

Bootstrap Techniques in Flat Space and

Cosmology



Jakub Supeł

Supervisor: Dr. Enrico Pajer

Department of Applied Mathematics and Theoretical Physics University of Cambridge

> This thesis is submitted for the degree of Doctor of Philosophy

Trinity College

August 2022

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis 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 except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Bootstrap Techniques in Flat Space and Cosmology

Jakub Supeł

Abstract

The scientific understanding of the Universe is evolving at a rapid pace. Each new experiment yields more and more accurate measurements of its fundamental parameters. The standard cosmological model postulates a very early period of fast expansion, but the details of its underlying mechanism remain hidden behind a veil of high energies that we cannot access in particle accelerators. Physics of the early Universe can instead be studied by identifying its effects on cosmological fluctuations produced in the early Universe, which are responsible for the anisotropies of the Cosmic Microwave Background and the development of cosmic structure we can observe today.

On the other hand, General Relativity, in its classical formulation, is not fully compatible with the principles of quantum mechanics, and a theory connecting the two realms remains to be discovered. Direct experimental verification of such a theory is challenging due to the extremely high energies required. Therefore, cosmological perturbations provide an excellent window into the perturbative regime of quantum gravity. Since the primordial perturbations were produced in the highly energetic early Universe, they can in principle be used to distinguish between different quantum gravity models. It is therefore essential to develop methods of deriving their statistics from specific features of the models.

This thesis focuses on the cosmological bootstrap, a research program that attempts to derive features of cosmological fluctuations from simple physical principles expected to be satisfied in the early Universe. I study the effect of background curvature on standard soft theorems and its impact on observables in the context of the Effective Field Theory of inflation. I extend flat spacetime bootstrap methods to settings where the boost symmetry is violated. I also employ several well-known cosmological bootstrap methods to constrain graviton correlators at the end of inflation.

But God made the earth by his power; he founded the world by his wisdom and stretched out the heavens by his understanding.

Jeremiah 10:12, NIV

Acknowledgements

At the beginning of this thesis I would like to express my gratitude to many people who have helped me in its completion.

First and foremost, I would like to thank my supervisor Enrico Pajer, without whom undertaking the research presented here would not have been possible. His advice helped me realise my fullest potential and many of the results of this thesis were inspired by his ideas. I would also like to thank him for reading parts of this manuscript and providing valuable feedback.

In the course of my PhD I had the pleasure of working with several excellent researchers. Thus I would like to thank Enrico Pajer, Guus Avis, Sadra Jazayeri, David Stefanyszyn, Giovanni Cabass and Ayngaran Thavanesan for effective collaborations. I am grateful to Paolo Creminelli, James Fergusson, Mehrdad Mirbabayi, Brando Bellazzini, Paolo Benincasa, Tanguy Grall, Scott Melville, Dong-Gang Wang, Maria Gutierrez Guillen, João Melo, Daniel Baumann, James Bonifacio, Carlos Duaso Pueyo, Harry Goodhew, Aaron Hillman, Austin Joyce, Gui Pimentel and Jacopo Salvalaggio for discussions related to the research presented here and comments on some of the publications. I should also acknowledge the help and guidance of Jorge Santos.

Finally, I would like to thank Tanguy Grall, a fellow PhD student, for many fruitful conversations about interesting physics topics and for reading parts of this manuscript.

Preface

Chapters 5, 6, 8 and Section 7.3 are based on articles published as [1–4] and constitute a collaborative effort as is common in modern physics research. A summary of my contributions to these works can be found at the end of Chapter 1.

Table of contents

1 Introduction

| 2 | Stan | dard cosmology and perturbation theory | 9 | | |
|---|---------------------------------------|--|----|--|--|
| | 2.1 | Background evolution | 9 | | |
| | 2.2 | Inflation | 14 | | |
| | 2.3 | Inflationary perturbation theory | 17 | | |
| | 2.4 | Evolution of perturbations after inflation | 25 | | |
| 3 | Revi | Review of bootstrap methods | | | |
| | 3.1 | Spinor-helicity formalism | 29 | | |
| | 3.2 | S-matrix bootstrap | 35 | | |
| | 3.3 | The wavefunction formalism | 42 | | |
| | 3.4 | The Manifestly Local Test | 47 | | |
| | 3.5 | The Cosmological Optical Theorem | 51 | | |
| | 3.6 | Adiabatic modes and soft theorems | 55 | | |
| 4 | Adiabatic modes in flat FLRW universe | | | | |
| | 4.1 | Vector spherical harmonics | 61 | | |
| | 4.2 | The constraints | 62 | | |
| | 4.3 | Solutions | 64 | | |
| | 4.4 | Comparison with (Pajer and Jazayeri, 2018) | 70 | | |
| 5 | Spatial Curvature at Sound Horizon | | | | |
| | 5.1 | Introduction and summary | 73 | | |
| | 5.2 | Residual diffeomorphism | 79 | | |

1

| | 5.3 | Soft theorems | 89 | | |
|---|---|---|-----|--|--|
| | 5.4 | Curvature corrections to the power spectrum and bispectrum | 91 | | |
| | 5.5 | Observational constraints on curvature | 105 | | |
| | 5.6 | Discussion and conclusion | 113 | | |
| | 5.7 | Appendices | 115 | | |
| 6 | The | Boostless Bootstrap: Amplitudes without Lorentz boosts | 129 | | |
| | 6.1 | Introduction and summary | 130 | | |
| | 6.2 | On-shell methods: symmetries and bootstrap techniques | 140 | | |
| | 6.3 | Three-particle amplitudes | 149 | | |
| | 6.4 | Four-particle amplitudes and the four-particle test | 157 | | |
| | 6.5 | Mind the gap: amplitudes and the flat-space limit of cosmology | 188 | | |
| | 6.6 | Discussion and conclusion | 195 | | |
| | 6.7 | Appendices | 197 | | |
| 7 | Bootstrap via BCFW Momentum Shifts 2 | | | | |
| | 7.1 | The Britto-Cachazo-Feng-Witten (BCFW) method | 213 | | |
| | 7.2 | Lorentz-invariant four-particle amplitudes | 215 | | |
| | 7.3 | The Boostless Bootstrap and BCFW Momentum Shifts | 220 | | |
| 8 | Bootstrapping Large Graviton non-Gaussianities 22 | | | | |
| | 8.1 | Introduction | 228 | | |
| | 8.2 | Bootstrap techniques from symmetries, locality and unitarity | 234 | | |
| | 8.3 | Unitarity constraints on <i>n</i> -point cosmological correlators | 249 | | |
| | 8.4 | Bootstrapping <i>all</i> graviton bispectra | 254 | | |
| | 8.5 | Graviton bispectra and symmetry breaking patterns | 279 | | |
| | 8.6 | Summary and future directions | 306 | | |
| | 8.7 | Appendix | 308 | | |
| 9 | Disc | ussion | 315 | | |
| | 9.1 | Summary of the main results | 315 | | |
| | 9.2 | Outlook | 317 | | |

Bibliography

319

Chapter 1

Introduction

The long history of studying the Universe has been marked by a gradual rejection of the belief in our planet's unique location. This process began with the heliocentric model, first proposed by ancient Greek philosophers and revived in the 16th century by Copernicus. According to heliocentrism, the Earth is one of many bodies simultaneously orbiting the Sun, which remains at the centre of the cosmos. Soon after, astronomers such as Iohannes Kepler formalized the laws of this orbital motion, laying the groundwork for Newtonian dynamics and the law of universal gravity.

As the understanding of nature progressed, it has been accepted that the same laws that govern everyday life also apply to heavenly bodies and that the Sun is not dissimilar from the so-called fixed stars, which may accommodate their own planets. All stars visible with the naked eye are gravitationally bound in the spiral structure of the Milky Way galaxy, which is again not unique but one of many galaxies in the observable Universe. Therefore, the solar system is not in the centre of the Universe, and our cosmic neighbourhood appears to be indistinguishable from other points in space, in a sense that I will later make precise. This postulate is known as the *Copernican principle*.

While the Copernican principle concerning spatial dimensions conforms to the observations quite well on large scales, the same cannot be said about its temporal version. On the contrary, based on the expansion of the Universe demonstrated by Lemaitre and Hubble [5, 6], it has been suggested that the Universe has not always been the same but used to be much denser and hotter in the past. This hypothesis of a hot Big Bang has been since confirmed by multiple lines of evidence, from Cosmic Microwave Background experiments to observed abundances of light elements that match the predictions of Big Bang nucleosynthesis theory [7].

At sufficiently early times, any massive particle known from the Standard Model was ultrarelativistic, its energy far exceeding the rest mass. Simple thermodynamics dictates that as we go back even further in time and the length scales become smaller, such radiation permeating the Universe is blueshifted, and its energy increases. The energy scale of a hot Big Bang is the temperature of this radiation. Beyond $T_c \approx 158$ MeV, our modelling of the physics becomes uncertain, as hadrons are now unstable and are replaced by the quark-gluon plasma, which is poorly understood [8]. At present, we also lack experimental access to energies beyond a couple of TeV. Thus, we do not know what the correct description of the very early Universe is. However, there is growing evidence that if we continued going back in time, we would eventually reach the era of cosmological inflation: a period of accelerated expansion which occurred in the very early phase of cosmic history. I will discuss inflation in Section 2.2, where I will show how the exponential growth of the scale factor can explain the approximate flatness and homogeneity of the Universe.

The framework of inflation can also account for the spatial density fluctuations, which can be observed in the Cosmic Microwave Background (CMB) maps (Fig. 1.1). Although the fluctuations are fundamentally unpredictable due to their quantum origin, their statistical properties are tied to the details of the theory. Thus, a free (scalar or spinning) field φ generates *Gaussian* primordial fluctuations, for which the expectation value of the second power of fields $\langle \varphi_{\mathbf{k}} \varphi_{\mathbf{k}'} \rangle$ at the end of inflation is non-zero, while the connected components of all higher powers vanish. An interacting field would, in turn, generate primordial *non-Gaussianities*, which are higher moments of the distribution function.

Hypothetical observation of non-Gaussianities would give us an unprecedented opportunity to study the high energy physics of the very early Universe. The primordial quantum fluctuations - that become classical on length scales beyond the cosmological horizon - seeded the nonuniformities in the CMB and the large-scale structures that we observe today. In other words, short wavelength fluctuations in the far past have been stretched to cosmological scales during inflation and are, in principle, observable today as classical inhomogeneities. Therefore, the present observational data may, in principle, be used to reconstruct the inflationary correlators, which originate from interactions at energies much greater¹ than anything we could ever hope to achieve in terrestrial accelerators such as the Large Hadron Collider. Such a *cosmological collider* could probe the regime of perturbative quantum gravity, and by measuring primordial gravitational waves, either directly or via the imprints

¹The energy scale of inflation itself is the Hubble scale, which could be as large as $10^{-6}M_{\rm pl}$; $M_{\rm pl}$ being the Planck mass.



Figure 1.1: The anisotropy of the Cosmic Microwave Background observed by Planck is one of the relicts of quantum field fluctuations during inflation [9]. Reproduced with permission, ©ESA and the Planck Collaboration

they left on CMB polarization, we could constrain quantum corrections to the Einstein-Hilbert action.

For this reason, it is crucial to develop a good understanding of the *phenomenology* of inflation, that is to say, the correspondence between inflationary theories and their predictions. A standard method of deriving the statistics of primordial fluctuations (in perturbation theory) from a given action is discussed in Section 2.3.2. However, there are almost innumerable models of inflation that substantially differ from one another. To begin with, they might generate distinct *background evolution*, which provides the first connection to observations. They may also differ in the *field content* (the type and number of fields, and therefore particles, that describe the high energy physics of inflation), the *symmetry breaking patterns* and the *interactions* that give rise to primordial non-Gaussianities.

The proliferation and diversity of inflationary models motivate us to seek universal or semiuniversal *consistency relations* that the observables are expected to obey and which are derived from a simple set of assumptions. If a consistency relation is observed to be violated, at least one of the assumptions would need to be rejected. For example, the violation of Maldacena's soft theorem (discussed in Section 3.6, Equation 3.132) would indicate that one of the following must hold: (a) more than one field was relevant in the inflationary era, (b) the Equivalence Principle is false, (c) inflation underwent a non-attractor phase [10] or (d) inflation was not slow-roll (see Section 2.2). Still, the most recent observations are consistent with Maldacena's relation and, therefore, with the listed assumptions [11]. This and other soft theorems are discussed in Section 3.6. Chapter 5 (published as [1]) discusses soft theorems as well as leading order model-dependent corrections to correlators in the context of a spatially curved Universe.

Soft theorems do not exhaust the methods available to us when constraining cosmological observables. Indeed, there exist other rules based on simple principles that are extremely well-tested or theoretically motivated so that they apply to large classes of models and can serve as input in the cosmological *bootstrap program*. A major goal in recent years has been to find bootstrap rules that represent, in particular, the *unitary* dynamics of quantum fields directly on the level observables. Progress in this direction has been made by [12–14] (see Section 3.5).

An ambitious project is then to use bootstrap rules to find all possible non-Gaussianities consistent with a given background evolution, field content and symmetry-breaking pattern. A simplified strategy involves writing down an ansatz for a given observable whose form is dictated by a few simple rules and then using the remaining rules to maximally constrain the solution. This method has been used in [4], included as part of this thesis (Chapter 8), to construct tree-level graviton bispectra, with the emphasis on parity-odd bispectra.

As late-time observers, all data we have access to are the observables at the end of inflation. Current data suggests that the Hubble parameter was approximately constant during inflation (see Section 2.3.1), which means that its background dynamics can be approximated by a de Sitter spacetime. Thus, in this approximation, we can only observe the future spacelike infinity of de Sitter space. Our situation is therefore reminiscent of the anti-de Sitter/conformal field theory (AdS/CFT) correspondence [15, 16]. AdS/CFT stipulates that a theory with gravity in AdS spacetime of D + 1 dimensions is formally equivalent to a conformal field theory on the D-dimensional boundary of AdS. A major task in this research program is to develop dictionaries that relate the AdS fields, whose asymptotic behaviour is known, to boundary operators. Asymptotic AdS observables could then be represented as CFT correlators. Strongly coupled physics in the bulk spacetime (AdS) can be studied by computing operator expectation values in the weakly coupled regime of the boundary CFT, allowing for significant progress in our understanding of AdS physics [16–21].

