e^{-\pi\mu_s/2}\int_0^\infty \mathrm{d}x\,x^{s-5/2}(1+ic_\pi k_1 x)G^{(s)*}_{i\mu_s}(k_1x)e^{-ic_\pi k_1 x}\,, \tag{C.61}$$

with $G^{(s)}_{i\mu_s}$ introduced in (C.7). The integrals in (C.60) and (C.61) can be computed analytically for arbitrary c_π. To derive the results below, we will use the formula

$$e^{-\pi\mu_s/2}\int_0^\infty \mathrm{d}x\,x^n H_{i\mu_s}(bx)e^{iax} = \frac{(i/2)^n}{\sqrt{\pi}\,b^{n+1}}F_{21}(n+3/2,\mu_s,(b-a)/2b)\,, \tag{C.62}$$

where

$$F_{21}(a,\mu_s,z) \equiv \frac{\Gamma(a-\frac{1}{2}-i\mu_s)\Gamma(a-\frac{1}{2}+i\mu_s)}{\Gamma(a)}\,{}_2F_1\Big(a-\frac{1}{2}-i\mu_s,a-\frac{1}{2}+i\mu_s,a+s,z\Big)\,. \tag{C.63}$$

In the squeezed limit, $k_1 \ll c_\pi k_3$, the result for (C.60) is

$$\lim_{k_1 \ll c_\pi k_3} \mathcal{P}_\pm^{(s)}(k_1, c_\pi k_3) = \frac{(-i)^{1/2} e^{(1\mp 1)\pi\mu_s/2}\Gamma(\frac{1}{2}+s+i\mu_s)\Gamma(\frac{1}{2}+s-i\mu_s)}{4\pi(\pm 2c_\pi k_3)^{1/2+s}} \times \left(\frac{k_1}{4c_\pi k_3}\right)^{i\mu_s} (5+2s+2i\mu_s)\frac{\Gamma(-i\mu_s)}{\Gamma(\frac{1}{2}-i\mu_s)} \mp i e^{-(1\mp 1)\pi\mu_s} \times c.c.\,. \quad \text{(C.64)}$$

Since we cannot take a soft limit of the integral (C.61), its general expression is rather complicated. For simplicity, let us display the results for the two limiting cases, $c_\pi = 1$ and $c_\pi \ll 1$, for which (C.61) reduces to

$$\mathcal{Q}^{(s)*}(c_\pi = 1, k_1) = f^{(s)}(1) \times \frac{i(2ik_1)^{3/2-s}}{\sqrt{\pi}\Gamma(s)} \frac{\Gamma(\frac{1}{2}+s-i\mu_s)\Gamma(\frac{1}{2}+s+i\mu_s)}{(s-\frac{3}{2})^2+\mu_s^2}\,, \quad \text{(C.65)}$$

$$\mathcal{Q}^{(s)*}(c_\pi \ll 1, k_1) = f^{(s)}(0) \times \frac{2i(ik_1/2)^{3/2-s}}{\pi} \frac{\Gamma[\frac{1}{2}(\frac{1}{2}+s+i\mu_s)]\Gamma[\frac{1}{2}(\frac{1}{2}+s-i\mu_s)]}{(s-\frac{3}{2})^2+\mu_s^2}\,. \quad \text{(C.66)}$$

Notice that the mixing integral becomes independent of c_π in the small c_π limit. The function $f^{(s)}(c_\pi)$ is precisely the difference between evaluating the integral (C.61) with the mode function $G^{(s)}_{i\mu_s}$ and a simple Hankel function $H_{i\mu_s}$. Since the mode function is a linear combination of Hankel functions, $f^{(s)}$ is a simple polynomial. The result for $s = 2$ is

$$f^{(2)}(1) = -\frac{985 - 664\mu_2^2 + 16\mu_2^4}{576}\,, \quad \text{(C.67)}$$

$$f^{(2)}(0) = -\frac{23 - 4\mu_2^2}{12}\,. \quad \text{(C.68)}$$

Summing (C.65) and (C.60) and focusing on terms which are non-analytic in momentum, we find

$$\lim_{k_1 \ll c_\pi k_3} \mathcal{I}^{(s)}(\mu_s, c_\pi, k_1, k_3, k_3) = \frac{A_s}{k_1^3 k_3^3}\left(\frac{k_1}{k_3}\right)^{3/2} \cos\left[\mu_s \ln\left(\frac{k_1}{k_3}\right) + \phi_s\right], \quad \text{(C.69)}$$

where the amplitude and the phase are given by

$$A_s = |\tilde{A}_s| \times \begin{cases} f^{(s)}(1) \times \dfrac{\sqrt{\pi}}{2^{2s-2}\Gamma(s)} \dfrac{\Gamma(\frac{1}{2}+s-i\mu_s)\Gamma(\frac{1}{2}+s+i\mu_s)}{(s-\frac{3}{2})^2+\mu_s^2} \propto e^{-\pi\mu_s} & c_\pi = 1 \\ f^{(s)}(0) \times \dfrac{\Gamma[\frac{1}{2}(\frac{1}{2}+s+i\mu_s)]\Gamma[\frac{1}{2}(\frac{1}{2}+s-i\mu_s)]}{(s-\frac{3}{2})^2+\mu_s^2} \propto e^{-\pi\mu_s/2} & c_\pi \ll 1 \end{cases}, \quad \text{(C.70)}$$

$$\phi_s \equiv \arg\tilde{A}_s - \mu_s \ln 4c_\pi\,, \quad \text{(C.71)}$$

with

$$\tilde{A}_s \equiv \frac{i^s\pi^3 s!}{8(2s-1)!!}\frac{(5+2s+2i\mu_s)(1+i\sinh\pi\mu_s)}{\cosh\pi\mu_s}\frac{\Gamma(-i\mu_s)}{\Gamma(\frac{1}{2}-i\mu_s)}\,, \quad \text{(C.72)}$$

for even spins, whereas the result for odd spins is given by replacing $1 + i\sinh\pi\mu_s \to i\cosh\pi\mu_s$.

The final answer is then obtained by summing the permutations $(\mathbf{k}_2 \leftrightarrow \mathbf{k}_3)$ in (C.50). Momentum conservation implies

$$\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_2 = -\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3 - \frac{k_1}{k_3}\left[1 - (\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3)^2\right] + \mathcal{O}(k_1^2/k_3^2)\,. \tag{C.73}$$

Writing the spin as $s = 2\ell + 1$ for odd spins, with ℓ an integer, we get

$$\begin{aligned} &P_{2\ell+1}(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3) + (\mathbf{k}_2 \leftrightarrow \mathbf{k}_3) \\ &\qquad = -(2\ell+1)\frac{k_1}{k_3}\left[P_{2\ell}(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3) - (\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3)P_{2\ell+1}(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3)\right] + \mathcal{O}(k_1^2/k_3^2)\,. \end{aligned} \tag{C.74}$$

For odd spins, the leading terms cancel in the sum over the two permutations, and the squeezed limit scales as $(k_1/k_3)^{5/2}$. Note that the right-hand side of (C.74) is an even-degree polynomial of the angle. For even spin $s = 2\ell$, we have instead (see also [262])

$$\begin{aligned} &P_{2\ell}(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3) + (\mathbf{k}_2 \leftrightarrow \mathbf{k}_3) = \\ &\qquad = 2P_{2\ell}(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3) - 2\ell\frac{k_1}{k_3}\left[P_{2\ell-1}(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3) - (\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3)P_{2\ell}(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3)\right] + \mathcal{O}(k_1^2/k_3^2)\,, \end{aligned} \tag{C.75}$$

where the leading terms add up and thus scale as $(k_1/k_3)^{3/2}$. The next-to-leading term at order $(k_1/k_3)^{5/2}$ is an odd-degree polynomial of the angle.

$\langle\gamma\zeta\zeta\rangle$. We also studied the tensor-scalar-scalar bispectrum $\langle\gamma\zeta\zeta\rangle$. Its squeezed limit can be written as

$$\lim_{k_1 \ll k_3} \frac{\langle \gamma^\lambda_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3}\rangle'}{\Delta_\gamma \Delta_\zeta^3} = \alpha_s \sqrt{r}\Delta_\zeta^{-1}\, \mathcal{E}_2^\lambda(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3)\hat{P}_s^\lambda(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3)\mathcal{B}^{(s)}(\mu_s, c_\pi, k_1, k_3, k_3) + (\mathbf{k}_2 \leftrightarrow \mathbf{k}_3)\,, \tag{C.76}$$

where

$$\mathcal{B}^{(s)} \equiv -\frac{\pi^3 c_\pi^{s-3/2}}{8}\sum_{\pm\pm}(\pm i k_1^{s-3})(\pm i c_\pi^2 k_3^{s-4})\,\mathcal{B}_{\pm\pm}\,, \tag{C.77}$$

$$\mathcal{B}^{(s)}_{\pm\pm} \equiv \int_{-\infty}^0 \frac{\mathrm{d}\eta}{\eta^{2-2s}}(1 \mp i c_\pi k_3 \eta)e^{\pm 2 i c_\pi k_3 \eta}\int_{-\infty}^0 \mathrm{d}\tilde{\eta}\, e^{\pm i k_1 \tilde{\eta}}\,\tilde{G}_{\pm\pm}(k_1, \eta, \tilde{\eta})\,, \tag{C.78}$$

with $\tilde{G}_{\pm\pm}$ the Green's functions for the helicity-± 2 mode, $\sigma^{\pm 2}_{s,s}$. Following the same steps as in the scalar case, we obtain the following factorized form of the integrals

$$\mathcal{B}^{(s)} = \tilde{N}^{(s)}\sum_{\pm}(\pm k_1^{s-3}k_3^{s-4})\tilde{\mathcal{P}}_\pm(k_1, c_\pi k_3)\tilde{\mathcal{Q}}^*(k_1) + c.c.\,, \tag{C.79}$$

where

$$\tilde{N}^{(s)} \equiv -\frac{9\pi^4 c_\pi^{s+1/2}}{32}\frac{(s-2)!s!}{(s+2)!(2s-1)!!}\frac{\Gamma(\frac{5}{2} - i\mu_s)\Gamma(\frac{5}{2} + i\mu_s)}{\Gamma(\frac{1}{2} + s - i\mu_s)\Gamma(\frac{1}{2} + s + i\mu_s)}\,, \tag{C.80}$$

$$\tilde{\mathcal{P}}_{\pm}(k_1, c_\pi k_3) \equiv e^{-\pi\mu_s/2}\int_0^\infty \mathrm{d}x\, x^{s-1/2}(1 \mp i c_\pi k_3 x)\tilde{G}^{(s)}_{i\mu_s}(k_1 x)e^{\pm 2ic_\pi k_3 x}\,, \tag{C.81}$$

$$\tilde{\mathcal{Q}}^*(k_1) \equiv e^{-\pi\mu_s/2}\int_0^\infty \mathrm{d}x\, x^{s-5/2}\tilde{G}^{(s)*}_{i\mu_s}(k_1 x)e^{-ik_1 x}\,, \tag{C.82}$$

with $\tilde{G}^{(s)}_{i\mu_s}$ defined in (C.7). In the squeezed limit, $k_1 \ll c_\pi k_3$, the integral (C.81) becomes

$$\begin{aligned}\lim_{k_1\ll c_\pi k_3}\tilde{\mathcal{P}}^{(s)}_{\pm}(k_1, c_\pi k_3) = {}& \frac{i^{3/2}e^{(1\mp 1)\pi\mu_s/2}\Gamma(\frac{1}{2}+s+i\mu_s)\Gamma(\frac{1}{2}+s-i\mu_s)}{4\pi(\pm 2c_\pi k_3)^{1/2+s}} \\ & \times\left(\frac{k_1}{4c_\pi k_3}\right)^{i\mu_s}(5+2s+2i\mu_s)\frac{\Gamma(-i\mu_s)}{\Gamma(\frac{5}{2}-i\mu_s)} \mp i e^{-(1\mp 1)\pi\mu_s}\times c.c.\,.\end{aligned} \tag{C.83}$$

The integral (C.82) is given by

$$\tilde{Q}^{(s)*}(k_1) = \tilde{f}^{(s)}\times\frac{2i(2ik_1)^{3/2-s}}{\sqrt{\pi}\Gamma(s-1)}\frac{\Gamma(\frac{1}{2}+s-i\mu_2)\Gamma(\frac{1}{2}+s-i\mu_2)}{\big((s-\frac{3}{2})^2+\mu_s^2\big)\big((s-\frac{1}{2})^2+\mu_s^2\big)}\,, \tag{C.84}$$

where $\tilde{f}^{(s)}$ a polynomial of μ_s that encodes the difference between evaluating the integral with $\tilde{G}^{(s)}_{i\mu_s}$ and $H_{i\mu_s}$. For spin-2, this is simply $\tilde{f}^{(2)} = 1$. The bispectrum is then given by

$$\mathcal{B}^{(s)}(\mu_s, c_\pi, k_1, k_3, k_3) = \frac{|B_s|}{k_1^3 k_3^3}\left(\frac{k_1}{k_3}\right)^{3/2}\cos\left[\mu_s\ln\left(\frac{k_1}{k_3}\right)+\tilde{\phi}_s\right], \tag{C.85}$$

where

$$B_s \equiv -\tilde{f}^{(s)}\times\frac{9i^s\pi^{7/2}}{2^{2s+4}}\frac{(5+2s+2i\mu_s)(1+i\sinh\pi\mu_s)}{(s+1)(s+2)(2s-1)!!\cosh\pi\mu_s}\frac{\Gamma(-\frac{3}{2}+s-i\mu_s)\Gamma(-\frac{3}{2}+s+i\mu_s)}{\Gamma(-i\mu_s)^{-1}\Gamma(-\frac{3}{2}-i\mu_s)}\,, \tag{C.86}$$

$$\tilde{\phi}_s \equiv \arg B_s - \mu_s\ln 4c_\pi\,, \tag{C.87}$$

for even spins. The result for odd spins requires the replacement $1+i\sinh\pi\mu_s \to \cosh\pi\mu_s$. The final bispectrum is then obtained by summing over the permutations ($\mathbf{k}_2 \leftrightarrow \mathbf{k}_3$).

D. Tensors and conformal symmetry

In this appendix, we provide supplementary details of the analysis presented in Chapter 6. In §D.1, we make a few remarks on the breaking of conformal symmetry in theories with a non-trivial sound speed. In §D.2, we review the argument of [153], showing that a non-trivial tensor spectrum can be mapped to a non-trivial scalar spectrum by a disformal transformation [154].

D.1 Breaking conformal symmetry

We will show that dilatations and special conformal transformations are broken by an amount controlled by $\varepsilon = -\dot{H}/H^2$. Moreover, we will demonstrate that in theories with non-trivial sound speed, $c_s \ll 1$, special conformal symmetry is broken even in the limit $\varepsilon \to 0$.

- First, let us consider a massless scalar field φ in pure de Sitter space

$$S_{\rm dS} = \frac{1}{2}\int \mathrm{d}^4x\, \frac{\varphi'^2 - (\nabla\varphi)^2}{H^2\eta^2}\,. \tag{D.1}$$

  Conformal transformations (3.25) act on φ as follows:

$$\delta_\lambda\varphi = \lambda\left(\eta\varphi' + \mathbf{x}\cdot\nabla\varphi\right)\,, \tag{D.2}$$

$$\delta_b\varphi = 2\,\mathbf{b}\cdot\mathbf{x}\left(\eta\varphi' + \mathbf{x}\cdot\nabla\varphi\right) + (\eta^2 - x^2)\,\mathbf{b}\cdot\nabla\varphi\,. \tag{D.3}$$

  It is straightforward to check that the action (D.1) is invariant under these transformations.

- Next, let us repeat the analysis for an inflationary background. The action of a massless scalar in quasi-de Sitter space is

$$S_{\rm inf} = \frac{1}{2}\int \mathrm{d}^4x\, a^2(\eta)\left[\varphi'^2 - (\nabla\varphi)^2\right] \equiv \int \mathrm{d}^4x\, \mathcal{L}_{\rm inf}\,, \qquad a(\eta) = -\frac{1}{H\eta(1-\varepsilon)}\,. \tag{D.4}$$

  After integrations by parts, we obtain the following variations of the action

$$\delta_\lambda S_{\rm inf} = 2\lambda\int \mathrm{d}^4x\, \varepsilon\, \mathcal{L}_{\rm inf}\,, \tag{D.5}$$

$$\delta_b S_{\rm inf} = 2\int \mathrm{d}^4x\, (\mathbf{b}\cdot\mathbf{x})\, \varepsilon\, \mathcal{L}_{\rm inf}\,. \tag{D.6}$$

  As advertised, dilatations and SCTs are broken by an amount proportional to ε.

- Finally, we consider a massless scalar field with a non-trivial speed of sound in a de Sitter background

$$S_{c_s} = \frac{1}{2}\int \mathrm{d}^4x\, \frac{1}{c_s^2}\frac{\varphi'^2 - c_s^2(\nabla\varphi)^2}{H^2\eta^2}\,. \tag{D.7}$$

  Assuming c_s is a constant for simplicity, the variation of the action gives

$$\delta_\lambda S_{c_s} = 0\,, \tag{D.8}$$

$$\delta_b S_{c_s} = 2\int \mathrm{d}^4x\, \frac{1-c_s^2}{c_s^2}\,\frac{\eta\varphi'\,\mathbf{b}\cdot\nabla\varphi}{H^2\eta^2}\,. \tag{D.9}$$

  We see that scale invariance is retained, while special conformal invariance is broken. For time-dependent $c_s(t)$, dilatations would be broken as well.[1]

D.2 Disformal transformation

In [153], it was shown that a non-trivial tensor sound speed can be set to unity by a disformal transformation [154]. This is followed by a conformal transformation, which brings the action back to Einstein frame. The combined transformation is given by

$$g_{\mu\nu} \;\to\; c_t^{-1}\left[g_{\mu\nu} + (1-c_t^2)\,n_\mu n_\nu\right], \tag{D.10}$$

where $n_\mu \propto \partial_\mu\phi$ is the unit vector orthogonal to the constant-time hypersurfaces. The action in the new frame then has a trivial sound speed for tensors, $\tilde{c}_t = 1$, but a non-trivial sound speed for scalars, $\tilde{c}_s = c_t^{-1}$. In this section, we show that observables are the same in both frames. In particular, we will find that the modification to the consistency condition (6.15) is still present in the new frame.

Consider the action (6.1) in comoving gauge. At quadratic order in fluctuations and at leading order in slow-roll, we get

$$S = \frac{M_{\rm pl}^2}{8}\int \mathrm{d}t\,\mathrm{d}^3x\,\left(\mathcal{L}_\zeta + \mathcal{L}_\gamma\right), \qquad \text{with} \qquad \begin{aligned}\mathcal{L}_\zeta &= 8a^3\varepsilon\left[\dot{\zeta}^2 - a^{-2}(\nabla\zeta)^2\right],\\ \mathcal{L}_\gamma &= a^3 c_t^{-2}\left[\dot{\gamma}_{ij}^2 - a^{-2}c_t^2(\nabla\gamma_{ij})^2\right].\end{aligned} \tag{D.11}$$

After performing the disformal transformation (D.10), the background line element becomes $\mathrm{d}s^2 = -c_t\,\mathrm{d}t^2 + c_t^{-1}a^2\mathrm{d}\mathbf{x}^2$. Rescaling the time and the scale factor,

$$\mathrm{d}\tilde{t} = c_t^{1/2}(t)\,\mathrm{d}t\,, \qquad \tilde{a}(\tilde{t}) = c_t^{-1/2}(t)\,a(t)\,, \tag{D.12}$$

we get $\mathrm{d}s^2 = -\mathrm{d}\tilde{t}^2 + \tilde{a}^2\mathrm{d}\mathbf{x}^2$. The curvature perturbation ζ and the tensor fluctuations γ_{ij}

[1] Moreover, one can show that scale invariance remains unbroken by the inclusion of higher-order interactions that break Lorentz invariance, such as φ'^3 and $\varphi'(\nabla\varphi)^2$.

transform as spacetime scalars, so the action (D.11) takes the form

$$S = \frac{1}{8}M_{\rm pl}^2 \int \mathrm{d}\tilde{t}\,\mathrm{d}^3x \left(\tilde{\mathcal{L}}_{\tilde{\zeta}} + \tilde{\mathcal{L}}_{\tilde{\gamma}}\right), \qquad \text{with} \qquad \begin{aligned} \tilde{\mathcal{L}}_{\tilde{\zeta}} &= 8\tilde{a}^3\varepsilon\left[c_t^2(\partial_{\tilde{t}}\tilde{\zeta})^2 - \tilde{a}^{-2}(\nabla\tilde{\zeta})^2\right], \\ \tilde{\mathcal{L}}_{\tilde{\gamma}} &= \tilde{a}^3\left[(\partial_{\tilde{t}}\tilde{\gamma}_{ij})^2 - \tilde{a}^{-2}(\nabla\tilde{\gamma}_{ij})^2\right]. \end{aligned} \tag{D.13}$$

Hence, in the new frame, the tensors propagate with a trivial sound speed, $\tilde{c}_t = 1$, but the scalars have a modified sound speed, $\tilde{c}_s = c_t^{-1}$. Notice that $c_t < 1$ implies $\tilde{c}_s > 1$. It is not unusual that a non-local field redefinition maps a purely luminal theory to one with apparent superluminality (e.g. [263–265]). In such a situation, the presence of a superluminal mode does not imply a violation of relativistic causality.