It is tempting to suggest a similar correspondence for the de Sitter spacetime, which would potentially enable us to study inflation via a dynamical principle formulated on the future spacelike boundary. Such a theory would have to be "without time", as the future boundary has a Euclidean signature and would be non-unitary, despite encoding a unitary time evolution in the bulk. Early attempts towards formulating such a de Sitter *holography* include [22–25]. It is well known that de Sitter and AdS are related by the analytic continuation $t_{dS} \mapsto ir_{AdS}$. [26, 27] used holographic methods based on a more general correspondence between cosmological and the so-called domain-wall, Euclidean spacetimes, to compute primordial inflationary correlators. All of these attempts constitute an important part of the de Sitter bootstrap program.

In the 1960s, parallel to the developments in cosmology, particle physicists were invested in the S-matrix theory, which attempted to give the S-matrix, rather than the action, a fundamental status. In this theory, basic principles obeyed by the S-matrix serve as a starting point for constructing a detailed description of observables (scattering amplitudes) without invoking the Lagrangian. These basic principles include complex analyticity, the interpretation of poles in the complex plane as representing exchanged particles, and branch cuts as exchanged massive particles or loops. While this has not been sufficient to solve the conundrums of interactions and bound states observed in particle accelerators, and the project has been superseded by quantum chromo-dynamics (QCD), it inspired the later efforts associated with the S-matrix bootstrap. Among the most intriguing results of this bootstrap programme is the proof that GR is the unique theory of massless spin 2 particles whose interactions are second order in derivatives, along with the construction of tree-level graviton amplitudes for any number of particles from the building block of the cubic interaction. Furthermore, the gravitational coupling strength can be shown to be equal to the coupling strength of any massless elementary particle with spin less than 2 to gravity, while those with spin 2 or more cannot minimally couple to gravity [28–31]. A more complete introduction will be given in Section 3.2; an in-depth review can be found in [32].

S-matrix methods have thus been extensively used to develop a deeper understanding of gravity and other theories around Minkowski space, and it has been a major goal for theoretical cosmology to find similar techniques for de Sitter space. Since the latter has different symmetries, causal structure and asymptotics than flat space, the set of allowed interactions might be different, especially because the spectra of free particles in both cases do not coincide. However, if the case of Minkowski is any guide, we might hope to derive recursion relations for cosmological correlators in (quasi) de Sitter, which would allow us to (i) use lower-point correlators to explicitly construct higher-point ones and, more ambitiously, (ii) constrain the allowed lower-point interactions in de Sitter through consistency



Figure 1.2: Penrose diagrams representing the causal structure of AdS, Minkowski and de Sitter spacetime. Each point on the Minkowski and de Sitter Penrose diagrams corresponds to a 2-dimensional sphere. In AdS, the boundary \mathscr{I} has a Lorentzian signature. In Minkowski, the boundary is comprised of the past and future null infinity, denoted by \mathscr{I}^- and \mathscr{I}^+ . In de Sitter, there is past and future spacelike infinity \mathscr{I}^- and \mathscr{I}^+ which have Euclidean signature. One of the four Poincaré patches of de Sitter, described by the (η, \mathbf{x}) coordinates of Section 2.2, is represented by the shaded region.

relations.

On the one hand, we do not know of any proof, beyond perturbation theory, of a correspondence between specific physical processes and analytic properties of de Sitter observables. If we are presented with an analytic expression for a correlator, we do not know how to identify *in full generality* the physical processes that are responsible for it. On the other hand, in recent years, there has been some progress in the understanding of this matter on the level of perturbation theory. Some of these methods will be introduced in Chapter 3.

Chapters 5-8 of this thesis, based on collaborative work and published as [1–4] are part of the effort outlined in this introduction and constitute a step towards answering the following important questions:

How can unitary and local physics in the bulk spacetime be encoded in full generality in the boundary data, either in perturbation theory or beyond? Given a primordial correlator, how can we ascertain whether it could be produced by a unitary and local theory in a quasi de Sitter spacetime?

How can we match specific features of the theory to specific properties of the boundary data?

Given a theoretical model, how can we bypass the cumbersome bulk computations and calculate the observables in the most efficient manner?

The remainder of this thesis is intended to provide a partial resolution to the above problems. It is structured as follows. Chapter 2 consists of an introduction to the current cosmological and inflationary paradigm, which a reader familiar with the field may skip. In Chapter 3, I present the main methods used in this thesis. Chapter 4 consists of an original derivation of adiabatic modes in a spatially flat Universe in the vector spherical harmonics basis. The remaining chapters have been published as standalone articles, each constituting a collaborative effort. In Chapter 5 (published as [1]), I discuss a violation of the leading order soft theorem in a curved Universe and determine the leading order correction to the scalar power spectrum and bispectrum in the presence of curvature. In Chapters 6-7 (published as [2, 3]), I constrain the flat space S-matrix for a certain class of Lagrangians, in the absence of Lorentz boost invariance. Even though this constitutes a marked departure from the de Sitter context, it contributes to the bootstrap project by deriving powerful constraints on the Minkowski space boundary data while allowing for the violation of boost invariance - the de Sitter analogue of Lorentz boosts being a symmetry that is strongly broken in inflation. In Chapter 8 (published as [4]), I bootstrap graviton non-Gaussianities (both parity-even and parity-odd) that can be large enough to be observable, using the techniques developed in [33, 2].

Chapter 5 was published as [1] and coauthored by myself, Guus Avis, Sadra Jazayeri and Enrico Pajer. Sections 5.2.4 and 5.3 are primarily my own work. Sections 5.4 and 5.5 were written in close collaboration with Sadra Jazayeri. I verified all other sections and can attest to their accuracy as a coauthor.

Chapter 6 was published as [2] in collaboration with David Stefanyszyn and Enrico Pajer. Section 6.2 is primarily my own work; I also derived all the results in Sections 6.3 and 6.4, cross-checking my results with those of David Stefanyszyn. Appendices 6.7.1, 6.7.2, 6.7.3 and 6.7.4 are my own contributions. I contributed to a lesser degree to other parts of the paper and take full responsibility for any inaccuracies.

Section 7.3 is based on [3] which is a work completed in close collaboration with David Stefanyszyn.

Chapter 8 published as [4] was coauthored by myself, Giovanni Cabass, Enrico Pajer and David Stefanyszyn. The results of Sections 8.4 (except for the final subsection) and 8.5.3 are primarily

my work, as are parts of Section 8.2, although other authors also made substantial contributions and independently derived all results to provide a reliable verification. Section 8.5.5 results from collaborative work. I am familiar with all the other sections that were primarily written by Giovanni Cabass, Enrico Pajer and David Stefanyszyn.

Chapter 2

Standard cosmology and perturbation theory

2.1 Background evolution

The cosmological principle is one of the fundamental postulates of modern cosmology. It states that there exists a family of observers according to whom our Universe is spatially homogeneous and isotropic. Although this assumption does not hold on all scales, it is a good approximation at distances beyond a few hundred Mpc [34] and is confirmed by the nearly uniform temperature of the Cosmic Microwave Background [11]. The evolution of the Universe on large scales can be therefore approximated, to leading order, as an evolution of a spatially homogeneous and isotropic spacetime. Assuming General Relativity [35] to be applicable on cosmological scales, dynamics can be described by the Einstein equation:

$$G_{\mu\nu} = 8\pi T_{\mu\nu} \,, \tag{2.1}$$

where we assume the unit convention c = G = 1. It must be noted that many modifications have been proposed to (2.1), mainly as an alternative to dark matter and dark energy or as a solution to the Hubble tension problem [36, 37]. While some of these modifications are equivalent to merely adding extra sources to the stress-energy tensor (for example, a cosmological constant term is equivalent to $T^{\Lambda}_{\mu\nu} = -\frac{\Lambda}{8\pi}g_{\mu\nu}$), others modify the gravitational action itself, for example by introducing new fields beyond $g_{\mu\nu}$. In this thesis, I assume that (2.1) holds as presented for the average Einstein and energy-momentum tensors.

Spatially homogeneous and isotropic spacetimes can be described by the *Friedmann-Lemaitre-Robertson-Walker* (FLRW) metric

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0\\ 0 & a^2(t)f^2(r) & 0 & 0\\ 0 & 0 & a^2(t)f^2(r) & 0\\ 0 & 0 & 0 & a^2(t)f^2(r) \end{pmatrix},$$
 (2.2)

where $f(r) = \frac{1}{1+\frac{1}{4}Kr^2}$ and $r^2 = x^2 + y^2 + z^2$. The only function that is not constrained by homogeneity and isotropy is the *scale factor* a(t). A constant parameter K corresponds to the spatial curvature: K = -1 for a universe with negative curvature, K = 0 for a flat universe and K = +1 if the curvature is positive.

Spatially homogeneous and isotropic stress-energy tensors must take the form

$$T_{\mu\nu} = \begin{pmatrix} \bar{\rho} & 0 & 0 & 0 \\ 0 & a^2(t)f^2(r)\bar{P} & 0 & 0 \\ 0 & 0 & a^2(t)f^2(r)\bar{P} & 0 \\ 0 & 0 & 0 & a^2(t)f^2(r)\bar{P} \end{pmatrix},$$
(2.3)

where $\bar{\rho}$ and \bar{P} are the background values of energy density and pressure, respectively. It is now straightforward to express (2.1) in terms of background quantities and find the relationship between matter and background spacetime. The nontrivial components of (2.1) are the 00 and the diagonal components *ii*, giving a system of equations that describe the background dynamics, known as *Friedmann* equations [35]

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}\bar{\rho} - \frac{K}{a^2},\tag{2.4}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}(\bar{\rho} + 3\bar{P}). \tag{2.5}$$

From the two equations above, we can derive another one, which can be interpreted as the conservation of energy:

$$\dot{\bar{\rho}} = -3\frac{\dot{a}}{a}\left(\bar{\rho} + \bar{P}\right). \tag{2.6}$$

The ratio $H \equiv \dot{a}/a$ is, of course, the *Hubble parameter*, which corresponds to the rate of the homogeneous and isotropic expansion of the Universe. Next, it is convenient to define

$$\Omega_a = \frac{8\pi}{3H_0^2}\bar{\rho}_a, \qquad (2.7)$$

$$\Omega_k = -\frac{K}{H_0^2}, \qquad (2.8)$$

where H_0 is the present value of the Hubble parameter, and rewrite the first Friedmann equation (2.4) as follows:

$$\sum_{a} \Omega_a + \frac{\Omega_k}{a^2} = \left(\frac{H}{H_0}\right)^2.$$
(2.9)

Here, the *a* index labels distinct contributions to the stress-energy tensor. Only three distinct contributions are sufficient to explain almost all of the cosmological data:

- Cold matter, also known as *dust*, which has negligible pressure ($\bar{P}_D \approx 0$). This contribution consists of ordinary (baryonic and leptonic) matter and dark matter.
- Radiation $(\bar{P}_R = \frac{1}{3}\bar{\rho}_R)$. This includes massless particles such as photons and gravitons, but also massive particles that are in the ultrarelativistic regime $(v \approx c)$.¹
- Dark energy that behaves as a cosmological constant $(\bar{P}_{\Lambda} = -\bar{\rho}_{\Lambda})$.

In a simplified scenario of a spatially flat universe with only one significant contribution to the stress-energy tensor, we can easily solve the Friedmann equations for a(t). The dynamics of the scale factor then depends on whether the energy density is dominated by pressureless matter, radiation or dark energy (cosmological constant).

- In the matter-dominated era ($\bar{P} \approx 0$), we have $a(t) \propto t^{2/3}$.
- In the radiation-dominated era $(\bar{P} \approx \frac{1}{3}\bar{\rho})$, we have $a(t) \propto t^{1/2}$.
- In the era dominated by dark energy with the equation of state $\bar{P}_{\Lambda} = -\bar{\rho}_{\Lambda}$, we have $a(t) \propto e^{Ht}$, where $H = \sqrt{\frac{8\pi}{3}\bar{\rho}_{\Lambda}} = const. \bar{\rho}_{\Lambda}$ is related to the cosmological constant via $\bar{\rho}_{\Lambda} = \frac{\Lambda}{8\pi}$.

Current data, for example the Planck satellite observations of the Cosmic Microwave Background [39], indicate that spatial curvature is at most of order 10^{-3} and is actually consistent with zero:

¹The possibility that at least some species of the Big Bang neutrinos remain relativistic despite having been redshifted is consistent with current constraints on neutrino masses [38].

| Pressureless matter, Ω_m | Radiation*, Ω_r | Dark energy, Ω_{Λ} | Spatial curvature, Ω_k |
|---------------------------------|--------------------------------------|---------------------------------|-------------------------------|
| 0.3111 ± 0.0056 | $(9.2311 \pm 0.0077) \times 10^{-5}$ | 0.6847 ± 0.0073 | 0.0007 ± 0.0019 |

Table 2.1: Recent estimates of the main components of the Universe, with 68% confidence intervals [39], [40]. The photon and the neutrino energy densities are included in *radiation*, assuming the existence of exactly three neutrino species.

 $\Omega_{k,0} = 0.0007 \pm 0.0019$ (68% CI, Planck + BAO). The relative energy densities of all components are listed in Table 2.1.

The flatness problem

The physical value of spatial curvature is [1]

$$\Omega_K = \frac{K}{a^2 H^2}.\tag{2.10}$$

The first Friedmann equation (2.4) can be rewritten in terms of the values of Ω_a at present, $\Omega_{a,0}$, as

$$\frac{\Omega_{r,0}}{a^2} + \frac{\Omega_{m,0}}{a} + a^2 \Omega_{\Lambda,0} - \frac{K}{H_0^2} = \left(\frac{aH}{H_0}\right)^2.$$
 (2.11)

Now, if Ω_{Λ} is negligible, as it has been for a significant part of the cosmic history, then the left-hand side is decreasing as the Universe expands. Therefore, aH must decrease, while $|\Omega_K| = \frac{|K|}{a^2H^2}$ must increase as the Universe expands. More precisely, if we assume $|\Omega_K|$ has always been small, then

$$\left|\frac{\Omega_K(t_E)}{\Omega_{K,0}}\right| \approx a(t_E)^2 \frac{1}{\Omega_{r,0}},\tag{2.12}$$

where t_E is some early time when radiation was the dominant component of the Universe. At the time of neutron decoupling $(a(t_{dec}) \sim 10^{-9})$, the physics of which is well understood, we have

$$\left|\frac{\Omega_K(t_{dec})}{\Omega_{K,0}}\right| \approx 6.8 \times 10^{-15}.$$
 (2.13)

During the hot Big Bang, the mean physical curvature Ω_K was therefore many orders of magnitude closer to zero than it is today. The fact that spatial curvature was equal to zero to such a great precision requires an explanation. This apparent fine-tuning is known as the *flatness problem*, a more detailed

discussion of which can be found in [41].

The horizon problem

Since at least the measurements by COBE [42], confirmed by many other experiments such as the WMAP and Planck satellites [11, 39, 43–48], it has been known that the CMB temperature fluctuations are very small. A typical amplitude of relative temperature fluctuations is $(\Delta T/T) \sim 10^{-5}$ to 10^{-4} , depending on the angular scale [39]. The three-dimensional Large Scale Structure (LSS) of the Universe has a similar property of being exceedingly uniform on the largest scales, i.e. those exceeding about 100 Mpc [49, 50]. This has long been interpreted as a problem for standard cosmology because in a universe dominated by matter and radiation, the particle horizon - the largest comoving distance an observer could receive any signals from - shrinks as we go into the past. The ratio of particle horizon at the time when the CMB was emitted to the currently observed radius of the CMB sphere, i.e. the sphere of last scattering (42 billion light years) is

$$\frac{r_p(t_{CMB})}{r_{CMB}} \approx 0.063. \tag{2.14}$$

This implies that the size of the particle horizon at the time of CMB emission corresponds to the angle of about ~ 1.2 degrees on the last scattering surface. If the assumptions behind this argument are correct, then regions larger than about 1 degree on the last scattering surface could not have been in causal contact before the CMB was emitted. Yet we know that the CMB is nearly uniform even on the largest angular scales. Because such a uniform background would again represent a high degree of inexplicable fine-tuning, a natural hypothesis is that distant regions of the CMB had been, in fact, in causal contact before the Universe became transparent to photons, so those different regions had reached approximate thermal equilibrium by then. Note that this argument does not assume, and is not limited to, any specific physics responsible for energy exchange between neighbouring regions of the Universe - the only assumption is that such physics is not superluminal.

Both the flatness and the horizon problems can be addressed by postulating a period of accelerated expansion of the Universe, known as *inflation*. In the next section, I will discuss the fundamental physics of inflation and demonstrate how it answers the problems presented above while simultaneously explaining features of cosmological perturbations (Fig. 2.1) that have been observed *after* the theory of inflation was proposed (Section 2.3.1).

2.2 Inflation

The hypothesis of cosmological inflation was originally put forward to address the magnetic monopole problem [51], and it was soon realised that it would also solve the problems of standard cosmology discussed in the previous section [52]. Inflation is defined as a sustained period during which the background evolution satisfies the inequality $\ddot{a} > 0$. Such an acceleration of the expansion rate can be achieved in a simple theoretical model with a cosmological constant Λ that dominates over other types of energy.

It is convenient to describe the kinematics of inflationary expansion using conformal time and comoving coordinates. We define the conformal time by

$$d\eta = \frac{dt}{a(t)}.$$
(2.15)

In terms of η , the background metric takes a simplified form. In the flat case (K = 0) we have

$$ds^{2} = a(\eta)^{2} \left(-d\eta^{2} + d\mathbf{x}^{2} \right).$$
(2.16)

Note that light cones in these coordinates have a constant slope equal to one. In other words, radial light rays travel along the lines of constant $\eta \pm r$ in the (η, r) spacetime diagram.

Given the new time coordinate, I should distinguish between time derivatives with respect to tand those taken with respect to η . I will use the notation

$$\dot{f} \equiv \frac{df}{dt},$$
 (2.17)

$$f' \equiv \frac{df}{d\eta} = \frac{dt}{d\eta}\dot{f} = a\dot{f}.$$
(2.18)

2.2.1 Motivation for the theory of inflation

Let us now see how a prolonged period of inflation solves the flatness and the horizon problem. Suppose the cosmological constant and the curvature term are the only contributions to the Friedmann equations, while $K = \pm 1$. The solutions are

$$a(t) = \begin{cases} \frac{1}{H_{\infty}} \sinh(H_{\infty}t) & \text{if } K = -1, \\ \frac{1}{H_{\infty}} \cosh(H_{\infty}t) & \text{if } K = +1. \end{cases}$$
(2.19)

$$\Omega_{K}(t) = \begin{cases}
-\frac{1}{\cosh^{2}(H_{\infty}t)} & \text{if } K = -1, \\
\frac{1}{\sinh^{2}(H_{\infty}t)} & \text{if } K = +1.
\end{cases}$$
(2.20)

where $H_{\infty} = \sqrt{\Lambda/3}$ is the late-time limit of the Hubble parameter. In both the open and the closed case, $|\Omega_K|$ decays as $\exp(-2H_{\infty}t)$ for large t. In other words, the Universe is becoming flatter. The fact that Ω_K is very close to zero is then explained, provided that inflation lasts sufficiently long.

As for the horizon problem, it will suffice to compute the particle horizon r_p near the end of inflation. For K = -1, 0, we have

$$r_p = \int_{a=a_i}^{a=a_f} \frac{dt}{a(t)} \to +\infty \quad \text{as} \quad a_i \to 0,$$
(2.21)

while for K = 1, r_p is also large provided that physical curvature at the end of inflation is sufficiently small. Therefore, we see that for a sufficiently long period of inflation, any two regions currently in the observable Universe could have exchanged signals before the CMB was emitted and therefore could have reached thermal equilibrium. Deviations from uniform CMB temperature and matter density can then be explained by invoking the theory of quantum fluctuations (Section 2.3).

Note that if the cosmological constant Λ were indeed exactly constant over time, then all matter and radiation would soon be heavily diluted, and inflation would last forever into the future, with the scale factor given by $a(t) = \exp(Ht)$, $H = \sqrt{\Lambda/3}$. Such an eternal, exponential expansion is certainly not what we observe. Therefore, multiple models have been proposed to provide an end to the inflationary era, either locally (through a spontaneous phase transition from an eternally inflating background) or globally in space, albeit at slightly different times (as the inflaton field potential naturally reaches the minimum). Such models usually entail a deviation from exponential expansion, so H becomes dependent on time. For later convenience, let us investigate the basic kinematics of inflation with weakly time-dependent Λ and H.

First, I define two *slow-roll* parameters ($\tilde{\eta}$ should not be confused with the conformal time η).

$$\epsilon = -\frac{\dot{H}}{H^2}, \qquad (2.22)$$

$$\tilde{\eta} = \frac{\frac{d}{dt} (\log \epsilon)}{H}.$$
(2.23)

There are several important classes of inflationary kinematics, grouped according to the behaviour of slow-roll parameters:

- Exact de Sitter. This corresponds to ε = η̃ = 0, so H is exactly constant. As remarked above, such a scenario is eternal and not realistic. Nevertheless, it is still useful for computing approximate values of those inflationary observables that are nonvanishing in the limit ε, η̃ → 0.
- Slow-roll inflation. This small deviation from de Sitter background is characterized by |ε|, |η̃| ≪
 1. Because |η̃| ≪ 1, we can treat ε as approximately constant. In that case, we have

$$a(\eta) = -\frac{1}{H_1 \eta^{\frac{1}{1-\epsilon}}},$$
(2.24)

$$H(\eta) = \frac{H_1}{1-\epsilon} \eta^{\frac{\epsilon}{1-\epsilon}}.$$
(2.25)

Because the deviation from constant H is small, I often refer to this spacetime as *quasi-de* Sitter.

• Ultra-slow-roll inflation. Characterized by $|\epsilon| \ll 1$, $\tilde{\eta} \approx -6$, which can be achieved by means of a very flat inflaton potential [53].

This thesis focuses on *field theoretical* models of inflation. The existence of a Lagrangian description of the dynamics in terms of *fields* is generally assumed (usually one in which inflation is driven only by a single scalar field). A UV cutoff at some energy scale Λ_c (not to be confused with the cosmological constant Λ) is also introduced, even if implicitly, to reflect our ignorance of high energy physics and prevent the rise of UV loop divergences. The description assumed throughout the thesis is *field theoretical* in the sense that I remain agnostic about the fundamental realisation of the physics that would require a qualitatively different description, such as vibrating strings, branes, spin foams, etc.

2.3 Inflationary perturbation theory

One of the main tasks of the theory of inflation is to connect a given model (or a class thereof) to specific observational consequences. The statistics of primordial fluctuations, generated by quantum effects during inflation, are the prime candidate to facilitate such a connection.

The remainder of this Chapter is organised as follows. In 2.3.1, I derive the power spectra of scalar and tensor curvature perturbations under specific assumptions and compare the results with recent data. In 2.3.2, I introduce the in-in method for computing higher order primordial correlators. Once inflation ends, these primordial seed fluctuations evolve classically. Although this post-inflationary evolution is not the main focus of this work, in 2.4 I briefly introduce the principles behind it.

2.3.1 Power spectra

Consider a field X with a vanishing expectation value,² $\langle X \rangle = 0$. The *power spectrum* $P_X(k)$ is defined as the second moment of X:

$$\langle X(\mathbf{k})X(\mathbf{k}')\rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k} + \mathbf{k}')P_X(k).$$
(2.26)

Note that in the presence of rotation symmetry, P_X may only depend on $k = |\mathbf{k}|$.

Power spectrum of ζ

A key observable is the scalar curvature perturbation and its power spectrum [54, 22]. First, note that in a vacuum (in the absence of any inflaton field), the metric's scalar perturbation modes can be eliminated by a redefinition of the comoving coordinates x_i , such that in the new coordinates, the scalar curvature becomes exactly zero. However, in single-clock inflation, there also exists a scalar field ϕ whose perturbations cannot be eliminated simultaneously with those of the scalar curvature. This suggests that we can redefine the coordinates so that $\delta \phi = 0$, but then scalar curvature cannot be made to vanish. This choice of gauge is called the *comoving gauge* and is convenient for deriving the statistics of the ζ fluctuations.

Neglecting tensor modes and assuming the mean curvature K = 0, the metric may be represented as

²This can always be ensured by subtracting off the vacuum expectation value from a field with a nonvanishing vev.

$$ds^{2} = -N^{2}dt^{2} + a^{2}e^{2\zeta}(dx^{i} + N^{i}dt)^{2}.$$
(2.27)

The inflaton field can be taken to be uniform due to our choice of gauge: $\phi(t, \mathbf{x}) = \overline{\phi}(t)$. Now we expand the action³

$$S = \frac{1}{8\pi} \int d^4x \sqrt{-g} \left(\frac{1}{2}R + \mathcal{L}_{\phi}\right), \qquad (2.28)$$

to second order in perturbation ζ [22],

$$S^{(2)} = M_{\rm pl}^2 \int dt d^3x \ a^3 \epsilon \left(\dot{\zeta}^2 - a^{-2} (\partial_i \zeta)^2 \right).$$
(2.29)

The equation of motion for ζ is

$$\ddot{\zeta} + 3H \left(1 + \tilde{\eta}\right) \dot{\zeta} - a^{-2} \nabla^2 \zeta = 0,$$
(2.30)

or, in conformal time,

$$\zeta'' - \frac{2}{\eta} \left(1 + \frac{3}{2} \tilde{\eta} \right) \zeta' - \nabla^2 \zeta = 0.$$
(2.31)

In the late time limit, when - due to expansion - every mode becomes a superhorizon mode, the dominant solution is $\zeta(t) \sim \text{const}$; the other solution decays approximately as η^3 , provided that $|\tilde{\eta}|$ is small. This means that once the scalar curvature perturbation has a wavelength larger than the current Hubble horizon, its amplitude remains constant - we say that the mode is *frozen*.⁴ To find the power spectrum of ζ , it is therefore sufficient to determine the statistics of quantum fluctuations at horizon exit, i.e. at the moment when k = aH. From that point on, expanding fluctuations cannot communicate and thus evolve classically, according to the free equation of motion.

To calculate the statistics of quantum fluctuations at the horizon exit, we decompose ζ in individual modes:

$$\phi(\eta; \mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} \left(a^{\dagger}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} f_k^+(\eta) + a(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}} f_k^-(\eta) \right),$$
(2.32)

³I take G = c = 1, hence $M_{pl}^2 = \frac{1}{8\pi}$. In the final result (2.39), one can reintroduce the physical value of M_{pl} in a straightforward way.