At leading order in slow-roll, the tensor power spectrum takes the standard form

$$\Delta_\gamma^2 = \frac{2}{\pi^2}\frac{\tilde{H}^2}{M_{\rm pl}^2}, \tag{D.14}$$

in terms of the new Hubble parameter $\tilde{H} \equiv \partial_{\tilde{t}} \ln \tilde{a} \approx c_t^{-1/2} H$. The tensor tilt is hence also of the usual form, $n_t = -2\tilde{\varepsilon}$, and all non-trivial features have been moved to the scalar sector. The power spectrum of curvature perturbations is

$$\Delta_\zeta^2 = \frac{1}{8\pi^2}\frac{1}{\varepsilon\tilde{c}_s}\frac{\tilde{H}^2}{M_{\rm pl}^2}, \tag{D.15}$$

where

$$\varepsilon = \tilde{\varepsilon} + \frac{1}{2}\tilde{\varepsilon}_s, \quad \text{with} \quad \tilde{\varepsilon}_s = -\varepsilon_t. \tag{D.16}$$

If we neglect the small shift in the amplitude due to $\tilde{c}_s \approx 1$, then the tensor-to-scalar ratio is

$$r \approx 16\left[\tilde{\varepsilon} + \frac{1}{2}\tilde{\varepsilon}_s\right]. \tag{D.17}$$

Hence, although the tensor tilt is standard in the new frame, the tensor-to-scalar ratio now is non-standard. The tensor consistency condition is then given by

$$-\frac{8n_t}{r} = \frac{\tilde{\varepsilon}}{\tilde{\varepsilon} + \frac{1}{2}\tilde{\varepsilon}_s}\left(= 1 + \frac{1}{2}\frac{\varepsilon_t}{\varepsilon}\right). \tag{D.18}$$

We see that the consistency condition is still modified in the new frame, but now the effect comes from the time dependence of a scalar sound speed, $\tilde{\varepsilon}_s \neq 1$. Substituting $\tilde{\varepsilon}$ and $\tilde{\varepsilon}_s$ in terms of the parameters in the original frame, ε and ε_t, we find complete agreement with our previous result (6.15).

E. Details of the superhorizon test

In this appendix, we provide supplementary material to Chapter 7. In §E.1, we provide details of a similar analysis in harmonic space. In §E.2, we present the derivation of the effective noise in multi-frequency experiments. The instrumental specifications of the CMB experiments considered in Chapter 7 are listed in §E.3.

E.1 Analysis in harmonic space

The analysis in §7.4 and §7.5 was presented mostly in real space. In this section, we give a few details of an equivalent formulation in harmonic space.

Superhorizon estimator. Transforming (7.32) to harmonic space, we obtain an estimator of the superhorizon part of the $\mathcal{B}$-mode power spectrum

$$\widehat{\mathcal{S}}_\ell = \sum_{\ell'} M_{\ell\ell'} \widehat{\mathcal{C}}_{\ell'} \ , \tag{E.1}$$

where $M_{\ell\ell'}$ denotes a generalization of the kernel (7.12) to the interval $\Theta = [\theta_{\rm min}, \theta_{\rm max}]$,

$$M_{\ell\ell'} \equiv \frac{2\ell'+1}{2} \underbrace{\int_{\cos\theta_{\rm max}}^{\cos\theta_{\rm min}} P_\ell(x) P_{\ell'}(x)\, {\rm d}x}_{\equiv I_{\ell\ell'}} \ . \tag{E.2}$$

The off-diagonal terms of $I_{\ell\ell'}$ are given by

$$I_{\ell\ell'} = \left[\frac{(\ell-\ell') x P_\ell P_{\ell'} + \ell' P_\ell P_{\ell'-1} - \ell P_{\ell-1} P_{\ell'}}{\ell(\ell+1) - \ell'(\ell'+1)} \right]_{\cos\theta_{\rm max}}^{\cos\theta_{\rm min}} \ , \tag{E.3}$$

while the diagonal terms still obey the recursion relation (7.14). The covariance matrix of the estimator (E.1) is then given by

$$\mathscr{C}[\widehat{\mathcal{S}}_\ell, \widehat{\mathcal{S}}_{\ell'}] = \sum_{ll'} \mathscr{C}[\widehat{\mathcal{C}}_l, \widehat{\mathcal{C}}_{l'}]\, M_{\ell l} M_{\ell' l'} \ , \tag{E.4}$$

where $\mathscr{C}[\widehat{\mathcal{C}}_l, \widehat{\mathcal{C}}_{l'}]$ was given in (7.31). Figure E.1 shows the filtered superhorizon and subhorizon $\mathcal{B}$-mode spectra projected onto the interval $[2.6°, 6.0°]$.

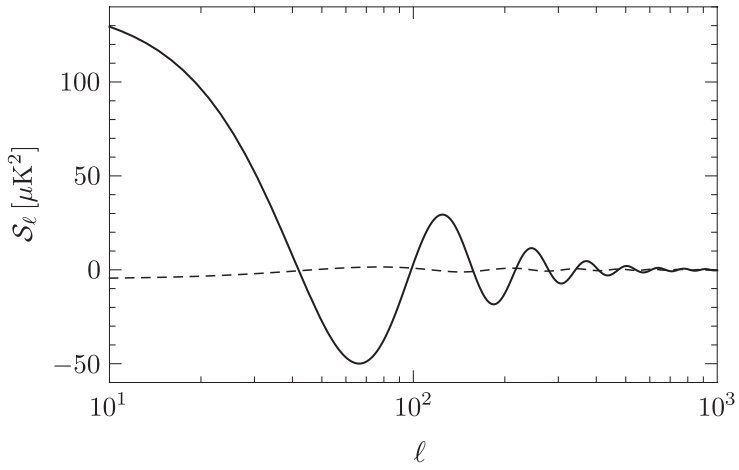


Figure E.1: Local $\mathcal{B}$-mode power spectrum for $r = 0.13$ using the Gaussian filter (7.20) with $\ell_s = 200$ projected onto the interval $[2.6°, 6.0°]$. The solid and dashed lines correspond to the superhorizon and subhorizon modes, respectively.

Signal-to-noise. We define the binned signal as

$$\widehat{\mathcal{S}}_b \equiv \sum_\ell B_{b\ell}\,\widehat{\mathcal{S}}_\ell \ , \tag{E.5}$$

where $B_{b\ell}$ is a binning matrix with uniform weight:

$$B_{b\ell} \equiv \begin{cases} (\ell_{(b+1)} - \ell_{(b)})^{-1} & \ell_{(b)} \le \ell < \ell_{(b+1)} \\ 0 & \text{otherwise} \end{cases} \ . \tag{E.6}$$

The binned covariance matrix is given by

$$\mathscr{C}_{bb'} \equiv \sum_{\ell\ell'} \mathscr{C}[\widehat{\mathcal{S}}_\ell, \widehat{\mathcal{S}}_{\ell'}]\, B_{b\ell} B_{\ell' b'} \ , \tag{E.7}$$

and the signal-to-noise is

$$(\mathrm{S/N})^2 = \sum_{bb'} \widehat{\mathcal{S}}_b \mathscr{C}^{-1}_{bb'} \widehat{\mathcal{S}}_{b'} \ , \tag{E.8}$$

where $\mathscr{C}^{-1}_{bb'}$ is the inverse of (E.7). As in the real space treatment, one has to choose the binning sensibly in order to sample the signal and the covariance well. A natural bandwidth in this case is $\Delta\ell \simeq 180°/(\theta_{\max} - \theta_{\min})$.

We have computed the signal-to-noise (E.8) and compared it with the real space results quoted in §7.5. Since both treatments produce very similar results, we have chosen only to present the real space analysis in the main text. The agreement is expected as (E.1) is an exact harmonic counterpart of the estimator (7.32). A slight difference arises from the choice of binning, since uniform binning in real space does not correspond to uniform binning in harmonic space.

E.2 Multi-frequency effective noise

Observations of the CMB anisotropies at multiple frequencies allow for foreground cleaning because the foreground contaminations have spectral distributions that are different from the Planck spectrum of the primordial CMB signal. In general, the different frequency channels have different noise power spectra, and the effective noise level of a multi-frequency experiment is given by taking a weighted combination which minimizes the variance [200]. In this section, we assume that foreground cleaning has been performed down to a given level, and derive the effective noise for the combined foreground-cleaned CMB map.

The harmonic coefficients of a CMB map measured at frequencies ν_i can be written as

$$a_{i,\ell m} = a^{\rm CMB}_{i,\ell m} + a^{\rm R}_{i,\ell m} + a^{\rm N}_{i,\ell m} \ , \tag{E.9}$$

where $a^{\rm CMB}_{i,\ell m}$ denotes the sum of the primary CMB and the lensing-induced signal (which are both frequency-independent in thermodynamic temperature units), while $a^{\rm R}_{i,\ell m}$ and $a^{\rm N}_{i,\ell m}$ stand for the foreground residuals and instrumental noise, respectively. We assume that the CMB signals, the foreground residuals, and the instrumental noise are uncorrelated, i.e. for any frequency channels i and j, we have

$$\langle a^{\rm CMB}_{i,\ell m} a^{\rm N}_{j,\ell m}\rangle = \langle a^{\rm CMB}_{i,\ell m} a^{\rm R}_{j,\ell m}\rangle = \langle a^{\rm N}_{i,\ell m} a^{\rm R}_{j,\ell m}\rangle = 0 \ . \tag{E.10}$$

Moreover, we assume that instrumental noise at different channels are uncorrelated, so that the noise cross-power spectrum is defined as

$$\langle a^{\rm N}_{i,\ell m} a^{\rm N*}_{j,\ell' m'}\rangle = N_{i,\ell}\delta_{\ell\ell'}\delta_{mm'}\delta_{ij} \ . \tag{E.11}$$

Although foregrounds are in general correlated among different channels—a fact that is exploited in component separation methods for foreground subtraction—we treat the foreground *residuals* as an extra source of *uncorrelated* noise. Thus, we have

$$\langle a^{\rm R}_{i,\ell m} a^{\rm R*}_{j,\ell' m'}\rangle = R_{i,\ell}\delta_{\ell\ell'}\delta_{mm'}\delta_{ij} \ . \tag{E.12}$$

Defining an estimator of the cross-power spectrum of the primary CMB as

$$\widehat{C}_{ij,\ell} \equiv \sum_m \frac{a_{i,\ell m} a^*_{j,\ell m}}{2\ell+1} - L_\ell - \delta_{ij}(N_{i,\ell} + R_{i,\ell}) \ , \tag{E.13}$$

the covariance matrix is

$$\mathscr{C}[\widehat{C}_{ij,\ell}, \widehat{C}_{i'j',\ell'}] = \frac{\delta_{\ell\ell'}}{2\ell+1}\Big\{\big[C_\ell + L_\ell + \delta_{ii'}(N_{i,\ell}+R_{i,\ell})\big]\big[C_\ell + L_\ell + \delta_{jj'}(N_{j,\ell}+R_{j,\ell})\big] + i' \leftrightarrow j'\Big\}. \tag{E.14}$$

The estimator of the power spectrum of a linearly combined foreground-cleaned map can then be

expressed as

$$\widehat{\mathbb{C}}_\ell \equiv \frac{1}{Z_\ell} \sum_{ij} \omega_{ij,\ell} \widehat{C}_{ij,\ell} \ , \tag{E.15}$$

where $Z_\ell \equiv \sum_{ij} \omega_{ij,\ell}$. The optimal weights $\omega_{ij,\ell}$ are determined by minimizing the variance of $\widehat{\mathbb{C}}_\ell$. A straightforward computation leads to

$$\omega_{ij,\ell} = \frac{1}{(N_{i,\ell} + R_{i,\ell})(N_{j,\ell} + R_{j,\ell})} \ . \tag{E.16}$$

The minimum variance of the combined CMB map is then

$$\mathrm{Var}[\widehat{\mathbb{C}}_\ell] = \frac{2}{2\ell+1} \left(C_\ell + L_\ell + N_\ell^{\mathrm{eff}}\right)^2 \ , \tag{E.17}$$

where the effective noise power spectrum is defined as

$$N_\ell^{\mathrm{eff}} \equiv \frac{1}{\sqrt{Z_\ell}} = \left[\sum_i \frac{1}{N_{i,\ell} + R_{i,\ell}}\right]^{-1} \ . \tag{E.18}$$

This recovers eq. (117) of [199].

E.3 Experimental specifications

	$f_{\rm sky}$ [%]	ν [GHz]	θ_b [′]	$N_{\rm det}$	Δ_P [μK′]	$\Delta_{P,\rm eff}$ [μK′]	$s_{P,\rm eff}$ [μK′]
BICEP2	2.4	150	29	512	5.2		
Keck Array		95	29	576	9.0		
		150	29	2560	2.3		
		220	29	1536	10.2		
BICEP3		95	29	2560	2.0	1.4	9.0
SPTPol	6	90	1.6	360	9.0		
		150	1.0	1176	5.0	4.4	17.8
PolarBeaR-2	20	95	5.2	3794	15.1		
		150	3.5	3794	15.1	10.7	23.9
Simons Array	20	95	5.2	7588	10.7		
		150	3.5	11382	8.7		
		220	2.7	3794	16.7	6.3	14.1
LiteBIRD	70	60	32	304	10.3		
		78	58	304	6.5		
		100	45	304	4.7		
		140	32	370	3.7		
		195	24	370	3.1		
		280	16	370	3.8	1.8	2.2
COrE	70	45	23.3	64	9.1		
		75	14.0	300	4.7		
		105	10.0	400	4.6		
		135	7.8	550	4.6		
		165	6.4	750	4.6		
		195	5.4	1150	4.5		
		225	4.7	1800	4.6		
		255	4.1	575	10.5		
		285	3.7	375	17.4		
		315	3.3	100	46.6		
		375	2.8	64	19.0		
		435	2.4	64	258.0		
		555	1.9	64	626.0		
		675	1.6	64	3640.0		
		795	1.3	64	22200.0	1.8	2.2

Table E.1: Instrumental specifications for current and planned CMB polarization experiments [190–196].

Bibliography

[1] D. Baumann, D. Green, H. Lee, and R. A. Porto, "Signs of Analyticity in Single-Field Inflation," arXiv:1502.07304 [hep-th].

[2] D. Baumann, H. Lee, and G. Pimentel, "High-Scale Inflation and the Tensor Tilt," *JHEP* **01** (2016) 101, arXiv:1507.07250 [hep-th].

[3] H. Lee, D. Baumann, and G. L. Pimentel, "Non-Gaussianity as a Particle Detector," *JHEP* **12** (2016) 040, arXiv:1607.03735 [hep-th].

[4] H. Lee, S. C. Su, and D. Baumann, "The Superhorizon Test of Future B-mode Experiments," *JCAP* **1502** no. 02, (2015) 036, arXiv:1408.6709 [astro-ph.CO].

[5] A. Einstein, "The Foundation of the General Theory of Relativity," *Annalen Phys.* **49** (1916) 769–822. [Annalen Phys.14,517(2005)].

[6] A. Friedmann, "Über die Krümmung des Raumes," *Zeitschrift fur Physik* **10** (1922) 377–386.

[7] G. Lemaitre, "A Homogeneous Universe of Constant Mass and Growing Radius Accounting for the Radial Velocity of Extragalactic Nebulae," *Annales Soc. Sci. Brux. Ser. I Sci. Math. Astron. Phys.* **A47** (1927) 49–59.

[8] E. Hubble, "A relation between distance and radial velocity among extra-galactic nebulae," *Proc. Nat. Acad. Sci.* **15** (1929) 168–173.

[9] R. A. Alpher, H. Bethe, and G. Gamow, "The origin of chemical elements," *Phys. Rev.* **73** (1948) 803–804.

[10] A. A. Penzias and R. W. Wilson, "A Measurement of excess antenna temperature at 4080-Mc/s," *Astrophys. J.* **142** (1965) 419–421.

[11] **COBE** Collaboration, G. F. Smoot *et al.*, "Structure in the COBE differential microwave radiometer first year maps," *Astrophys. J.* **396** (1992) L1–L5.

[12] E. Komatsu and D. N. Spergel, "Acoustic signatures in the primary microwave background bispectrum," *Phys. Rev.* **D63** (2001) 063002, arXiv:astro-ph/0005036 [astro-ph].

[13] **Planck** Collaboration, R. Adam *et al.*, "Planck 2015 results. I. Overview of products and scientific results," *Astron. Astrophys.* **594** (2016) A1, arXiv:1502.01582 [astro-ph.CO].

[14] **Supernova Search Team** Collaboration, A. G. Riess *et al.*, "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *Astron. J.* **116** (1998) 1009–1038, arXiv:astro-ph/9805201 [astro-ph].

[15] **Supernova Cosmology Project** Collaboration, S. Perlmutter *et al.*, "Measurements of Omega and Lambda from 42 high redshift supernovae," *Astrophys. J.* **517** (1999) 565–586, arXiv:astro-ph/9812133 [astro-ph].

[16] A. H. Guth, "The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems," *Phys. Rev.* **D23** (1981) 347–356.

[17] **CMS** Collaboration, S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett.* **B716** (2012) 30–61, arXiv:1207.7235 [hep-ex].

[18] A. D. Linde, "A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems," *Phys. Lett.* **B108** (1982) 389–393.

[19] A. Albrecht and P. J. Steinhardt, "Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking," *Phys. Rev. Lett.* **48** (1982) 1220–1223.

[20] A. D. Linde, "Chaotic Inflation," *Phys. Lett.* **B129** (1983) 177–181.

[21] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck 2015 results. XIII. Cosmological parameters," *Astron. Astrophys.* **594** (2016) A13, arXiv:1502.01589 [astro-ph.CO].

[22] N. A. Chernikov and E. A. Tagirov, "Quantum theory of scalar fields in de Sitter space-time," *Ann. Inst. H. Poincare Phys. Theor.* **A9** (1968) 109.

[23] T. S. Bunch and P. C. W. Davies, "Quantum Field Theory in de Sitter Space: Renormalization by Point Splitting," *Proc. Roy. Soc. Lond.* **A360** (1978) 117–134.

[24] R. L. Arnowitt, S. Deser, and C. W. Misner, "The Dynamics of General Relativity," *Gen. Rel. Grav.* **40** (2008) 1997–2027, arXiv:gr-qc/0405109 [gr-qc].

[25] J. M. Maldacena, "Non-Gaussian Features of Primordial Fluctuations in Single-Field Inflationary Models," *JHEP* **0305** (2003) 013, arXiv:astro-ph/0210603 [astro-ph].

[26] K. Hinterbichler, L. Hui, and J. Khoury, "An Infinite Set of Ward Identities for Adiabatic Modes in Cosmology," *JCAP* **1401** (2014) 039, arXiv:1304.5527 [hep-th].

[27] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, "Spontaneous Creation of Almost Scale - Free Density Perturbations in an Inflationary Universe," *Phys. Rev.* **D28** (1983) 679.

[28] D. S. Salopek and J. R. Bond, "Nonlinear evolution of long wavelength metric fluctuations in inflationary models," *Phys. Rev.* **D42** (1990) 3936–3962.

[29] P. J. E. Peebles and J. T. Yu, "Primeval adiabatic perturbation in an expanding universe," *Astrophys. J.* **162** (1970) 815–836.

[30] E. R. Harrison, "Fluctuations at the threshold of classical cosmology," *Phys. Rev.* **D1** (1970) 2726–2730.

[31] Ya. B. Zeldovich, "A Hypothesis, unifying the structure and the entropy of the universe," *Mon. Not. Roy. Astron. Soc.* **160** (1972) 1P–3P.

[32] A. A. Starobinsky, "Spectrum of relict gravitational radiation and the early state of the universe," *JETP Lett.* **30** (1979) 682–685. [Pisma Zh. Eksp. Teor. Fiz.30,719(1979)].

[33] D. Fixsen, "The Temperature of the Cosmic Microwave Background," *Astrophys.J.* **707** (2009) 916–920, arXiv:0911.1955 [astro-ph.CO].

[34] S. Dodelson, "Coherent phase argument for inflation," *AIP Conf. Proc.* **689** (2003) 184–196, arXiv:hep-ph/0309057 [hep-ph]. [,184(2003)].

[35] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck 2015 results. XX. Constraints on inflation," *Astron. Astrophys.* **594** (2016) A20, arXiv:1502.02114 [astro-ph.CO].

[36] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, "The Effective Field Theory of Inflation," *JHEP* **0803** (2008) 014, arXiv:0709.0293 [hep-th].

[37] L. Senatore and M. Zaldarriaga, "A Naturally Large Four-Point Function in Single-Field Inflation," *JCAP* **1101** (2011) 003, arXiv:1004.1201 [hep-th].

[38] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, "Limits on non-gaussianities from wmap data," *JCAP* **0605** (2006) 004, arXiv:astro-ph/0509029 [astro-ph].

[39] L. Senatore, K. M. Smith, and M. Zaldarriaga, "Non-Gaussianities in Single Field Inflation and their Optimal Limits from the WMAP 5-year Data," *JCAP* **1001** (2010) 028, arXiv:0905.3746 [astro-ph.CO].

[40] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck 2015 results. XVII. Constraints on primordial non-Gaussianity," *Astron. Astrophys.* **594** (2016) A17, arXiv:1502.01592 [astro-ph.CO].

[41] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, "Second order cosmological perturbations from inflation," *Nucl. Phys.* **B667** (2003) 119–148, arXiv:astro-ph/0209156 [astro-ph].

[42] D. Baumann, D. Green, and R. A. Porto, "B-modes and the Nature of Inflation," arXiv:1407.2621 [hep-th].

[43] **BICEP2, Keck Array** Collaboration, P. A. R. Ade *et al.*, "Improved Constraints on Cosmology and Foregrounds from BICEP2 and Keck Array Cosmic Microwave Background Data with Inclusion of 95 GHz Band," *Phys. Rev. Lett.* **116** (2016) 031302, arXiv:1510.09217 [astro-ph.CO].

[44] **CMB-S4** Collaboration, K. N. Abazajian *et al.*, "CMB-S4 Science Book, First Edition," arXiv:1610.02743 [astro-ph.CO].