⁴Note that if $|\tilde{\eta}| > 1$, the dominant mode grows with time. In ultra-slow-roll inflation, $\tilde{\eta} = -6$, and $\zeta \sim \eta^{-15}$.

where a, a^{\dagger} are the usual creation-annihilation operators. There exists a *vacuum state*, which satisfies

$$a(\mathbf{k})|0\rangle = 0. \tag{2.33}$$

The mode functions f_k^{\pm} both satisfy the free equation of motion for ζ . Specifying the vacuum state $|0\rangle$ is equivalent to choosing the boundary conditions for f_k^{\pm} , but in de Sitter space, there is some ambiguity in this procedure. One option is to choose the following boundary conditions that are also consistent with the canonical commutation relations:

$$f_k^{\pm}(\eta)e^{\mp ik\eta} \sim \mp \frac{iH\eta}{2M_{pl}\sqrt{\epsilon k}} \quad \text{as} \quad \eta \to -\infty,$$
 (2.34)

$$\lim_{\eta \to -\infty(1+i\epsilon)} f_k^-(\eta) = 0, \tag{2.35}$$

$$\lim_{\eta \to -\infty(1-i\epsilon)} f_k^+(\eta) = 0, \qquad (2.36)$$

which implies

$$f_k^{\pm}(\eta) = \frac{H}{2M_{pl}\sqrt{\epsilon k^3}} \left(1 \mp ik\eta\right) e^{\pm ik\eta}.$$
(2.37)

The mode function f_k^+ (f_k^-) is now interpreted as the negative (positive) frequency mode in the far past. With this choice, $|0\rangle$ is the Bunch-Davies vacuum, defined as the state annihilated by all the negative frequency modes. Such an initial state can be motivated as follows. As we go back in time, every particle is deep inside the Hubble horizon and does not experience the effects of expansion, so it propagates as if it were in flat space. The Bunch-Davies vacuum condition then requires that in the limit $\eta \to -\infty$, no particles are present in this (locally) flat space. I should note that this explanation holds for a spatially flat universe, but if mean curvature is present, then particles would experience its effects at arbitrarily early times and could never be described as propagating in approximately flat space.

Assuming that the quantum field's initial state was the Bunch-Davies vacuum $|0\rangle$, we can compute the power spectrum of ζ as follows.

$$P_{\zeta}(k) := \langle 0|\zeta_{\mathbf{k}}(\eta=0)\zeta_{\mathbf{k}'}(\eta=0)|0\rangle' = |f_k^+(0)|^2 = \frac{H^2}{4\epsilon M_{pl}^2 k^3}.$$
(2.38)

In the above, I used the approximation of constant H. To account for the slow variation of the

Hubble parameter, H in the above formula should be evaluated approximately at the time when a given mode crossed the horizon. This is because once a mode becomes much longer than the horizon scale, it stops evolving and is not affected by the variation of the Hubble parameter. The required H is found by solving $k = a(\eta_c)H(\eta_c)$, which leads to

$$P_{\zeta}(k) = \frac{H(\eta_c)^2}{4\epsilon M_{pl}^2 k^3} \propto k^{-3} k^{n_s - 1}, \qquad (2.39)$$

$$n_s - 1 \approx -2\epsilon - \tilde{\eta}.$$
 (2.40)

The parameter $n_s - 1$ is called the *scalar spectral tilt*. It quantifies the deviation of the power spectrum from scale invariance,⁵ and in slow-roll models it is predicted to be small, which is consistent with observations (see Figure 2.1 and Table 2.2).

4.00 Marginalised 3.75 3.50 $\ln(10^{10}\mathcal{P}_{\mathcal{R}})$ 3.25 3.00 2.75 2.50 2.25 2.00 10^{-1} -3 10^{-2} 10 $[Mpc^{-1}]$ k

Power spectrum of γ_{ij}

Linearised general relativity is a theory of a massless tensor $h_{\mu\nu}$ accompanied by a diffeomorphism symmetry. Quantum fluctuations of $h_{\mu\nu}$ during inflation have a tensor component which is physical and thus cannot be eliminated by a gauge transfor-

Figure 2.1: The primordial power spectrum likelihood reconstructed from the Planck CMB maps [44]. The scalar tilt is nearly constant. N. Aghanim et al., A&A 641 (2020) A1, p. 32, reproduced with permission ©ESO.

mation. If the characteristic energy of inflation was sufficiently large, it should have generated tensor perturbations of the metric - primordial gravitational waves. These are potentially observable, either directly in gravitational wave detectors [55] or indirectly, by producing B-mode polarised waves in

⁵Consider

$$\langle \zeta^2(\mathbf{x}) \rangle = \int \frac{d^3k}{(2\pi)^3} P_{\zeta}(k) = \frac{1}{2\pi^2} \int d\log(k) \left(k^3 P_{\zeta}(k)\right).$$
 (2.41)

The contribution to the variance of $\zeta(\mathbf{x})$ from wavenumbers between k_1 and k_2 can be therefore estimated as

$$\frac{1}{2\pi^2} \int_{\log k_1}^{\log k_2} d\log(k) \left(k^3 P_{\zeta}(k)\right).$$
(2.42)

This depends only on the raio k_2/k_1 , and is therefore independent of the scale, provided that $k^3 P_{\zeta}(k) = const$; hence $P_{\zeta}(k) \propto k^{-3}$ is a *scale invariant* spectrum.
the CMB [56].⁶

I will now calculate the power spectrum of primordial tensor fluctuations in slow-roll inflation, assuming the Einstein-Hilbert action is a valid description of the linearised physics at inflationary energy scales. The metric is of the form

$$ds^{2} = -dt^{2} + a^{2}(e^{\gamma})_{ij}dx^{i}dx^{j}.$$
(2.43)

All degrees of freedom except for the two tensor modes are eliminated by imposing the conditions $\gamma_{ii} = 0, \gamma_{ij} = \gamma_{ji}$ and $\partial_j \gamma_{ij} = 0$. I have already dropped the lapse N and shift N_i because they can be solved for using constraint equations, and at leading (linear) order, they are independent of γ_{ij} . Similarly, the scalar perturbation ζ studied in the previous section (and equivalent to the trace of γ_{ij}) does not couple to γ_{ij} at the quadratic level, so can also be ignored.⁷

Expanding the action (2.28) to second order in γ_{ij} , we obtain

$$S^{(2)} = \frac{1}{8} M_{pl}^2 \int dt d^3 x a^3 \left(\dot{\gamma}_{ij}^2 - a^{-2} (\partial_k \gamma_{ij})^2 \right).$$
(2.44)

I will call this action the *canonical* quadratic action for gravity. Recall that it was derived from the Einstein-Hilbert action by perturbing the metric.

The tensor perturbations may be decomposed into two helicity components, $\gamma_{ij}(\mathbf{k}) = \sum_{\lambda} e_{ij}^{\lambda}(\mathbf{k}) \gamma_{\mathbf{k}}^{\lambda}$. The polarisation tensors satisfy the following conditions:

$$e_{ii}^{h}(\mathbf{k}) = k^{i} e_{ij}^{h}(\mathbf{k}) = 0$$
 (transverse and traceless), (2.45)

$$e_{ij}^{h}(\mathbf{k}) = e_{ji}^{h}(\mathbf{k})$$
 (symmetric), (2.46)

$$e_{ii}^{h}(\mathbf{k})e_{ik}^{h}(\mathbf{k}) = 0 \qquad \text{(lightlike)}, \qquad (2.47)$$

$$e_{ij}^{h}(\mathbf{k})e_{ij}^{h'}(\mathbf{k})^{*} = 2\delta_{hh'} \qquad \text{(normalization)}, \qquad (2.48)$$

$$e_{ij}^{h}(\mathbf{k})^{*} = e_{ij}^{h}(-\mathbf{k}) \qquad (\gamma_{ij}(x) \text{ is real}).$$
(2.49)

Then

⁶It must be noted that primordial gravitational waves may also be generated, for example, by early phase transitions and bubble collisions. This is, however, beyond the scope of this section.

⁷It is impossible to construct a nonvanishing scalar or a vector at linear order in perturbations, because γ_{ii} , $\partial_i \partial_j \gamma_{ij} = 0$ and $\partial_i \gamma_{ij} = 0$.

$$S^{(2)} = \frac{1}{4} M_{pl}^2 \int dt \frac{d^3k}{(2\pi)^3} a^3 \sum_{\lambda} \left(\dot{\gamma}_{\mathbf{k}}^{\lambda} \dot{\gamma}_{-\mathbf{k}}^{\lambda} - a^{-2} k^2 \gamma_{\mathbf{k}}^{\lambda} \gamma_{-\mathbf{k}}^{\lambda} \right).$$
(2.50)

Following the same quantisation procedure as before, we find that the power spectrum is

$$P_{\gamma}^{\lambda\lambda'}(k) = \langle \gamma_{\mathbf{k}}^{\lambda} \gamma_{-\mathbf{k}}^{\lambda'} \rangle' = \delta_{\lambda\lambda'} \frac{H^2}{M_{nl}^2 k^3}.$$
(2.51)

Therefore, the total power spectrum (summed over the polarisations) is

$$P_{\gamma}(k) = \langle \gamma_{\mathbf{k},ij} \gamma_{-\mathbf{k},ij} \rangle' = 2 \times 2 \times \frac{H^2}{M_{pl}^2 k^3} = \frac{4H^2}{M_{pl}^2 k^3}.$$
 (2.52)

(One factor of two arises because of two polarisations, while the second one is due to our normalisation of e_{ij} .) As for the case of scalar perturbations, the Hubble parameter in the above power spectrum should be evaluated at horizon exit because after that moment, fluctuations evolve classically and are frozen. This will again lead to a slight deviation from a scale-invariant power spectrum:

$$k^3 P_{\gamma}(k) \propto k^{n_t}, \qquad (2.53)$$

$$n_t \approx -2\epsilon,$$
 (2.54)

where n_t is the *tensor tilt*.

There are several important assumptions in the above derivation:

- Absence of other tensor fields beside the graviton.
- Quantum corrections to the GR action do not modify the quadratic Lagrangian.
- Validity of perturbation theory around the background, quasi de Sitter solution and validity of the tree-level approximation for the power spectrum (so that loop constributions to the PS can be neglected).
- Approximate de Sitter spacetime.

Current constraints on the primordial power spectra

Let us summarise the predicted power spectra for scalar and tensor curvature fluctuations in a theory of single-clock inflation with classical GR. We have

| Amplitude of P_{ζ} | $\ln(10^{10}A_s) = 3.044 \pm 0.014$ |
|--------------------------|--|
| Tensor-to-scalar ratio | $r_{0.002} < 0.056 \; (95\% \; \mathrm{CL})$ |
| Scalar spectral tilt | $n_s = 0.9649 \pm 0.0042$ |

Table 2.2: Current estimates of the power spectrum parameters [11, 57]. Constant n_s is consistent with the data at a 2σ level.

$$k^{3}P_{\zeta}(k) = A_{s}\left(\frac{k}{k_{*}}\right)^{n_{s}-1},$$
 (2.55)

$$k^{3}P_{\gamma}(k) = A_{t}\left(\frac{k}{k_{*}}\right)^{n_{t}}, \qquad (2.56)$$

where A_s , A_t are the amplitudes of the scalar and tensor power spectra at $k = k_*$ and k_* is the reference scale (*pivot scale*), equal to $0.05 \,\mathrm{Mpc}^{-1}$. A_t is usually represented in terms of the *tensor-to-scalar ratio*, $r = \frac{P_{\gamma}(k)}{P_{\zeta}(k)}$. Neglecting the weak dependence on k, we should have $r = -8n_t = 16\epsilon$, as can be seen by examining the power spectra.

Observations of the scalar spectrum are consistent with the above result, while tensor power spectra have not been detected yet, so we only have an upper limit on r (see Table 2.2).

2.3.2 The in-in formalism

The standard method of deriving the statistics of primordial fluctuations is the in-in formalism [58, 59]. Since a late-time cosmological observer cannot perform scattering experiments but only has (indirect) access to the late-time limit of the inflationary universe, observables are limited to the expectation values $\langle O_i(\eta = 0) \rangle$ of operators on a late-time asymptotic state in (quasi) de Sitter spacetime. Let us label the initial (early-time) state of the inflationary universe as $| in \rangle$. This state can then be evolved until the end of inflation. In the interaction picture, the evolved state is

$$\mathcal{T}\exp\left(-i\int_{-\infty}^{0}H^{I}(\eta,\mathbf{x})d^{3}xd\eta\right)\mid\mathrm{in}\rangle,\tag{2.57}$$

We should evaluate the operator of interest on this vacuum state or, equivalently, evaluate the interaction picture operator on the evolved state:

$$\langle \operatorname{in}|\mathcal{O}_{i}(\eta=0)|\operatorname{in}\rangle = \langle \operatorname{in}|\bar{\mathcal{T}}\exp\left(i\int_{-\infty}^{0}H^{I}(\eta,\mathbf{x})d^{3}xd\eta\right).$$
(2.58)

$$\mathcal{O}_{i}^{I}(\eta=0)\mathcal{T}\exp\left(-i\int_{-\infty}^{0}H^{I}(\eta,\mathbf{x})d^{3}xd\eta\right)\left|\mathrm{in}\right\rangle,\qquad(2.59)$$

where \mathcal{T} denotes the time-ordered product, and $\overline{\mathcal{T}}$ - the anti-time ordered product. To compute the above integrals (either up to tree level or including loop corrections), we expand the operator $\mathcal{O}_i^I(\eta = 0)$ and H^I in all relevant fields and use de Sitter mode expansion for these fields. For a scalar ϕ (assuming a constant sound speed $c_s \leq 1$),

$$\phi(\eta; \mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} \left(a^{\dagger}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} f_k^+(\eta) + a(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}} f_k^-(\eta) \right),$$
(2.60)

where a, a^{\dagger} are the annihilation and creation operators, respectively. As before, we identify the initial state with the Bunch-Davies vacuum. The corresponding mode functions for the massless case with $c_s = 1$ are given in (2.37). In a more general case considered here, the mode functions must satisfy the free equation of motion for the scalar field,

$$\phi''(\eta) - \frac{2}{\eta}\phi'(\eta) + \left(c_s^2k^2 + \frac{m^2}{H^2\eta^2}\right)\phi(\eta) = 0,$$
(2.61)

with boundary and asymptotic conditions (2.34) - (2.36). The solutions are

$$f^{+}(k,\eta) = \frac{i\sqrt{\pi}}{2} H e^{-\frac{i\pi}{2}(\nu+1/2)} (-\eta)^{3/2} H_{\nu}^{(2)}(-c_s k\eta), \quad f^{-}(k,\eta) = \left(f^{+}(k,\eta)\right)^*, \quad (2.62)$$

where $\nu = \sqrt{\frac{9}{4} - \frac{m^2}{H^2}}$ [12]. In most cases, these functions are too complicated to be useful, and as a consequence, the final result of the in-in calculation cannot be expressed in terms of elementary functions. However, in cases of special interest, $f^{\pm}(k, \eta)$ are simple. For a massless scalar, we have

$$f^{+}(k,\eta) = \frac{H}{\sqrt{2c_s^3 k^3}} (1 - ic_s k\eta) e^{ic_s k\eta},$$
(2.63)

while for the special value of the mass, $m^2 = 2H^2$, and with $c_s = 1$,

$$f^+(k,\eta) = -\frac{iH}{\sqrt{2k}}\eta e^{ik\eta}.$$
(2.64)

This case is known as a conformally coupled (scalar) field.

2.4 Evolution of perturbations after inflation

The power spectra computed in Section 2.3.1 and the higher-point correlators, which can be computed through the in-in method, describe the statistics of field fluctuations at the end of inflation. However, two practical issues arise as one tries to connect such predictions to observations. First, the curvature perturbation ζ is not observable directly but only through its effect on matter and radiation perturbations. The latter can only be observed on the two-dimensional last scattering surface. Secondly, the seed fluctuations undergo a nontrivial classical evolution once inflation ends. Initially, this evolution is linear, but later it becomes nonlinear, with the nonlinearities becoming particularly strong as the ratio of density fluctuation to average density becomes of order one.

In this short section, I briefly review the problem and present approximate solutions. A more complete introduction can be found in [60].

2.4.1 Density perturbations in the early universe

To describe small density perturbations, it is convenient to define the *density contrast* $\delta(\mathbf{x}, t)$ as

$$\delta(\mathbf{x},t) \equiv \frac{\delta\rho(\mathbf{x},t)}{\bar{\rho}(t)}, \qquad (2.65)$$

where $\bar{\rho}$ is the mean mass density and $\delta \rho = \rho - \bar{\rho}$ is the absolute density perturbation. We can define analogous quantities for each component of the universe separately, $\delta_a \equiv \delta \rho_a / \bar{\rho}_a$. We are especially interested in the radiation density contrast δ_r , which is directly related to the temperature of CMB, and the total matter density contrast δ_m , which is related to the galaxy density distribution.

Let us begin by studying the influence of gravity. Every particle experiences the same gaugeinvariant gravitational potential⁸ Φ . The evolution equation for the total density contrast δ is

$$ds^{2} = a(\eta)^{2} \left(-(1+2A)d\eta^{2} + 2\partial_{i}Bdx^{i}d\eta + \left[(1+2C)\delta_{ij} + 2(\partial_{i}\partial_{j} - \frac{1}{3}\nabla^{2})E \right] dx^{i}dx^{j} \right),$$
(2.66)

the potential Φ is defined as

$$\Phi = -C - \mathcal{H}(B - E') + \frac{1}{3}\nabla^2 E,$$
(2.67)

where $\mathcal{H} = \frac{a'}{a}$.

⁸Writing the perturbed metric as

$$\nabla^2 \Phi - 3\mathcal{H}(\Phi' + \mathcal{H}\Phi) = \frac{3}{2}\mathcal{H}^2\delta, \qquad (2.68)$$

where $\mathcal{H} = a^{-1}(\eta)\partial_{\eta}a(\eta)$ is the comoving Hubble parameter.

2.4.2 Scalar curvature and density contrasts

When inflation ended, the curvature power spectrum was given by (2.39), with higher moments of scalar curvature perturbations given by the respective primordial correlators. As a first step, these perturbations can be related to fluctuations of the potential Φ . On superhorizon scales, ζ is constant, and we have [61, 54]

$$\Phi = \frac{3+3w}{5+3w}\zeta, \qquad (2.69)$$

where $P = w\rho$. Thus, we have $\Phi_{RD} = \frac{2}{3}\zeta$, $\Phi_{MD} = \frac{3}{5}\zeta$. Taking (2.68) and neglecting all derivatives, we find that the density contrast is

$$\delta = -2\Phi. \tag{2.70}$$

At the end of inflation, the universe is radiation-dominated, and δ can be identified with δ_r . (For adiabatic perturbations, it follows that $\delta_m = \frac{3}{4}\delta_r \approx -\frac{3}{2}\Phi$.) We then need to evolve δ_r until recombination, i.e. the moment of the last scattering. δ_r is constant on superhorizon scales, but evolves as $\cos(k\eta/\sqrt{3})$ on subhorizon scales. Under the approximations made here, the power spectrum of radiation density contrast δ_r is given by [61]

$$P_r(k) \propto \begin{cases} k^{n_s - 4} & (k < k_*), \\ k^{n_s - 4} \cos^2\left(\frac{k\eta_*}{\sqrt{3}}\right) & (k < k_*). \end{cases}$$
(2.71)

where $k_* = 1/\eta_*$ is the wavenumber of the mode that entered the horizon at the time of matterradiation equality η_* . The temperature of radiation is related to δ_r via

$$\frac{\delta T}{T} = \frac{1}{4}\delta_r.$$
(2.72)

However, this is *not* the observed temperature of CMB since, after photons decoupled from matter, they still underwent nontrivial dynamics, which we need to take into account. An important effect is the redshift of photons emitted from a well of a gravitational potential Φ , and the related Sachs-Wolfe effect [62, 54], which give

$$\frac{\delta T_{G+SW}}{T} = \frac{1}{3}\Phi.$$
(2.73)

Other effects that impact the statistics of the observed CMB include baryon acoustic oscillations [63] (in the above treatment, I neglected the coupling of photons to baryons), the remaining Sachs-Wolfe effect, which occurs due to the CMB photons passing through evolving density perturbations [64], gravitational lensing [65] and Sunayev-Zel'dovich effect [66]. It must also be noted that the last scattering surface is two-dimensional; temperature fluctuations on that surface are related to the three-dimensional statistics via

$$\frac{\langle \delta T \delta T \rangle}{T^2} = \sum_l \frac{2l+1}{4\pi} C_l P_l(\cos \theta), \qquad (2.74)$$

$$C_{l} = 16\pi T^{2} \int dk \, k^{2} \left\langle \left(\frac{1}{3}\Phi_{\mathbf{k}} + \frac{1}{4}\delta_{r,\mathbf{k}}\right) \left(\frac{1}{3}\Phi_{-\mathbf{k}} + \frac{1}{4}\delta_{r,-\mathbf{k}}\right) \right\rangle' j_{l}^{2}(kr) \,.$$
(2.75)

2.4.3 Linearised evolution of matter

In the first approximation - that of a linear evolution - the density contrast of each component at time t will be related to the density contrast at the end of inflation t_{in} through a function of the comoving wavenumber k:

$$\delta_i(\mathbf{k}, t) = T_i(k; t, t_{in})\delta_i(\mathbf{k}, t_{in}).$$
(2.76)

For example, if only one component dominates, then in the Newtonian approximation, the density perturbation δ satisfies the linearised evolution equation

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} + c_s^2 \left(\frac{k^2}{a^2} - 4\pi G\bar{\rho}/c_s^2\right)\delta_{\mathbf{k}} = 0.$$
(2.77)

where $c_s^2 = \frac{\delta P}{\delta \rho}$. The transfer function should therefore satisfy

$$\partial_t^2 T_i(k;t,t_{in}) + 2H\partial_t T_i(k;t,t_{in}) + c_s^2 \left(\frac{k^2}{a^2} - 4\pi G\bar{\rho}/c_s^2\right) T_i(k;t,t_{in}) = 0.$$
(2.78)

In a matter-dominated universe, taking $c_s = 0$ (which corresponds to the case of pressureless dust), the transfer function satisfies $T(k) \sim t^{2/3} \sim a$. The perturbations, therefore, grow proportionally to the scale factor. This represents the leading order approximation for the evolution of matter density perturbations, which is valid in the regime where such perturbations are small and where matter pressure can be neglected, as is the case on all but the smallest length scales.

When the approximation $\delta \ll 1$ becomes violated in the late universe, nonlinear effects become significant. Indeed, as matter coalesces into structures such as galaxies and galaxy groups, density contrast becomes of order 1 or larger, especially on small scales. Thus, the distribution of matter and galaxies in the late universe should not be expected to conform to the predictions of (2.76) on all scales. Furthermore, the nonlinear evolution implies that even purely Gaussian primordial perturbations will generate higher moments of density distribution over time, complicating the reconstruction of higher-order primordial correlators from late time data.

There are generally two ways to improve the prediction of (2.76): extend perturbation theory to include nonlinear effects order by order [67–72] or run computationally expensive gravitational N-body simulations [73, 74]. Improving the accuracy of both methods constitutes an active area of research, but a further review is beyond the scope of this thesis.

Chapter 3

Review of bootstrap methods

The purpose of this chapter is to introduce the main methods used in the remainder of this thesis. In Section 3.1, I introduce the spinor-helicity formalism, a representation of massless particle kinematics that has proven useful in bootstrapping elementary processes [75]. In Section 3.2, I discuss the flat space S-matrix bootstrap which is employed and generalized in Chapters 6 and 7. In Sections 3.3, 3.4 and 3.5 I introduce the cosmological wavefunction formalism along with two recently formulated methods of bootstrapping the wavefunction [33, 12]. Finally, in Section 3.6 I discuss adiabatic modes and the related idea of cosmological soft theorems, which can also be used as a bootstrap method. This chapter constitutes a review and does not include any original work.

3.1 Spinor-helicity formalism

Note: in this and the following section, I use the mostly-minus metric signature convention (+ - -), which is widely used in particle physics. The same convention is used in Chapters 6 and 7. In cosmology, the more common convention is (- + ++) which is the one used in Chapter 8.

Spinor helicity formalism [75] is a convenient representation of a lightlike four-vector $p_{\mu} = (p, \mathbf{p})$ as a *spinor pair* $(\lambda, \tilde{\lambda})$. To that end, I begin by defining a 2 × 2 matrix associated with p_{μ} ,

$$p_{\alpha\dot{\alpha}} = (\sigma^{\mu})_{\alpha\dot{\alpha}} \, p_{\mu} \,, \tag{3.1}$$

where $\sigma^{\mu} = (1, \sigma), \bar{\sigma}^{\mu} = (1, -\sigma); \sigma^{i}$ being the Pauli matrices,¹

¹The matrices $\sigma^{\mu}_{\alpha\dot{\alpha}}$ coincide with the identity and the Pauli matrices - note that the $\alpha, \dot{\alpha}$ indices are **down**. The matrices $\bar{\sigma}^{\mu,\alpha\dot{\alpha}}$ - with indices **up** - coincide with 1 and minus the Pauli matrices.

$$\sigma^{1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^{2} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^{3} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$
(3.2)

Importantly, the matrix $p_{\alpha\dot{\alpha}}$ is degenerate,

$$\det\left(p_{\alpha\dot{\alpha}}\right) = p^{\mu}p_{\mu} = 0, \qquad (3.3)$$

which implies that it can be written as a product of two vectors, known as *spinors*: $p_{\alpha\dot{\alpha}} = \lambda_{\alpha}\tilde{\lambda}_{\dot{\alpha}}$. Each spinor has 2 (complex) components, but a simultaneous rescaling of both leaves p^{μ} invariant, so only 3 components are physical. Using the relation $\bar{\sigma}_{\nu}^{\dot{\alpha}\alpha}\sigma_{\alpha\dot{\alpha}}^{\mu} = 2\delta_{\nu}^{\mu}$, we can invert (3.1) and write p^{μ} as

$$p_{\mu} = \frac{1}{2} \left(\bar{\sigma}_{\mu} \right)^{\alpha \dot{\alpha}} \lambda_{\alpha} \tilde{\lambda}_{\dot{\alpha}} , \qquad (3.4)$$

 λ and $\tilde{\lambda}$ are sometimes referred to as the *holomorphic* and *anti-holomorphic* spinors, respectively, owing to their transformation properties under the Lorentz group: λ transforms in the (1/2, 0), while $\tilde{\lambda}$ transforms in the (0, 1/2) representation of the SL $(2, \mathbb{C})$.

Any scalar observable in a Lorentz symmetric theory must be a function of only Lorentz invariant objects. If a set of such objects captures all asymptotic data of a scattering process (up to spacetime symmetries), then this set includes all variables that a scattering amplitude can possibly depend on. Certainly, the set of all spinor pairs, supplemented with particle spin and all the necessary quantum numbers that identify particle species and carry information about their charges, is sufficient to determine the amplitude. In fact, any Lorentz invariant can be constructed from just two basic objects, known as spinor-helicity brackets [28, 29]. These are defined as

$$\langle \lambda_i \lambda_j \rangle \equiv \epsilon^{\alpha \beta} \lambda_{i,\alpha} \lambda_{j,\beta} , \qquad (3.5)$$

$$[\tilde{\lambda}_i \tilde{\lambda}_j] \equiv \epsilon^{\dot{\alpha}\dot{\beta}} \tilde{\lambda}_{i,\dot{\alpha}} \tilde{\lambda}_{j,\dot{\beta}}.$$
(3.6)

The *i*, *j* indices label distinct particles, while the Greek letters refer to spinor indices. Usually, one uses a shorthand notation, $\langle ij \rangle = \langle \lambda_i \lambda_j \rangle$ and $[ij] = [\tilde{\lambda}_i \tilde{\lambda}_j]$. The angle and square brackets are antisymmetric,

$$\langle ji \rangle = -\langle ij \rangle, \quad [ji] = -[ij].$$
 (3.7)

The scalar product of two four-vectors p_1 and p_2 can be written in terms of the brackets as

$$p_{1}^{\mu}p_{2,\mu} = \frac{1}{4} \left(\bar{\sigma}^{\mu}\right)^{a\dot{a}} \left(\bar{\sigma}_{\mu}\right)^{b\dot{b}} \lambda_{1,a}\tilde{\lambda}_{1,\dot{a}}\lambda_{2,b}\tilde{\lambda}_{2,\dot{b}}$$
(3.8)

$$= \frac{1}{2} \epsilon^{ab} \epsilon^{\dot{a}\dot{b}} \lambda_{1,a} \tilde{\lambda}_{1,\dot{a}} \lambda_{2,b} \tilde{\lambda}_{2,\dot{b}} = \frac{1}{2} \langle 12 \rangle [12] \,. \tag{3.9}$$

In the above derivation I used the identity $(\bar{\sigma}^{\mu})^{a\dot{a}} (\bar{\sigma}_{\mu})^{b\dot{b}} = 2\epsilon^{ab}\epsilon^{\dot{a}\dot{b}}$ [75].