[45] **CORE** Collaboration, F. Finelli *et al.*, "Exploring Cosmic Origins with CORE: Inflation," arXiv:1612.08270 [astro-ph.CO].

[46] S. Weinberg, "Effective Field Theory for Inflation," *Phys.Rev.* **D77** (2008) 123541, arXiv:0804.4291 [hep-th].

[47] E. J. Copeland, A. R. Liddle, D. H. Lyth, E. D. Stewart, and D. Wands, "False vacuum inflation with Einstein gravity," *Phys. Rev.* **D49** (1994) 6410–6433, arXiv:astro-ph/9401011 [astro-ph].

[48] R. Kallosh, A. D. Linde, D. A. Linde, and L. Susskind, "Gravity and global symmetries," *Phys. Rev.* **D52** (1995) 912–935, arXiv:hep-th/9502069 [hep-th].

[49] D. Baumann and D. Green, "Inflating with Baryons," *JHEP* **04** (2011) 071, arXiv:1009.3032 [hep-th].

[50] D. H. Lyth, "What would we learn by detecting a gravitational wave signal in the cosmic microwave background anisotropy?," *Phys. Rev. Lett.* **78** (1997) 1861–1863, arXiv:hep-ph/9606387 [hep-ph].

[51] D. Baumann and L. McAllister, "Inflation and String Theory," arXiv:1404.2601 [hep-th].

[52] C. Armendariz-Picon, T. Damour, and V. F. Mukhanov, "k - inflation," *Phys. Lett.* **B458** (1999) 209–218, arXiv:hep-th/9904075 [hep-th].

[53] P. Creminelli, M. A. Luty, A. Nicolis, and L. Senatore, "Starting the Universe: Stable Violation of the Null Energy Condition and Non-Standard Cosmologies," *JHEP* **0612** (2006) 080, arXiv:hep-th/0606090 [hep-th].

[54] F. Piazza and F. Vernizzi, "Effective Field Theory of Cosmological Perturbations," *Class.Quant.Grav.* **30** (2013) 214007, arXiv:1307.4350.

[55] J. Goldstone, "Field Theories with Superconductor Solutions," *Nuovo Cim.* **19** (1961) 154–164.

[56] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries," *Phys. Rev.* **127** (1962) 965–970.

[57] H. Watanabe and H. Murayama, "Unified Description of Nambu-Goldstone Bosons without Lorentz Invariance," *Phys. Rev. Lett.* **108** (2012) 251602, arXiv:1203.0609 [hep-th].

[58] S. Weinberg, *The Quantum Theory of Fields*, vol. 2. Cambridge University Press, Cambridge, 2005.

[59] I. Low and A. V. Manohar, "Spontaneously broken space-time symmetries and Goldstone's theorem," *Phys. Rev. Lett.* **88** (2002) 101602, arXiv:hep-th/0110285 [hep-th].

[60] P. Creminelli, J. Noreña, and M. Simonovic, "Conformal Consistency Relations for Single-Field Inflation," *JCAP* **1207** (2012) 052, arXiv:1203.4595 [hep-th].

[61] K. Hinterbichler, L. Hui, and J. Khoury, "Conformal Symmetries of Adiabatic Modes in Cosmology," *JCAP* **1208** (2012) 017, arXiv:1203.6351 [hep-th].

[62] C. Cheung, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, "On the Consistency Relation of the Three-Point Function in Single-Field Onflation," *JCAP* **0802** (2008) 021, arXiv:0709.0295 [hep-th].

[63] N. Agarwal, R. H. Ribeiro, and R. Holman, "Why does the effective field theory of inflation work?," *JCAP* **1406** (2014) 016, arXiv:1311.0869 [hep-th].

[64] D. Baumann and D. Green, "Equilateral Non-Gaussianity and New Physics on the Horizon," *JCAP* **1109** (2011) 014, arXiv:1102.5343 [hep-th].

[65] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, "Causality, Analyticity and an IR Obstruction to UV Completion," *JHEP* **0610** (2006) 014, arXiv:hep-th/0602178 [hep-th].

[66] A. J. Tolley and M. Wyman, "The Gelaton Scenario: Equilateral Non-Gaussianity from Multi-Field Dynamics," *Phys.Rev.* **D81** (2010) 043502, arXiv:0910.1853 [hep-th].

[67] B. Bellazzini, L. Martucci, and R. Torre, "Symmetries, Sum Rules and Constraints on Effective Field Theories," *JHEP* **1409** (2014) 100, arXiv:1405.2960 [hep-th].

[68] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*. Addison-Wesley, Reading, USA, 1995.

[69] R. J. Eden *et al.*, *The Analytic S-Matrix*. Cambridge University Press, 2002.

[70] S. B. Giddings and R. A. Porto, "The Gravitational S-matrix," *Phys.Rev.* **D81** (2010) 025002, arXiv:0908.0004 [hep-th].

[71] J. Bros, H. Epstein, and V. Glaser, "A Proof of the Crossing Property for Two-Particle Amplitudes in General Quantum Field Theory," *Comm. Math. Phys.* **1** (1965) 240.

[72] M. Froissart, "Asymptotic Behavior and Subtractions in the Mandelstam Representation," *Phys.Rev.* **123** (1961) 1053–1057.

[73] A. Martin, "Unitarity and High-Energy Behavior of Scattering Amplitudes," *Phys.Rev.* **129** (1963) 1432–1436.

[74] Y. Jin and A. Martin, "Number of Subtractions in Fixed-Transfer Dispersion Relations," *Phys.Rev.* **135** (1964) B1375–B1377.

[75] W. D. Goldberger and I. Z. Rothstein, "An Effective field theory of gravity for extended objects," *Phys. Rev.* **D73** (2006) 104029, arXiv:hep-th/0409156 [hep-th].

[76] R. A. Porto, "Post-Newtonian corrections to the motion of spinning bodies in NRGR," *Phys. Rev.* **D73** (2006) 104031, arXiv:gr-qc/0511061 [gr-qc].

[77] P. Creminelli and M. Zaldarriaga, "Single-Field Consistency Relation for the Three-Point Function," *JCAP* **0410** (2004) 006, arXiv:astro-ph/0407059 [astro-ph].

[78] D. Lopez Nacir, R. A. Porto, L. Senatore, and M. Zaldarriaga, "Dissipative Effects in the Effective Field Theory of Inflation," *JHEP* **1201** (2012) 075, arXiv:1109.4192 [hep-th].

[79] D. Lopez Nacir, R. A. Porto, and M. Zaldarriaga, "The Consistency Condition for the Three-Point Function in Dissipative Single-Clock Inflation," *JCAP* **1209** (2012) 004, arXiv:1206.7083 [hep-th].

[80] R. Flauger, D. Green, and R. A. Porto, "On Squeezed Limits in Single-Field Inflation," *JCAP* **1308** (2013) 032, arXiv:1303.1430 [hep-th].

[81] E. Silverstein and D. Tong, "Scalar Speed Limits and Cosmology: Acceleration from D-cceleration," *Phys.Rev.* **D70** (2004) 103505, arXiv:hep-th/0310221 [hep-th].

[82] M. Alishahiha, E. Silverstein, and D. Tong, "DBI in the Sky," *Phys.Rev.* **D70** (2004) 123505, arXiv:hep-th/0404084 [hep-th].

[83] **Planck** Collaboration, P. Ade *et al.*, "Planck 2015 Results. XVII. Constraints on Primordial Non-Gaussianity," arXiv:1502.01592 [astro-ph.CO].

[84] X. Chen and Y. Wang, "Quasi-Single-Field Inflation and Non-Gaussianities," *JCAP* **1004** (2010) 027, arXiv:0911.3380 [hep-th].

[85] A. Achucarro, J.-O. Gong, S. Hardeman, G. A. Palma, and S. P. Patil, "Mass Hierarchies and Non-Decoupling in Multi-Scalar Field Dynamics," *Phys.Rev.* **D84** (2011) 043502, arXiv:1005.3848 [hep-th].

[86] A. Achucarro, J.-O. Gong, S. Hardeman, G. A. Palma, and S. P. Patil, "Effective Theories of Single-Field Inflation when Heavy Fields Matter," *JHEP* **1205** (2012) 066, arXiv:1201.6342 [hep-th].

[87] A. Achucarro *et al.*, "Heavy Fields, Reduced Speeds of Sound and Decoupling during Inflation," *Phys.Rev.* **D86** (2012) 121301, arXiv:1205.0710 [hep-th].

[88] R. Gwyn, G. A. Palma, M. Sakellariadou, and S. Sypsas, "Effective Field Theory of Weakly Coupled Inflationary Models," *JCAP* **1304** (2013) 004, arXiv:1210.3020 [hep-th].

[89] S. Cespedes, V. Atal, and G. A. Palma, "On the Importance of Heavy Fields during Inflation," *JCAP* **1205** (2012) 008, arXiv:1201.4848 [hep-th].

[90] A. Avgoustidis *et al.*, "Decoupling Survives Inflation: A Critical Look at Effective Field Theory Violations During Inflation," *JCAP* **1206** (2012) 025, arXiv:1203.0016 [hep-th].

[91] X. Chen and Y. Wang, "Quasi-Single-Field Inflation with Large Mass," *JCAP* **1209** (2012) 021, arXiv:1205.0160 [hep-th].

[92] V. Assassi, D. Baumann, D. Green, and L. McAllister, "Planck-Suppressed Operators," *JCAP* **1401** no. 01, (2014) 033, arXiv:1304.5226 [hep-th].

[93] A. Achúcarro and Y. Welling, "Multiple Field Inflation and Signatures of Heavy Physics in the CMB," arXiv:1502.04369 [gr-qc].

[94] A. Nicolis, R. Rattazzi, and E. Trincherini, "Energy's and Amplitudes' Positivity," *JHEP* **1005** (2010) 095, arXiv:0912.4258 [hep-th].

[95] S. Roy, "Exact Integral Equation for Pion-Pion Scattering Involving only Physical Region Partial Waves," *Phys.Lett.* **B36** (1971) 353.

[96] K. M. Smith, L. Senatore, and M. Zaldarriaga, "Optimal Analysis of the CMB Trispectrum," arXiv:1502.00635 [astro-ph.CO].

[97] J. Fergusson, D. Regan, and E. Shellard, "Optimal Trispectrum Estimators and WMAP Constraints," arXiv:1012.6039 [astro-ph.CO].

[98] D. Regan, M. Gosenca, and D. Seery, "Constraining the WMAP9 Bispectrum and Trispectrum with Needlets," arXiv:1310.8617 [astro-ph.CO].

[99] L. Parker, "Particle Creation in Expanding Universes," *Physical Review Letters* **21** no. 8, (1968) 562.

[100] L. Parker, "Quantized Fields and Particle Creation in Expanding Universes. I," *Physical Review* **183** no. 5, (1969) 1057.

[101] L. Parker, "Quantized Fields and Particle Creation in Expanding Universes. II," *Physical Review D* **3** no. 2, (1971) 346.

[102] N. Arkani-Hamed and J. Maldacena, "Cosmological Collider Physics," arXiv:1503.08043 [hep-th].