Below I list other standard identities of spinor-helicity formalism [28]. Recall that in this section I only discuss massless particles on Lorentz invariant backgrounds. In Chapter 8, we discuss a generalization to a cosmological scenario where the zero component of particle momentum (the "energy") is not conserved.

$$\sum_{j} p_{j,\mu} = 0 \quad \Rightarrow \quad \sum_{j} \langle ij \rangle [jk] = 0.$$
(3.10)

Schouten identity (for any three spinors $\lambda_i, \lambda_j, \lambda_k$),

$$\langle ij\rangle\lambda_k + \langle jk\rangle\lambda_i + \langle ki\rangle\lambda_j = 0, \qquad (3.11)$$

Similarly for $\tilde{\lambda}_{i,j,k}$:

$$[ij]\tilde{\lambda}_k + [jk]\tilde{\lambda}_i + [ki]\tilde{\lambda}_j = 0.$$
(3.12)

In the case of four-particle kinematics, Mandelstam variables can be expressed as follows:²

$$s = (p_1 + p_2)^{\mu} (p_1 + p_2)_{\mu} = \langle 12 \rangle [12] = \langle 34 \rangle [34], \qquad (3.13)$$

$$t = (p_1 + p_3)^{\mu} (p_1 + p_3)_{\mu} = \langle 13 \rangle [13] = \langle 24 \rangle [24], \qquad (3.14)$$

$$u = (p_1 + p_4)^{\mu} (p_1 + p_4)_{\mu} = \langle 14 \rangle [14] = \langle 23 \rangle [23].$$
(3.15)

Note that

$$s + t + u = 0. (3.16)$$

²In many texts, the definitions of t and u are interchanged. I follow the conventions of [76] and [77].

The little group

The little group is a subgroup of the Lorentz symmetry group that leaves the four-momentum p^{μ} unchanged [78]. Suppose $|\mathbf{p}, \sigma\rangle$ is a quantum state of a single particle with a specific momentum \mathbf{p} , such that the index σ corresponds to all information about the particle beyond the momentum itself. Then, any element of the little group acts linearly on single-particle states, moving them within the space of states with the same momentum:

$$|\mathbf{p},\sigma\rangle \mapsto D(\sigma,\sigma')|\mathbf{p},\sigma'\rangle.$$
 (3.17)

For massless particles, the little group is isomorphic to ISO(2), the isometry group of 2-dimensional Euclidean space. In nature, all elementary particles are observed to transform trivially under the translation component of this group - the presence of a continuous quantum number that would correspond to the translation subgroup is simply inconsistent with experimental evidence. Thus, once divided by the translation subgroup, the little group for massless particles reduces to SO(2). We should then allow a state to be mapped to minus itself under a full rotation (one-particle states are representations of the projective group), as is the case for particles with half-integer angular momentum. The eigenvalue of a particle state under the little group transformation must therefore be an integer or a half-integer, $h = 0, \pm \frac{1}{2}, \pm 1, \ldots$. This quantum number is known as *helicity* and is equal to the projection of the spin onto the direction of spatial momentum:

$$h = \mathbf{S}.\hat{\mathbf{p}}.\tag{3.18}$$

The little group transformation, also known as the helicity transformation, is thus simply a rotation of a state around its momentum vector. It can be represented as a rescaling of the spinors, which does not change the momentum but multiplies a one-particle state by a phase,

$$g(\theta): |\psi\rangle \mapsto e^{ih\theta} |\psi\rangle,$$
 (3.19)

$$g(\theta): (\lambda, \tilde{\lambda}) \mapsto (e^{-i\theta/2}\lambda, e^{i\theta/2}\tilde{\lambda}).$$
 (3.20)

The generator of the helicity transformation, \hat{H} , known as the *helicity operator*, has a simple form when expressed in terms of the spinors:

$$g(\theta) = \exp\left(i\theta\hat{H}\right),$$
 (3.21)

$$\hat{H} = \frac{1}{2} \left(\tilde{\lambda} \frac{\partial}{\partial \tilde{\lambda}} - \lambda \frac{\partial}{\partial \lambda} \right) .$$
(3.22)

The above property entails a powerful constraint on observables such as scattering amplitudes.³ These will be discussed in the next section. Before we continue, we need to find expressions for the circular polarization tensors in terms of spinor-helicity variables.

Polarization tensors

Polarization tensors of massless particles with helicity $\pm h \in \mathbb{Z}$ have rank h and satisfy the following definitional relations,

$$(e^{h})^{\mu}_{\mu\dots}(\mathbf{k}) = k^{\mu}e^{h}_{\mu\nu\dots}(\mathbf{k}) = 0 \qquad (\text{transverse and traceless}), \qquad (3.23)$$

$$e^{h}_{\mu\nu\dots}(\mathbf{k}) = e^{h}_{\nu\mu\dots}(\mathbf{k}) \qquad (\text{symmetric}), \qquad (3.24)$$

$$(e^{h})^{\nu}_{\mu\dots}(\mathbf{k})e^{h}_{\nu\sigma\dots}(\mathbf{k}) = 0 \qquad (\text{lightlike}), \qquad (3.25)$$

$$(e^{h})^{\mu\nu\dots}(\mathbf{k})e^{-h'}_{\mu\nu\dots}(\mathbf{k}) = c_{h}\delta_{hh'} \qquad (\text{normalization}), \qquad (3.26)$$

$$e^{h}_{\mu\nu\dots}(\mathbf{k})^{*} = (-1)^{|h|} e^{-h}_{\mu\nu\dots}(\mathbf{k})$$
 (fields are real). (3.27)

Polarization tensors of higher rank can be represented as tensor products of spin 1 polarizations,

$$(e^{+h})_{\mu_1\dots\mu_h}(\mathbf{k}) = \sqrt{c_h} e^+_{\mu_1}(\mathbf{k}) e^+_{\mu_2}(\mathbf{k}) \dots e^+_{\mu_h}(\mathbf{k}), \qquad (3.28)$$

$$(e^{-h})_{\mu_1\dots\mu_h}(\mathbf{k}) = \sqrt{c_h} e^-_{\mu_1}(\mathbf{k}) e^-_{\mu_2}(\mathbf{k}) \dots e^-_{\mu_h}(\mathbf{k}), \qquad (3.29)$$

where $\sqrt{c_h}$ is a numerical constant that depends on the chosen normalization condition for helicity $\pm h$ polarization tensors, which frequently differs between publications.

In terms of spinor helicity variables, the spin 1 vectors take the form

$$e_{\alpha\dot{\alpha}}^{+} = \sqrt{2} \frac{\eta_{\alpha}\lambda_{\dot{\alpha}}}{\langle\eta\lambda\rangle}, \quad e_{\alpha\dot{\alpha}}^{-} = \sqrt{2} \frac{\lambda_{\alpha}\tilde{\eta}_{\dot{\alpha}}}{[\tilde{\lambda}\tilde{\eta}]}.$$
 (3.30)

The above form of polarization vectors can be justified by verifying the relations (3.23) - (3.27).

• First, I verify the transversality condition:

³Strictly speaking, it's the *differential cross sections* that are observable. They are directly related to the absolute value of the scattering amplitude, the complex phase of which is not observable.

$$p^{\mu}e^{+}_{\mu}(\mathbf{p}) = \frac{1}{2}\sqrt{2}\frac{\langle\lambda\eta\rangle[\tilde{\lambda}\tilde{\lambda}]}{\langle\eta\lambda\rangle} = 0,$$
 (3.31)

$$p^{\mu}e^{-}_{\mu}(\mathbf{p}) = \frac{1}{2}\sqrt{2}\frac{\langle\lambda\lambda\rangle[\lambda\tilde{\eta}]}{[\tilde{\lambda}\tilde{\eta}]} = 0.$$
(3.32)

• Higher helicity polarization tensors should be traceless. This is guaranteed by the following,

$$e^+_{\mu}(\mathbf{p})e^{+,\mu}(\mathbf{p}) = 0,$$
 (3.33)

and similarly for e_{μ}^{-} . This also ensures that these tensors are lightlike (3.25).

- Symmetry of polarization tensors follows directly from the definitions (3.28) (3.29).
- Next, I check that polarization vectors are correctly normalized,

$$(e^{-}(\mathbf{p}))^{\mu}e^{+}_{\mu}(\mathbf{p}) = \frac{\langle \lambda\eta\rangle[\tilde{\eta}\lambda]}{\langle \eta\lambda\rangle[\tilde{\lambda}\tilde{\eta}]} = 1.$$
(3.34)

• Finally, I check the condition (3.27),

$$(e_{\mu}^{+}(\mathbf{k}))^{*} = \sqrt{2} \frac{\eta_{\alpha}^{*}(\tilde{\lambda}_{\dot{\alpha}})^{*}}{\langle \eta^{*} \lambda^{*} \rangle} = \sqrt{2} \frac{\eta_{\alpha}^{*} \lambda_{\dot{\alpha}}}{[\eta^{*} \tilde{\lambda}]}$$
(3.35)

Taking $\eta^* = \lambda$ w.l.o.g., we get

$$(e^+_{\mu}(\mathbf{k}))^* = \sqrt{2} \frac{\lambda_{\alpha} \lambda_{\dot{\alpha}}}{|\lambda \tilde{\lambda}|}$$
(3.36)

which is exactly equal to $-e_{\mu}^{-}(\mathbf{k})$ with $\tilde{\eta} = \lambda$.

I was free to choose η and $\bar{\eta}$ in the last argument because these are not physically meaningful objects. Rather, they are known as *reference spinors* and can be chosen freely as long as η is not proportional to λ and $\tilde{\eta}$ is not proportional to $\tilde{\lambda}$. If we replace η with a linear combination $a\eta + b\lambda$, then we have

$$e_{\alpha\dot{\alpha}}^{+} = \sqrt{2} \frac{(a\eta_{\alpha} + b\lambda_{\alpha})\tilde{\lambda}_{\dot{\alpha}}}{\langle(a\eta + b\lambda),\lambda\rangle} = \sqrt{2} \frac{a\eta_{\alpha}\tilde{\lambda}_{\dot{\alpha}} + b\lambda_{\alpha}\tilde{\lambda}_{\dot{\alpha}}}{a\langle\eta\lambda\rangle} = \sqrt{2} \frac{\eta_{\alpha}\tilde{\lambda}_{\dot{\alpha}} + b\lambda_{\alpha}\tilde{\lambda}_{\dot{\alpha}}}{\langle\eta\lambda\rangle} + \sqrt{2} \frac{b}{a\langle\eta\lambda\rangle} p_{\alpha\dot{\alpha}} \,. \,(3.37)$$

The polarization tensor is thus modified by a term proportional to p^{μ} . Such a term is equivalent to a gauge transformation [78] and, in a consistent theory, should not affect observable quantities. It is also easy to check that the relations (3.23) - (3.27) are still satisfied. The independence of a scattering



Figure 3.1: A Feynman diagram representing one of the contributions to a $2 \rightarrow 3$ scattering amplitude.

amplitude on reference spinors is often a powerful constraint allowing us to bootstrap the amplitude (see Appendix F of [2]).

3.2 S-matrix bootstrap

The primary way to connect particle theory to experiment is to calculate scattering amplitudes for processes that can then be measured in the laboratory setting. In the Feynman approach, one obtains scattering amplitudes by adding up the contributions from various Feynman diagrams, which are calculated using Feynman rules. It must be noted that a Feynman diagram like the one shown in Fig. 3.1 does not represent on-shell particles travelling between interaction vertices.⁴ For this reason, the contribution of a particular Feynman diagram to the amplitude is, in general, not gauge invariant and cannot describe a measurable physical quantity. We obtain a gauge-invariant, physically meaningful result only after the contributions of all Feynman diagrams are summed together.

The contributions of individual Feynman diagrams will often take the form of relatively complicated mathematical expressions. However, after adding up all the Feynman diagrams, one tends to obtain a simple result. All the gauge dependence of individual contributions must also be cancelled out, suggesting that the gauge dependence introduced in the Feynman diagram calculation was avoidable.

Nowhere has this tendency been more apparent than in the case of strong interactions. A long expression for the 6-gluon amplitude derived by Parke and Taylor [80] that had to be evaluated

⁴According to one interpretation, a Feynman diagram does not represent particles at all but is only a convenient way of representing mathematical expressions [79].

numerically led to a compact, one-line result [81], strongly suggesting that there must be a simpler way of reaching it. Indeed, authors of [82] developed a recursive method that uses only on-shell data and does not introduce any unphysical quantities. It allowed [82] to derive the result of [80, 81] much more elegantly and generalize it to any n-gluon amplitude.

In this section, I will review the basics of this on-shell bootstrap, a method which has developed into a significant research field [83]. I will show how the principles of symmetry, little group scaling, unitarity and causality can be used to constrain the interactions in some simple cases. The section serves as an introduction to techniques used in [2, 3]. While [2, 3] apply to boost-violating theories in flat space, in this section I am concerned only with the fully Lorentz invariant scenario.

Consider an early time state $|in\rangle$ that is an eigenstate of the *free* Hamiltonian. We might ask what is the probability amplitude of measuring the system prepared in a state $|in\rangle$ to be in a state $|out\rangle$ at a late time. This can be represented by the S-matrix, defined as a collection of elements

$$S_{\text{in}\to\text{out}} = \langle \text{in}|S|\text{out} \rangle$$
 (3.38)

where S is the evolution operator. In the interaction picture, it can be written in terms of the interaction Hamiltonian $H_{int}(t)$ as

$$S = \lim_{t_1 \to -\infty} \lim_{t_2 \to +\infty} U(t_1, t_2) = \lim_{t_1 \to -\infty} \lim_{t_2 \to +\infty} \mathcal{T} \exp\left(-i \int_{t_1}^{t_2} H_{int}(t) dt\right) .$$
(3.39)

Following the cluster decomposition principle [78], suppose the states $|in\rangle$ and $|out\rangle$ are both separable into products of one-particle states. This is a valid assumption in most contexts, provided that at early and late times the system can be described as a collection of individual, stable particles that are sufficiently separated from each other and do not interact with one another. I define the scattering amplitude - henceforth denoted as A - via

$$S_{\rm in\to out} = \mathcal{A}({\rm in} \to {\rm out})(2\pi)^4 \delta^{(4)} \left(\sum_a p_a^{\mu}\right).$$
(3.40)

Let us return to our treatment of the little group and derive constraints on \mathcal{A} of massless particles from the action of the helicity operator. More concretely, consider \hat{H} acting on one of the incoming particles involved in a scattering process. Then we have

$$\hat{H}_i \langle \text{in} | \text{out} \rangle = h_i \langle \text{in} | \text{out} \rangle$$
, or: $\hat{H}_i \mathcal{A} = h_i \mathcal{A}$. (3.41)

Therefore, we can derive the dependence of \mathcal{A} on the spinors,

$$\frac{1}{2} \left(\tilde{\lambda}_i \frac{\partial}{\partial \tilde{\lambda}_i} - \lambda_i \frac{\partial}{\partial \lambda_i} \right) \mathcal{A} = h_i \mathcal{A}, \tag{3.42}$$

Henceforth, I will treat all particles involved in a scattering process as incoming. This is possible thanks to a general property known as crossing symmetry: an outgoing massless particle with energy E and momentum **p** can be treated as an incoming anti-particle with opposite helicity, energy -Eand momentum $-\mathbf{p}$ ([84], p. 155). This is a simple consequence of the CPT symmetry. Equation (3.42) enables us to fully constrain almost all three-particle amplitudes for massless particles, up to an overall proportionality constant. The only additional ingredient is the requirement that the mass dimension of such amplitude is non-negative, which is guaranteed for local three-particle vertices. Under this assumption, a fully general solution to (3.42) is

$$\mathcal{A}_{3}(\{\lambda^{(i)}, \tilde{\lambda}^{(i)}; h_{i}\}) = \begin{cases} g_{H} \langle 12 \rangle^{h_{3}-h_{1}-h_{2}} \langle 23 \rangle^{h_{1}-h_{2}-h_{3}} \langle 31 \rangle^{h_{2}-h_{3}-h_{1}}, & h \leq 0, \\ g_{AH} [12]^{h_{1}+h_{2}-h_{3}} [23]^{h_{2}+h_{3}-h_{1}} [31]^{h_{3}+h_{1}-h_{2}}, & h \geq 0. \end{cases}$$
(3.43)

As an instructive example, consider a massless spin 1 particle - i.e., a photon. Lorentz invariance and little group scaling would require the three-particle amplitudes to take the following form,

$$\mathcal{A}_3(1^+, 2^+, 3^+) = g_1[12][23][31], \qquad (3.44)$$

$$\mathcal{A}_3(1^+, 2^+, 3^-) = g_2 \frac{[12]^3}{[23][31]}.$$
 (3.45)

However, the particles involved are identical and therefore interchangeable. Since photons have bosonic statistics, the all-plus amplitude should be symmetric under the interchange of any two labels, while the + + - amplitude should be symmetric under the interchange of 1 and 2. Due to the antisymmetry of [ij], neither of the amplitudes has this property unless, of course, $g_1 = g_2 = 0$. We must therefore conclude that Lorentz invariance, together with little group scaling and the bosonic properties of spin 1 particles, forces the cubic interactions of a massless photon to vanish. This can be shown on the level of the Lagrangian as well by requiring gauge symmetry, which then precludes the existence of any nontrivial cubic operator. However, the bootstrap approach becomes a much more practical tool in more complicated examples, as we will soon see.

In order to restore the consistency of three-particle massless amplitudes for spin 1 particles, one could consider several different species of such particles. Let us add an extra label denoted by Latin indices a, b, c, \ldots , representing particle species. In general, cubic interactions involving such spin 1 particles will have the form

$$\mathcal{A}_3(1_a^+, 2_b^+, 3_c^+) = f_{abc}[12][23][31], \qquad (3.46)$$

$$\mathcal{A}_3(1_a^+, 2_b^+, 3_c^-) = g_{abc} \frac{[12]^3}{[23][31]} \,. \tag{3.47}$$

Assuming even parity, the following also holds:

$$\mathcal{A}_3(1_a^-, 2_b^-, 3_c^-) = f_{abc} \langle 12 \rangle \langle 23 \rangle \langle 31 \rangle , \qquad (3.48)$$

$$\mathcal{A}_3(1_a^-, 2_b^-, 3_c^+) = g_{abc} \frac{\langle 12 \rangle^3}{\langle 23 \rangle \langle 31 \rangle} \,. \tag{3.49}$$

The constants f_{abc} , g_{abc} can be referred to as *structure constants* and are not arbitrary. To begin with, we can still interchange a pair of particles without affecting the amplitude, provided we preserve the label of each. Thus, we have

$$\mathcal{A}_3(1_a^+, 2_b^+, 3_c^+) = f_{abc}[12][23][31] = f_{bac}[21][13][32] = \mathcal{A}_3(2_b^+, 1_a^+, 3_c^+), \qquad (3.50)$$

along with analogous relations for other pairs, which implies that f_{abc} must be completely antisymmetric:

$$f_{abc} = -f_{bac} = -f_{acb} = -f_{cba}.$$
(3.51)

We also have

$$\mathcal{A}_{3}(1_{a}^{+}, 2_{b}^{+}, 3_{c}^{-}) = g_{abc} \frac{[12]^{3}}{[23][31]} = g_{bac} \frac{[21]^{3}}{[13][32]} = \mathcal{A}_{3}(2_{b}^{+}, 1_{a}^{+}, 3_{c}^{-}).$$
(3.52)

Implying that g_{abc} must be antisymmetric at least in the first two indices. Unfortunately, an argument of this kind is insufficient to show that g_{abc} must be completely antisymmetric. Nonetheless, the antisymmetry follows from the natural assumption that the + + - amplitude separates into a part

dependent on a, b, c and the part that depends on the helicities. More precisely, writing

$$\mathcal{A}_{3}(1_{a}^{h_{1}}, 2_{b}^{h_{2}}, 3_{c}^{h_{3}}) = e_{\mu}^{h_{1}}(\mathbf{k}_{1})e_{\nu}^{h_{2}}(\mathbf{k}_{2})e_{\sigma}^{h_{3}}(\mathbf{k}_{3})A_{3,abc}^{\mu\nu\sigma}, \qquad (3.53)$$

and assuming that $A_{3,abc}^{\mu\nu\sigma}$ factorises,

$$A_{3,abc}^{\mu\nu\sigma} = g_{abc}B^{\mu\nu\sigma}, \qquad (3.54)$$

for some $B^{\mu\nu\sigma}$ independent of a, b, c, we get

$$\mathcal{A}_3(1_a^+, 3_c^-, 2_b^+) = g_{acb} \frac{[21]^3}{[13][32]}.$$
(3.55)

Equating this to $\mathcal{A}_3(1_a^+, 2_b^+, 3_c^-)$, which represents the same amplitude, we conclude that g_{abc} must be completely antisymmetric [85].

3.2.1 Four-particle amplitudes

Scattering amplitudes involving four or more particles have an additional feature which is a source of powerful constraints not only for many-particle amplitudes but also for three-particle ones. This feature is the factorization property of amplitudes.

The factorization theorem states that when many-particle amplitudes are considered as complex functions of the kinematic data, then all singularities can only be simple poles. Any such simple pole is in a one-to-one correspondence with a *factorization channel*, a limit where an internal leg of a Feynman diagram goes on-shell and can be interpreted as a physical particle travelling a macroscopic distance in space [86]. In this limit, we have

$$\lim_{s_I \to 0} (s_I \mathcal{A}) = \mathcal{A}_L \mathcal{A}_R. \tag{3.56}$$

where $s_I = (\sum_{i \in I} p_I)^2$, while A_L and A_R are the constituent amplitudes of the two disconnected diagrams created by replacing the internal line with two external lines (Fig. 3.2).

Condition (3.56) is a powerful constraint, especially when amplitudes are expressed in spinorhelicity variables. Let us discuss some illustrative examples in the case of tree-level⁵ four-particle

⁵Since the constraint on three-particle amplitudes derived in this section is valid to all orders in perturbation theory, one could work instead with a special class of diagrams with the property that cutting the exchange line separates the diagram



Figure 3.2: In the limit in which the highlighted line is on-shell $(p_I^2 = 0)$, the four-particle amplitude factorizes into a product of $1/p_I^2$ and the two constituent three-particle amplitudes.

amplitudes.

First, consider a spin 1 massless particle with an additional label a, b, c, \ldots , as in the previous subsection. To derive the constraints from factorization, I bootstrap the four-particle (+1, -1, +1, -1) amplitude from the exchange of spin 1. Let $\{A_4\}$ be the *bare mass dimension* of such amplitude - i.e., the mass dimension without taking the coupling constants into account. For exchange diagrams constructed out of (+1, +1, -1) and (-1, -1, +1) vertices and the propagator (whose mass dimension is -2), $\{A_4\}$ is

$$\{\mathcal{A}_4(1_a^-, 2_b^-, 3_c^+, 4_d^+)\} = 2\{\mathcal{A}_3\} - 2 = 0.$$
(3.57)

There can be at most three factorization channels for A_4 , each corresponding to the limit where one of the three Mandelstam variables vanishes:

$$s = (p_1 + p_2)^2, \quad t = (p_1 + p_3)^2, \quad u = (p_1 + p_4)^2.$$
 (3.58)

Therefore, the amplitude \mathcal{A}_4 must take the form

$$\mathcal{A}_4(1_a^-, 2_b^-, 3_c^+, 4_d^+) = \langle 12 \rangle^2 [34]^2 \left(\frac{c_{st}}{st} + \frac{c_{tu}}{tu} + \frac{c_{us}}{us}\right) \,. \tag{3.59}$$

In the s channel, the relation (3.56) entails

into two disconnected components.

$$\mathcal{A}_{4} \sim \sum_{e} g_{abe} g_{ecd} \frac{\langle 12 \rangle^{2} [34]^{2}}{t} = -\sum_{e} g_{abe} g_{ecd} \frac{\langle 12 \rangle^{2} [34]^{2}}{u} \quad \text{as} \quad s \to 0.$$
(3.60)

In the t and u channels, we have

$$\mathcal{A}_4 \sim \sum_e g_{ace} g_{edb} \frac{\langle 12 \rangle^2 [34]^2}{u} = -\sum_e g_{ace} g_{edb} \frac{\langle 12 \rangle^2 [34]^2}{s} \quad \text{as} \quad t \to 0.$$
(3.61)

$$\mathcal{A}_{4} \sim \sum_{e} g_{ade} g_{ebc} \frac{\langle 12 \rangle^{2} [34]^{2}}{s} = -\sum_{e} g_{ade} g_{ebc} \frac{\langle 12 \rangle^{2} [34]^{2}}{t} \quad \text{as} \quad u \to 0.$$
(3.62)

Hence

$$c_{st} - c_{us} = \sum_{e} g_{abe} g_{ecd} , \qquad (3.63)$$

$$c_{tu} - c_{st} = \sum_{e} g_{ace} g_{edb} , \qquad (3.64)$$

$$c_{us} - c_{tu} = \sum_{e} g_{ade} g_{ebc} \,. \tag{3.65}$$

This system of equations has a solution for $\{c_{xy}\}$ if and only if the sum of the right-hand sides is zero:

$$\sum_{e} g_{abe}g_{ecd} + \sum_{e} g_{ace}g_{edb} + \sum_{e} g_{ade}g_{ebc} = 0.$$
(3.66)

Note that this condition is nothing else than Jacobi identity for a set of coefficients that could be identified with the structure constants of some Lie algebra. I have thus shown, using bootstrap methods, that (+1, +1, -1) amplitudes for massless spin 1 fields are consistent only if they arise from an underlying Lie algebra, defined by the commutation relations of the generators, $[T^a, T^b] = Cg_{abc}T^c$ (for some constant C). If this Lie algebra is taken to be SU(N), we obtain the Yang-Mills theory of strong interactions at cubic order.

The four-particle amplitude must take the form

$$\mathcal{A}_4(1_a^-, 2_b^-, 3_c^+, 4_d^+) = \langle 12 \rangle^2 [34]^2 \left(\frac{\sum\limits_e g_{abe}g_{ecd}}{st} - \frac{\sum\limits_e g_{ade}g_{ebc}}{tu} \right) + \mathcal{A}_{4, \text{regular}}, \quad (3.67)$$

where the final term represents the contribution that is regular in all the Mandelstam variables. Such a

term can originate from a contact diagram via quartic interactions that I have not covered in the above analysis. This contribution cannot have any poles but should still be proportional to $\langle 12 \rangle^2 [34]^2$ to give the correct helicity scaling of the amplitude. Thus, we conclude that $\mathcal{A}_{4,\text{regular}}$ must be zero if its bare mass dimension is at most 3 - meaning that we have fully bootstrapped $\mathcal{A}_4(1_a^-, 2_b^-, 3_c^+, 4_d^+)$ in theories where quartic operators have at most 3 derivatives, finding only one consistent form of this four-particle amplitude.