[103] R. Flauger, M. Mirbabayi, L. Senatore, and E. Silverstein, "Productive Interactions: Heavy Particles and Non-Gaussianity," arXiv:1606.00513 [hep-th].

[104] T. Garidi, "What is Mass in de Sitterian Physics?," arXiv:hep-th/0309104 [hep-th].

[105] A. Higuchi, "Forbidden Mass Range for Spin-2 Field Theory in De Sitter Spacetime," *Nucl. Phys.* **B282** (1987) 397.

[106] S. Deser and R. Nepomechie, "Gauge Invariance Versus Masslessness in De Sitter Space," *Annals Phys.* **154** (1984) 396.

[107] L. Singh and C. Hagen, "Lagrangian Formulation for Arbitrary Spin. 1. The Boson Case," *Phys. Rev.* **D9** (1974) 898–909.

[108] L. Singh and C. Hagen, "Lagrangian Formulation for Arbitrary Spin. 2. The Fermion Case," *Phys. Rev.* **D9** (1974) 910–920.

[109] Y. Zinoviev, "On Massive High Spin Particles in AdS," arXiv:hep-th/0108192 [hep-th].

[110] S. Deser and A. Waldron, "Arbitrary Spin Representations in de Sitter from dS/CFT with Applications to dS Supergravity," *Nucl. Phys.* **B662** (2003) 379–392, arXiv:hep-th/0301068 [hep-th].

[111] E. Wigner, "On Unitary Representations of the Inhomogeneous Lorentz Group," *Annals of Mathematics* **40** no. 1, (1939) 149–204. http://www.jstor.org/stable/1968551.

[112] L. Thomas, "On Unitary Representations of the Group of De Sitter Space," *Annals of Mathematics* **42** no. 1, (1941) 113–126. http://www.jstor.org/stable/1968990.

[113] T. Newton, "A Note on the Representations of the De Sitter Group," *Annals of Mathematics* **51** no. 3, (1950) 730–733. http://www.jstor.org/stable/1969376.

[114] S. Deser and A. Waldron, "Gauge Invariances and Phases of Massive Higher Spins in (A)dS," *Phys. Rev. Lett.* **87** (2001) 031601, arXiv:hep-th/0102166 [hep-th].

[115] L. Senatore and M. Zaldarriaga, "The Effective Field Theory of Multifield Inflation," *JHEP* **04** (2012) 024, arXiv:1009.2093 [hep-th].

[116] T. Noumi, M. Yamaguchi, and D. Yokoyama, "EFT Approach to Quasi-Single-Field Inflation and Effects of Heavy Fields," *JHEP* **06** (2013) 051, arXiv:1211.1624 [hep-th].

[117] L. Delacretaz, T. Noumi, and L. Senatore, "Boost Breaking in the EFT of Inflation," arXiv:1512.04100 [hep-th].

[118] D. Baumann and D. Green, "Signatures of Supersymmetry from the Early Universe," *Phys.Rev.* **D85** (2012) 103520, arXiv:1109.0292 [hep-th].

[119] S. Pi and M. Sasaki, "Curvature Perturbation Spectrum in Two-Field Inflation with a Turning Trajectory," *JCAP* **1210** (2012) 051, arXiv:1205.0161 [hep-th].

[120] V. Assassi, D. Baumann, and D. Green, "On Soft Limits of Inflationary Correlation Functions," *JCAP* **1211** (2012) 047, arXiv:1204.4207 [hep-th].

[121] G. Pimentel, "Inflationary Consistency Conditions from a Wavefunctional Perspective," *JHEP* **02** (2014) 124, arXiv:1309.1793 [hep-th].

[122] L. Berezhiani and J. Khoury, "Slavnov-Taylor Identities for Primordial Perturbations," *JCAP* **1402** (2014) 003, arXiv:1309.4461 [hep-th].

[123] D. Binosi and A. Quadri, "The Cosmological Slavnov-Taylor Identity from BRST Symmetry in Single-Field Inflation," *JCAP* **1603** no. 03, (2016) 045, arXiv:1511.09309 [hep-th].

[124] T. Tanaka and Y. Urakawa, "Dominance of Gauge Artifact in the Consistency Relation for the Primordial Bispectrum," *JCAP* **1105** (2011) 014, arXiv:1103.1251 [astro-ph.CO].

[125] E. Pajer, F. Schmidt, and M. Zaldarriaga, "The Observed Squeezed Limit of Cosmological Three-Point Functions," *Phys. Rev.* **D88** no. 8, (2013) 083502, arXiv:1305.0824 [astro-ph.CO].

[126] P. Creminelli, G. D'Amico, M. Musso, and J. Noreña, "The (not so) Squeezed Limit of the Primordial Three-Point Function," *JCAP* **1111** (2011) 038, arXiv:1106.1462 [astro-ph.CO].

[127] M. Mirbabayi and M. Simonović, "Effective Theory of Squeezed Correlation Functions," *JCAP* **1603** no. 03, (2016) 056, arXiv:1507.04755 [hep-th].

[128] E. A. Lim, "Quantum Information of Cosmological Correlations," *Phys. Rev.* **D91** no. 8, (2015) 083522, arXiv:1410.5508 [hep-th].

[129] J. Martin and V. Vennin, "Quantum Discord of Cosmic Inflation: Can We Show That CMB Anisotropies Are of Quantum-Mechanical Origin?," *Phys. Rev.* **D93** no. 2, (2016) 023505, arXiv:1510.04038 [astro-ph.CO].

[130] J. Maldacena, "A Model with Cosmological Bell Inequalities," *Fortsch. Phys.* **64** (2016) 10–23, arXiv:1508.01082 [hep-th].

[131] S. Choudhury, S. Panda, and R. Singh, "Bell Violation in the Sky," arXiv:1607.00237 [hep-th].

[132] J. Liu, C.-M. Sou, and Y. Wang, "Cosmic Decoherence: Massive Fields," arXiv:1608.07909 [hep-th].

[133] N. Kundu, A. Shukla, and S. P. Trivedi, "Constraints from Conformal Symmetry on the Three-Point Scalar Correlator in Inflation," *JHEP* **1504** (2015) 061, arXiv:1410.2606 [hep-th].

[134] S. Giombi, S. Prakash, and X. Yin, "A Note on CFT Correlators in Three Dimensions," *JHEP* **07** (2013) 105, arXiv:1104.4317 [hep-th].

[135] E. Dimastrogiovanni, M. Fasiello, and M. Kamionkowski, "Imprints of Massive Primordial Fields on Large-Scale Structure," *JCAP* **1602** no. 02, (2016) 017, arXiv:1504.05993 [astro-ph.CO].

[136] P. D. Meerburg, J. Meyers, A. van Engelen, and Y. Ali-Haïmoud, "On CMB B-Mode Non-Gaussianity," arXiv:1603.02243 [astro-ph.CO].

[137] K. Abazajian *et al.*, "Neutrino Physics from the Cosmic Microwave Background and Large Scale Structure," arXiv:1309.5383 [astro-ph.CO].

[138] D. Blas, D. Comelli, F. Nesti, and L. Pilo, "Lorentz Breaking Massive Gravity in Curved Space," *Phys. Rev.* **D80** (2009) 044025, arXiv:0905.1699 [hep-th].

[139] L. Bordin, P. Creminelli, M. Mirbabayi, and J. Noreña, "Tensor Squeezed Limits and the Higuchi Bound," arXiv:1605.08424 [astro-ph.CO].

[140] H. Lee *et al.*, *work in progress*.

[141] J. M. Maldacena and G. L. Pimentel, "On Graviton Non-Gaussianities during Inflation," *JHEP* **1109** (2011) 045, arXiv:1104.2846 [hep-th].

[142] J. E. Lidsey *et al.*, "Reconstructing the Inflation Potential : An Overview," *Rev.Mod.Phys.* **69** (1997) 373–410, arXiv:astro-ph/9508078 [astro-ph].

[143] N. Kaloper, M. Kleban, A. E. Lawrence, and S. Shenker, "Signatures of Short Distance Physics in the Cosmic Microwave Background," *Phys.Rev.* **D66** (2002) 123510, arXiv:hep-th/0201158 [hep-th].

[144] A. Lue, L.-M. Wang, and M. Kamionkowski, "Cosmological Signature of New Parity Violating Interactions," *Phys.Rev.Lett.* **83** (1999) 1506–1509, arXiv:astro-ph/9812088 [astro-ph].

[145] S. Alexander and J. Martin, "Birefringent Gravitational Waves and the Consistency Check of Inflation," *Phys.Rev.* **D71** (2005) 063526, arXiv:hep-th/0410230 [hep-th].

[146] C. R. Contaldi, J. Magueijo, and L. Smolin, "Anomalous CMB Polarization and Gravitational Chirality," *Phys.Rev.Lett.* **101** (2008) 141101, arXiv:0806.3082 [astro-ph].

[147] T. Takahashi and J. Soda, "Chiral Primordial Gravitational Waves from a Lifshitz Point," *Phys.Rev.Lett.* **102** (2009) 231301, arXiv:0904.0554 [hep-th].

[148] S. Deser and A. Redlich, "String Induced Gravity and Ghost Freedom," *Phys.Lett.* **B176** (1986) 350.

[149] A. A. Starobinsky, "A New Type of Isotropic Cosmological Models Without Singularity," *Phys. Lett.* **B91** (1980) 99–102.

[150] M. J. Duff, "Twenty Years of the Weyl Anomaly," *Class. Quant. Grav.* **11** (1994) 1387–1404, arXiv:hep-th/9308075 [hep-th].

[151] R. J. Riegert, "A Nonlocal Action for the Trace Anomaly," *Phys. Lett.* **B134** (1984) 56–60.

[152] D. Baumann *et al.*, *work in progress*.

[153] P. Creminelli, J. Gleyzes, J. Noreña, and F. Vernizzi, "Resilience of the Standard Predictions for Primordial Tensor Modes," *Phys.Rev.Lett.* **113** no. 23, (2014) 231301, arXiv:1407.8439 [astro-ph.CO].

[154] J. D. Bekenstein, "The Relation between Physical and Gravitational Geometry," *Phys.Rev.* **D48** (1993) 3641–3647, arXiv:gr-qc/9211017 [gr-qc].

[155] S. Tsujikawa, "Disformal Invariance of Cosmological Perturbations in a Generalized Class of Horndeski Theories," *JCAP* **1504** no. 04, (2015) 043, arXiv:1412.6210 [hep-th].

[156] X. Chen, M.-x. Huang, S. Kachru, and G. Shiu, "Observational Signatures and Non-Gaussianities of General Single-Field Inflation," *JCAP* **0701** (2007) 002, arXiv:hep-th/0605045 [hep-th].

[157] J. J. Atick and E. Witten, "The Hagedorn Transition and the Number of Degrees of Freedom of String Theory," *Nuclear Physics B* **310** no. 2, (1988) 291–334.

[158] J. Caligiuri and A. Kosowsky, "Inflationary Tensor Perturbations After BICEP2," *Phys.Rev.Lett.* **112** (2014) 191302, arXiv:1403.5324 [astro-ph.CO].

[159] S. Dodelson, "How much can we learn about the physics of inflation?," *Phys.Rev.Lett.* **112** (2014) 191301, arXiv:1403.6310 [astro-ph.CO].

[160] T. L. Smith, M. Kamionkowski, and A. Cooray, "Direct Detection of the Inflationary Gravitational Wave Background," *Phys.Rev.* **D73** (2006) 023504, arXiv:astro-ph/0506422 [astro-ph].

[161] T. L. Smith, H. V. Peiris, and A. Cooray, "Deciphering Inflation with Gravitational Waves," *Phys.Rev.* **D73** (2006) 123503, arXiv:astro-ph/0602137 [astro-ph].

[162] S. Chongchitnan and G. Efstathiou, "Prospects for Direct Detection of Primordial Gravitational Waves," *Physical Review D* **73** no. 8, (2006) 083511.

[163] A. Strominger, "The dS/CFT Correspondence," *JHEP* **0110** (2001) 034, arXiv:hep-th/0106113 [hep-th].

[164] A. Zamolodchikov, "Renormalization Group and Perturbation Theory Near Fixed Points in Two-Dimensional Field Theory," *Sov.J.Nucl.Phys.* **46** (1987) 1090.

[165] E. Witten, "Quantum Gravity in de Sitter Space," arXiv:hep-th/0106109 [hep-th].

[166] B. S. DeWitt, "Quantum Theory of Gravity. 1. The Canonical Theory," *Phys.Rev.* **160** (1967) 1113–1148.

[167] J. P. Van Der Schaar, "Inflationary Perturbations from Deformed CFT," *Journal of High Energy Physics* **2004** no. 01, (2004) 070.

[168] H. Osborn and A. Petkou, "Implications of Conformal Invariance in Field Theories for General Dimensions," *Annals Phys.* **231** (1994) 311–362, arXiv:hep-th/9307010 [hep-th].

[169] I. Mata, S. Raju, and S. Trivedi, "CMB from CFT," *JHEP* **1307** (2013) 015, arXiv:1211.5482 [hep-th].

[170] E. S. Fradkin and A. A. Tseytlin, "Conformal Supergravity," *Physics Reports* **119** no. 4, (1985) 233–362.

[171] N. Berkovits and E. Witten, "Conformal Supergravity in Twistor-String Theory," *Journal of High Energy Physics* **2004** no. 08, (2004) 009.

[172] J. Maldacena, "Einstein Gravity from Conformal Gravity," *arXiv:1105.5632* (2011) .

[173] D. Wands, N. Bartolo, S. Matarrese, and A. Riotto, "Observational Test of Two-Field Inflation," *Physical Review D* **66** no. 4, (2002) 043520.

[174] D. Coulson, R. Crittenden, and N. Turok, "Polarization and Anisotropy of the Microwave Sky," *Phys.Rev.Lett.* **73** (1994) 2390–2393, arXiv:astro-ph/9406046 [astro-ph].

[175] **WMAP** Collaboration, H. V. Peiris *et al.*, "First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Implications for inflation," *Astrophys. J. Suppl.* **148** (2003) 213–231, arXiv:astro-ph/0302225 [astro-ph].

[176] D. Spergel and M. Zaldarriaga, "CMB Polarization as a Direct Test of Inflation," *Phys.Rev.Lett.* **79** (1997) 2180–2183, arXiv:astro-ph/9705182 [astro-ph].

[177] D. Baumann and M. Zaldarriaga, "Causality and Primordial Tensor Modes," *JCAP* **0906** (2009) 013, arXiv:0901.0958 [astro-ph.CO].

[178] M. Kamionkowski, A. Kosowsky, and A. Stebbins, "Statistics of Cosmic Microwave Background Polarization," *Phys.Rev.* **D55** (1997) 7368–7388, arXiv:astro-ph/9611125 [astro-ph].

[179] M. Zaldarriaga and U. Seljak, "An All-Sky Analysis of Polarization in the Microwave Background," *Phys.Rev.* **D55** (1997) 1830–1840, arXiv:astro-ph/9609170 [astro-ph].

[180] K. Smith and M. Zaldarriaga, "A General Solution to the E-B Mixing Problem," *Phys.Rev.* **D76** (2007) 043001, arXiv:astro-ph/0610059 [astro-ph].

[181] W. Hu and M. White, "CMB Anisotropies: Total Angular Momentum Method," *Phys.Rev.* **D56** (1997) 596–615, arXiv:astro-ph/9702170 [astro-ph].

[182] K. Jones-Smith, L. Krauss, and H. Mathur, "A Nearly Scale Invariant Spectrum of Gravitational Radiation from Global Phase Transitions," *Phys.Rev.Lett.* **100** (2008) 131302, arXiv:0712.0778 [astro-ph].

[183] U. Seljak, U.-L. Pen, and N. Turok, "Polarization of the Microwave Background in Defect Models," *Phys.Rev.Lett.* **79** (1997) 1615–1618, arXiv:astro-ph/9704231 [astro-ph].

[184] L. Pogosian and M. Wyman, "B-modes from Cosmic Strings," *Phys.Rev.* **D77** (2008) 083509, arXiv:0711.0747 [astro-ph].

[185] M. Zaldarriaga, "Fluctuations in the Cosmic Microwave Background," arXiv:astro-ph/9806122 [astro-ph].

[186] D. Baumann, "TASI Lectures on Inflation," arXiv:0907.5424 [hep-th].

[187] R. Durrer, *The Cosmic Microwave Background*. Cambridge Univ. Press, Cambridge, 2008.

[188] L. Knox, "Determination of inflationary observables by cosmic microwave background anisotropy experiments," *Phys.Rev.* **D52** (1995) 4307–4318, arXiv:astro-ph/9504054 [astro-ph].

[189] W. Wu *et al.*, "A Guide to Designing Future Ground-Based CMB Experiments," arXiv:1402.4108 [astro-ph.CO].

[190] **BICEP2** Collaboration, P. A. R. Ade *et al.*, "BICEP2 II: Experiment and Three-Year Data Set," *Astrophys. J.* **792** no. 1, (2014) 62, arXiv:1403.4302 [astro-ph.CO].

[191] J. Bock. http://burkeinstitute.caltech.edu/workshops/BICEP2.

[192] Z. Kermish *et al.*, "The POLARBEAR Experiment," arXiv:1210.7768 [astro-ph.IM].

[193] **COrE** Collaboration, F. R. Bouchet *et al.*, "COrE (Cosmic Origins Explorer) A White Paper," arXiv:1102.2181 [astro-ph.CO].

[194] J. Austermann *et al.*, "SPTPol: An Instrument for CMB Polarization Measurements with the South Pole Telescope," *Proc.SPIE Int.Soc.Opt.Eng.* **8452** (2012) 84520E, arXiv:1210.4970 [astro-ph.IM].

[195] T. Tomaru *et al.*, "The POLARBEAR-2 Experiment," *Proc. SPIE* **8452** (2012) 84521H–84521H–10.

[196] T. Matsumura *et al.*, "Mission Design of LiteBIRD," arXiv:1311.2847 [astro-ph.IM].

[197] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck intermediate results. XXII. Frequency dependence of thermal emission from Galactic dust in intensity and polarization," *Astron. Astrophys.* **576** (2015) A107, arXiv:1405.0874 [astro-ph.GA].

[198] **WMAP** Collaboration, L. Page *et al.*, "Three year Wilkinson Microwave Anisotropy Probe (WMAP) observations: polarization analysis," *Astrophys. J. Suppl.* **170** (2007) 335, arXiv:astro-ph/0603450 [astro-ph].

[199] D. Baumann *et al.*, "CMBPol Mission Concept Study: Probing Inflation with CMB Polarization," *AIP Conf.Proc.* **1141** (2009) 10–120, arXiv:0811.3919 [astro-ph].

[200] M. Tegmark, D. J. Eisenstein, W. Hu, and A. de Oliveira-Costa, "Foregrounds and Forecasts for the Cosmic Microwave Background," *Astrophys.J.* **530** (2000) 133–165, arXiv:astro-ph/9905257 [astro-ph].

[201] **Planck** Collaboration, A. Abergel *et al.*, "Planck 2013 results. XI. All-sky model of thermal dust emission," *Astron. Astrophys.* **571** (2014) A11, arXiv:1312.1300 [astro-ph.GA].

[202] J. Aumont, "Component separation for cmb polarization," *ESLAB conference* (2013) .

[203] M. Tucci *et al.*, "Limits on the detectability of the CMB B-mode polarization imposed by foregrounds," *Mon.Not.Roy.Astron.Soc.* **360** (2005) 935–949, arXiv:astro-ph/0411567 [astro-ph].

[204] L. Verde, H. Peiris, and R. Jimenez, "Optimizing CMB polarization experiments to constrain inflationary physics," *JCAP* **0601** (2006) 019, arXiv:astro-ph/0506036 [astro-ph].

[205] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck intermediate results. XIX. An overview of the polarized thermal emission from Galactic dust," *Astron. Astrophys.* **576** (2015) A104, arXiv:1405.0871 [astro-ph.GA].

[206] D. Finkbeiner, M. Davis, and D. Schlegel, "Extrapolation of galactic dust emission at 100 microns to CMBR frequencies using FIRAS," *Astrophys.J.* **524** (1999) 867–886, arXiv:astro-ph/9905128 [astro-ph].

[207] J. Dunkley *et al.*, "CMBPol Mission Concept Study: Prospects for Polarized Foreground Removal," arXiv:0811.3915 [astro-ph].

[208] H. Eriksen *et al.*, "CMB component separation by parameter estimation," *Astrophys.J.* **641** (2006) 665–682, arXiv:astro-ph/0508268 [astro-ph].

[209] R. Stompor, S. M. Leach, F. Stivoli, and C. Baccigalupi, "Maximum Likelihood algorithm for parametric component separation in CMB experiments," *Mon.Not.Roy.Astron.Soc.* **392** (2009) 216, arXiv:0804.2645 [astro-ph].

[210] M. Zaldarriaga and U. Seljak, "Gravitational Lensing Effect on Cosmic Microwave Background Polarization," *Phys.Rev.* **D58** (1998) 023003, arXiv:astro-ph/9803150 [astro-ph].

[211] W. Hu, "Weak Lensing of the CMB: A Harmonic Approach," *Phys.Rev.* **D62** (2000) 043007, arXiv:astro-ph/0001303 [astro-ph].

[212] T. Okamoto and W. Hu, "CMB Lensing Reconstruction on the Full Sky," *Phys.Rev.* **D67** (2003) 083002, arXiv:astro-ph/0301031 [astro-ph].

[213] C. Hirata and U. Seljak, "Reconstruction of lensing from the cosmic microwave background polarization," *Phys.Rev.* **D68** (2003) 083002, arXiv:astro-ph/0306354 [astro-ph].

[214] K. Smith *et al.*, "CMBPol Mission Concept Study: Gravitational Lensing," *AIP Conf.Proc.* **1141** (2009) 121, arXiv:0811.3916 [astro-ph].

[215] K. Smith, D. Hanson, M. LoVerde, C. Hirata, and O. Zahn, "Delensing CMB Polarization with External Datasets," *JCAP* **1206** (2012) 014, arXiv:1010.0048 [astro-ph.CO].

[216] U. Seljak and C. M. Hirata, "Gravitational Lensing as a Contaminant of the Gravity Wave Signal in CMB," *Phys.Rev.* **D69** (2004) 043005, arXiv:astro-ph/0310163 [astro-ph].

[217] M. Shimon, B. Keating, N. Ponthieu, and E. Hivon, "CMB Polarization Systematics Due to Beam Asymmetry: Impact on Inflationary Science," *Phys.Rev.* **D77** (2008) 083003, arXiv:0709.1513 [astro-ph].

[218] D. Hanson, G. Rocha, and K. Gorski, "Lensing reconstruction from PLANCK sky maps: inhomogeneous noise," *Mon.Not.Roy.Astron.Soc.* **400** (2009) 2169–2173, arXiv:0907.1927 [astro-ph.CO].

[219] D. Hanson, A. Lewis, and A. Challinor, "Asymmetric Beams and CMB Statistical Anisotropy," *Phys.Rev.* **D81** (2010) 103003, arXiv:1003.0198 [astro-ph.CO].

[220] D. Hanson, A. Challinor, G. Efstathiou, and P. Bielewicz, "CMB temperature lensing power reconstruction," *Phys.Rev.* **D83** (2011) 043005, arXiv:1008.4403 [astro-ph.CO].

[221] T. Namikawa, D. Hanson, and R. Takahashi, "Bias-Hardened CMB Lensing," *Mon.Not.Roy.Astron.Soc.* **431** (2013) 609–620, arXiv:1209.0091 [astro-ph.CO].

[222] A. Benoit-Levy *et al.*, "Full-sky CMB lensing reconstruction in presence of sky-cuts," arXiv:1301.4145 [astro-ph.CO].

[223] A. van Engelen *et al.*, "CMB Lensing Power Spectrum Biases from Galaxies and Clusters using High-angular Resolution Temperature Maps," *Astrophys.J.* **786** (2014) 13, arXiv:1310.7023 [astro-ph.CO].

[224] L. Boyle, K. M. Smith, C. Dvorkin, and N. Turok, "On testing and extending the inflationary consistency relation for tensor modes," arXiv:1408.3129 [astro-ph.CO].

[225] K. Smith, W. Hu, and M. Kaplinghat, "Weak Lensing of the CMB: Sampling Errors on B-modes," *Phys.Rev.* **D70** (2004) 043002, arXiv:astro-ph/0402442 [astro-ph].

[226] K. Smith, "Pseudo-C_ℓ Estimators Which Do Not Mix E- and B-modes," *Phys.Rev.* **D74** (2006) 083002, arXiv:astro-ph/0511629 [astro-ph].

[227] A. Lewis, A. Challinor, and A. Lasenby, "Efficient computation of CMB anisotropies in closed FRW models," *Astrophys. J.* **538** (2000) 473–476, arXiv:astro-ph/9911177 [astro-ph].

[228] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck 2013 results. XVI. Cosmological parameters," *Astron. Astrophys.* **571** (2014) A16, arXiv:1303.5076 [astro-ph.CO].

[229] R. Ogburn *et al.*, "BICEP2 and Keck Array Operational Overview and Status of Observations," arXiv:1208.0638 [astro-ph.IM].

[230] **BICEP3** Collaboration, Z. Ahmed *et al.*, "BICEP3: a 95GHz refracting telescope for degree-scale CMB polarization," *Proc. SPIE Int. Soc. Opt. Eng.* **9153** (2014) 91531N, arXiv:1407.5928 [astro-ph.IM].

[231] R. Flauger, C. Hill, and D. Spergel, "Toward an Understanding of Foreground Emission in the BICEP2 Region," arXiv:1405.7351 [astro-ph.CO].

[232] **SPTpol** Collaboration, D. Hanson *et al.*, "Detection of B-mode Polarization in the Cosmic Microwave Background with Data from the South Pole Telescope," *Phys. Rev. Lett.* **111** no. 14, (2013) 141301, arXiv:1307.5830 [astro-ph.CO].

[233] A. Lee. http://max.ifca.unican.es/EPI2013/epi2013_talks/Thursday/EPI_27062013_01_Lee.pdf.

[234] **CORE** Collaboration, E. Di Valentino *et al.*, "Exploring Cosmic Origins with CORE: Cosmological Parameters," arXiv:1612.00021 [astro-ph.CO].

[235] E. Leitch *et al.*, "DASI three-year cosmic microwave background polarization results," *Astrophys.J.* **624** (2005) 10–20, arXiv:astro-ph/0409357 [astro-ph].

[236] N. Katayama and E. Komatsu, "Simple foreground cleaning algorithm for detecting primordial B-mode polarization of the cosmic microwave background," *Astrophys.J.* **737** (2011) 78, arXiv:1101.5210 [astro-ph.CO].

[237] T. Namikawa and R. Nagata, "Lensing reconstruction from a patchwork of polarization maps," arXiv:1405.6568 [astro-ph.CO].

[238] K. Abazajian *et al.*, "Inflation Physics from the Cosmic Microwave Background and Large Scale Structure," arXiv:1309.5381 [astro-ph.CO].

[239] **BICEP2, Planck** Collaboration, P. A. R. Ade *et al.*, "Joint Analysis of BICEP2/*Keck Array* and *Planck* Data," *Phys. Rev. Lett.* **114** (2015) 101301, arXiv:1502.00612 [astro-ph.CO].

[240] X. Chen, "Primordial Non-Gaussianities from Inflation Models," *Adv. Astron.* **2010** (2010) 638979, arXiv:1002.1416 [astro-ph.CO].

[241] M. Alvarez *et al.*, "Testing Inflation with Large Scale Structure: Connecting Hopes with Reality," arXiv:1412.4671 [astro-ph.CO].

[242] A. Loeb and M. Zaldarriaga, "Measuring the Small-Scale Power Spectrum of Cosmic Density Fluctuations through 21 cm Tomography Prior to the Epoch of Structure Formation," *Phys. Rev. Lett.* **92** (2004) 211301, arXiv:astro-ph/0312134 [astro-ph].

[243] J. B. Muñoz, Y. Ali-Haïmoud, and M. Kamionkowski, "Primordial non-gaussianity from the bispectrum of 21-cm fluctuations in the dark ages," *Phys. Rev.* **D92** no. 8, (2015) 083508, arXiv:1506.04152 [astro-ph.CO].

[244] M. Kamionkowski and E. D. Kovetz, "The Quest for B Modes from Inflationary Gravitational Waves," *Ann. Rev. Astron. Astrophys.* **54** (2016) 227–269, arXiv:1510.06042 [astro-ph.CO].

[245] N. Arkani-Hamed, "Cosmological collider physics." Particle Physics Seminar Series, Perimeter Institute, 2014. http://pirsa.org/displayFlash.php?id=14110115.

[246] S. Caron-Huot, Z. Komargodski, A. Sever, and A. Zhiboedov, "Strings from Massive Higher Spins: The Asymptotic Uniqueness of the Veneziano Amplitude," arXiv:1607.04253 [hep-th].

[247] C. Brust and K. Hinterbichler, "Partially Massless Higher-Spin Theory," *JHEP* **02** (2017) 086, arXiv:1610.08510 [hep-th].

[248] M. A. Vasiliev, "Consistent equation for interacting gauge fields of all spins in (3+1)-dimensions," *Phys. Lett.* **B243** (1990) 378–382.

[249] M. A. Vasiliev, "More on equations of motion for interacting massless fields of all spins in (3+1)-dimensions," *Phys. Lett.* **B285** (1992) 225–234.

[250] M. A. Vasiliev, "Higher spin gauge theories in four-dimensions, three-dimensions, and two-dimensions," *Int. J. Mod. Phys.* **D5** (1996) 763–797, arXiv:hep-th/9611024 [hep-th].

[251] M. A. Vasiliev, "Higher spin gauge theories: Star product and AdS space," arXiv:hep-th/9910096 [hep-th].

[252] S. Raju, "New Recursion Relations and a Flat Space Limit for AdS/CFT Correlators," *Phys. Rev.* **D85** (2012) 126009, arXiv:1201.6449 [hep-th].

[253] D. Poland and D. Simmons-Duffin, "The conformal bootstrap," *Nature Phys.* **12** no. 6, (2016) 535–539.

[254] I. Antoniadis, P. O. Mazur, and E. Mottola, "Conformal Invariance, Dark Energy, and CMB Non-Gaussianity," *JCAP* **1209** (2012) 024, arXiv:1103.4164 [gr-qc].

[255] P. Creminelli, "Conformal Invariance of Scalar Perturbations in Inflation," *Phys.Rev.* **D85** (2012) 041302, arXiv:1108.0874 [hep-th].

[256] A. Bzowski, P. McFadden, and K. Skenderis, "Implications of conformal invariance in momentum space," *JHEP* **03** (2014) 111, arXiv:1304.7760 [hep-th].

[257] E. Pajer, G. L. Pimentel, and J. V. S. Van Wijck, "The Conformal Limit of Inflation in the Era of CMB Polarimetry," arXiv:1609.06993 [hep-th].

[258] C. Pitrou, J.-P. Uzan, and F. Bernardeau, "The cosmic microwave background bispectrum from the non-linear evolution of the cosmological perturbations," *JCAP* **1007** (2010) 003, arXiv:1003.0481 [astro-ph.CO].

[259] P. Creminelli, C. Pitrou, and F. Vernizzi, "The CMB bispectrum in the squeezed limit," *JCAP* **1111** (2011) 025, arXiv:1109.1822 [astro-ph.CO].

[260] N. Bartolo, S. Matarrese, and A. Riotto, "Non-Gaussianity in the Cosmic Microwave Background Anisotropies at Recombination in the Squeezed limit," *JCAP* **1202** (2012) 017, arXiv:1109.2043 [astro-ph.CO].

[261] S. Weinberg, *The Quantum Theory of Fields*, vol. 1. Cambridge University Press, Cambridge, 2005.

[262] V. Assassi, D. Baumann, E. Pajer, Y. Welling, and D. van der Woude, "Effective Theory of Large-Scale Structure with Primordial Non-Gaussianity," *JCAP* **1511** (2015) 024, arXiv:1505.06668 [astro-ph.CO].

[263] P. Creminelli, M. Serone, and E. Trincherini, "Nonlinear Representations of the Conformal Group and Mapping of Galileons," *JHEP* **1310** (2013) 040, arXiv:1306.2946 [hep-th].

[264] C. de Rham, M. Fasiello, and A. J. Tolley, "Galileon Duality," *Phys.Lett.* **B733** (2014) 46–51, arXiv:1308.2702 [hep-th].

[265] C. De Rham, L. Keltner, and A. J. Tolley, "Generalized Galileon Duality," *Phys. Rev.* **D90** no. 2, (2014) 024050, arXiv:1403.3690 [hep-th].