The consistent factorization technique that was covered in this section and the related technique of BCFW deformations introduced and used in Chapter 7 have been used to derive other profound results. In many cases, the IR properties of Lorentz invariant theories are simple consequences of symmetries, unitarity and causality. For example, if two copies of a massless particle are coupled to one graviton in the IR, the coupling strength must be the same as the strength of the graviton self-interaction. This is, of course, the equivalence principle, and we generalize this result in Chapter 6. Other notable results include (i) an observation that (+S, +S, -S) amplitudes must vanish for spin $S \ge 3$, (ii) a derivation of GR as the unique low-energy theory of a massless spin-2 particle (see Chapter 7), and (iii) a proof that the existence of massless spin-3/2 particles leads to supergravity. Chapters 6 and 7 extend the S-matrix bootstrap presented here to a certain class of theories, defined therein, that break Lorentz boost invariance. In Chapter 8, we apply the spinor-helicity formalism in the context of cosmology, and by combining it with cosmological bootstrap techniques, we derive a complete set of primordial graviton non-Gaussianities in the slow-roll approximation. Rather than directly constraining the non-Gaussianities, Chapter 8 uses the cosmological wavefunction formalism introduced in the next section.

3.3 The wavefunction formalism

The cosmological wavefunction formalism is a Schrodinger picture approach to cosmological correlators that presents a convenient alternative to the in-in calculations presented in Section 2.3.2. Here I give a basic introduction to this formalism, along the lines of Appendix A of [12], which may be consulted by the reader for further details.

The quantum wavefunction is essentially a state vector⁶ describing a pure state of a quantum system, represented in a specific basis of the Hilbert space. For example, a particle moving in

⁶Depending on one's favoured interpretation of quantum mechanics, the wavefunction may describe the *true* state of a system or merely represent an observer's *knowledge* about the system [87].

one dimension, described by a state $|\Psi(t)\rangle$, has the corresponding spatial representation of the wavefunction,

$$\Psi(t,x) = \langle x | \Psi(t) \rangle. \tag{3.68}$$

Where $|x\rangle$ is the (non-normalizable) quantum state which corresponds to a particle localized at x, satisfying $\hat{x}|x\rangle = x|x\rangle$. In analogy with the standard quantum mechanics, the cosmological wavefunction (more properly, the *wavefunctional*) is defined as the projection of the state vector $|\Psi(\eta)\rangle$ onto the basis of field configurations,

$$\Psi[\varphi,\eta] = \langle \varphi,\eta | \Psi(\eta) \rangle \tag{3.69}$$

where φ collectively denotes all degrees of freedom and $|\varphi, \eta\rangle$ represents the field configuration ϕ at time η . The states $|\varphi, \eta\rangle$ and $|\Psi(\eta)\rangle$ in the above relation are in the Heisenberg picture, meaning that the time evolution has been absorbed into the operators. Note that in our notation $|\varphi, \eta\rangle \neq |\varphi, \eta_0\rangle$ in general, since the former is to be interpreted as the time-independent state corresponding to a field configuration φ at time η , while the latter describes the same field configuration at η_0 .

Now, from (3.69), we have

$$\Psi[\bar{\varphi},\eta_0] = \langle \bar{\varphi},\eta_0 | \Psi(\eta_0) \rangle = \langle \bar{\varphi},\eta_0 | \Psi(\eta) \rangle$$
(3.70)

$$= \int d\varphi \langle \bar{\varphi}, \eta_0 | \varphi, \eta \rangle \langle \varphi, \eta | \Psi(\eta) \rangle .$$
(3.71)

It can then be shown (see Appendix A of [12]) that

$$\langle \bar{\varphi}, \eta_0 | \varphi, \eta \rangle = \int_{\varphi(\eta) = \varphi, \varphi(\eta_0) = \bar{\varphi}} \mathcal{D}\varphi \int \frac{\mathcal{D}\pi}{(2\pi)^n} \exp\left(i\left(\int_{\eta}^{\eta_0} \varphi' \pi - H_{\eta,\eta_0}(\varphi, \pi)\right)\right).$$
(3.72)

where π denotes all the momenta conjugate to the φ fields, and n is the number of degrees of freedom (number of fields) at each point in space. The action S_{η,η_0} is evaluated between the times indicated in the subscript. We wish to take the limit $\eta \to -\infty(1 - i\epsilon)$, and the early time vacuum is chosen such that $\langle \varphi, \eta | \Psi(\eta) \rangle$ does not depend on φ in this limit, i.e. it is a constant. Then

$$\Psi[\bar{\varphi},\eta_0] = \mathcal{N}\lim_{\eta \to -\infty(1-i\epsilon)} \int_{\varphi(\eta_0) = \bar{\varphi}} \mathcal{D}\varphi \int \frac{\mathcal{D}\pi}{(2\pi)^n} \exp\left(i\left(\int_{\eta}^{\eta_0} \varphi'\pi - H_{\eta,\eta_0}(\varphi,\pi)\right)\right). \quad (3.73)$$

Thus, we see that the cosmological wavefunction can be computed in terms of the path integral. $\Psi[\bar{\varphi}, \eta_0]$ can be written as a formal series expansion in powers of $\bar{\varphi}$,

$$\Psi[\bar{\varphi},\eta_0] = \Psi[0] \exp\left(-\sum_{n=2} \frac{1}{n!} \int \prod_{i=1}^n \left(\frac{d^3 k_i}{(2\pi)^3} \bar{\varphi}_{\mathbf{k}_i}\right) (2\pi)^3 \delta^{(3)}\left(\sum \mathbf{k}_i\right) \psi_n(\{k_i\},\{\mathbf{k}_i\})\right). (3.74)$$

Note that the ψ_n coefficients *do not* include the momentum-conserving delta function.

How can we calculate the ψ_n for a given theory? In perturbation theory, the expansion (3.74) can be considered an expansion around the Gaussian state. Terms with n > 2 would then be treated as *small*. A perturbative solution can then be formulated using Feynman rules for wavefunction coefficients. We start by equating (3.73) to (3.74) and taking the semiclassical limit,

$$\log \Psi[0] - \sum_{n=2} \frac{1}{n!} \int \prod_{i=1}^{n} \left(\frac{d^3 k_i}{(2\pi)^3} \right) \tilde{\delta}^{(3)} \left(\sum \mathbf{k}_i \right) \psi_n(\{k_i\}, \{\mathbf{k}_i\}) \bar{\varphi}_{\mathbf{k}_1} \dots \bar{\varphi}_{\mathbf{k}_n} \approx i S[\varphi_{cl}], \quad (3.75)$$

where $\tilde{\delta}_{\Sigma \mathbf{q}_i}^{(3)} = (2\pi)^3 \delta^{(3)}(\Sigma \mathbf{q}_i)$, $S[\varphi] = \int_{\eta}^{\eta_0} (\varphi' \pi(\varphi, \varphi') - H_{\eta, \eta_0}(\varphi, \pi(\varphi, \varphi')))$ and φ_{cl} is the solution to the full equations of motion with the boundary condition $\varphi_{cl}(\eta_0) = \bar{\varphi}$. The wavefunction coefficients can be obtained by evaluating

$$\psi_n(\{k_i\}, \{\mathbf{k}_i\}) = -i \frac{\delta^n S[\varphi_{cl}]}{\delta \bar{\varphi}_{\mathbf{k}_1} \dots \delta \bar{\varphi}_{\mathbf{k}_n}}.$$
(3.76)

This expression can be evaluated perturbatively using a diagrammatic representation analogous to flat space Feynman diagrams. There are two relevant types of propagators: the bulk-to-boundary propagator K and the bulk-to-bulk propagator G, which satisfy

$$\mathcal{O}(k,\eta)K(k,\eta) = 0, \tag{3.77}$$

$$\mathcal{O}(k,\eta)G(k;\eta,\eta') = -\delta(\eta-\eta'), \qquad (3.78)$$

where $\mathcal{O}(k,\eta)$ is a shorthand notation for the linearized equations of motion. For a scalar field φ ,

$$\mathcal{O}(k,\eta)\varphi = \frac{\partial}{\partial\eta}\frac{\delta\mathcal{L}_2}{\delta\varphi'} - \frac{\delta\mathcal{L}_2}{\delta\varphi}.$$
(3.79)

The boundary conditions are

$$\lim_{\eta \to \eta_0} K(k,\eta) = 1, \qquad \qquad \lim_{\eta \to -\infty(1-i\epsilon)} K(k,\eta) = 0, \qquad (3.80)$$

$$\lim_{\eta,\eta'\to\eta_0} G(k;\eta,\eta') = 0, \qquad \lim_{\eta,\eta'\to-\infty(1-i\epsilon)} G(k;\eta,\eta') = 0.$$
(3.81)

If $f_k^+(\eta)$ is the positive-frequency mode function for the field under consideration, then the solutions are

$$K(k,\eta) = \frac{f_k^+(\eta)}{f_k^+(\eta_0)},$$
(3.82)

$$G(k;\eta,\eta') = 2P(k) \left[\text{Im} \, K(k,\eta) K(k,\eta') \theta(\eta-\eta') + K(k,\eta) \text{Im} \, K(k,\eta') \theta(\eta'-\eta) \right] \,.$$
(3.83)

By careful examination of the relevant formulas, we obtain the following Feynman rules for wavefunction coefficients [14]:⁷

- Draw a diagram and assign momenta \mathbf{k}_a to each external line (flowing out of the diagram) and momenta \mathbf{p}_b (in an arbitrary direction) to each internal line. Assign a conformal time η_A to each vertex.
- For every external line: include a factor of $K(k_a, \eta_A)$ (the bulk-to-boundary propagator) for an external line attached to the vertex A.
- For every internal line: include a factor of $G(p_b; \eta_A, \eta_B)$ (the bulk-to-bulk propagator) for an internal line with momentum \mathbf{p}_b connecting vertices A and B.
- For every vertex: include a factor of +ip_{b,i} for every spatial derivative acting on a leg with momentum p_b flowing *out* of the vertex, and a factor of -ip_{b,i} if the momentum is flowing *into* the vertex. For every conformal time derivative in the interaction operator, include ∂_{η_A} acting on the propagator of the appropriate line. Finally, include the coupling constant of the operator, an appropriate power of a(η_A) and the momentum (but not energy) conserving delta function δ⁽³⁾ (∑ σ_ip_i), where σ_i ∈ {-1, 1} account for the direction of momentum. Sum over all permutations that are consistent with the structure of the diagram.

$$\mathcal{O}(k,\eta)G(k;\eta,\eta') = i\delta(\eta-\eta'),$$

⁷Some authors use an alternative convention in which $G(k; \eta, \eta')$ is defined by

In this alternative convention, the Feynman rules are slightly modified: each vertex carries a factor of i, while every bulk-to-bulk propagator includes an extra factor of -i. Since V - I = 1 - L, both conventions give the same result for ψ_n .

- Integrate over all internal momenta \mathbf{p}_b .
- Include a factor i^{1-L} , where L is the number of loops in the diagram.
- Finally, integrate over all internal momenta \mathbf{p}_b and over all conformal times η_A .

Once wavefunction coefficients are known, they can be used to calculate the correlators at a given conformal time η_0 . Let us define

$$\mathcal{P}_n(\{k_i\}, \{\mathbf{k}_i\}) = \psi_n(\{k_i\}, \{\mathbf{k}_i\}) + \psi_n(\{k_i\}, \{-\mathbf{k}_i\})^*.$$
(3.84)

I will now temporarily suppress dependence on the energies, as these always have the same sign in the expressions that follow and can be determined from the spatial momenta. I also continue to suppress the extra indices that are normally used to label particle species and their spins - it should always be understood that φ , ψ_n and \mathcal{P}_n may carry such labels, and the expressions are summed over those extra indices whenever they are repeated. Assuming a hierarchy $\psi_n(\{\mathbf{k}_a\}) \gg \int_{\mathbf{q}} \frac{\psi_{n+2}(\{\mathbf{k}_a\},\mathbf{q},-\mathbf{q})}{\psi_2(q)}$ for all n and assuming, without loss of generality, the normalization $\Psi[0] = 1$, we have

$$\left\langle \prod_{a=1}^{N} \varphi_{\mathbf{k}_{a}} \right\rangle = \int \left(\prod_{\mathbf{p}} d\varphi_{\mathbf{p}} \right) |\Psi[\gamma, \eta_{0}]|^{2} \prod_{a}^{N} \varphi_{\mathbf{k}_{a}}$$

$$= \int \left(\prod_{\mathbf{p}} d\varphi_{\mathbf{p}} \right) \exp \left\{ -\sum_{n \ge 2} \frac{1}{n!} \int \left(\prod_{i} \frac{d^{3}q_{i}}{(2\pi)^{3}} \varphi_{\mathbf{q}_{i}} \right) \tilde{\delta}_{\Sigma \mathbf{q}_{i}}^{(3)} \psi_{n}(\{\mathbf{q}_{i}\}) + c.c. \right\} \prod_{a=1}^{N} \varphi_{\mathbf{k}_{a}}$$

$$= \int \left(\prod_{\mathbf{p}} d\varphi_{\mathbf{p}} \right) \exp \left\{ -\sum_{n \ge 2} \frac{1}{n!} \int \left(\prod_{i} \frac{d^{3}q_{i}}{(2\pi)^{3}} \varphi_{\mathbf{q}_{i}} \right) \tilde{\delta}_{\Sigma \mathbf{q}_{i}}^{(3)} \left(\psi_{n}(\{\mathbf{q}_{i}\}) + \psi_{n}(\{-\mathbf{q}_{i}\})^{*} \right) \right\} \prod_{a=1}^{N} \varphi_{\mathbf{k}_{a}}$$

$$\approx \int \left(\prod_{\mathbf{p}} d\varphi_{\mathbf{p}} \right) \exp \left\{ -\sum_{n=2}^{N} \frac{1}{n!} \int \left(\prod_{i} \frac{d^{3}q_{i}}{(2\pi)^{3}} \varphi_{\mathbf{q}_{i}} \right) \tilde{\delta}_{\Sigma \mathbf{q}_{i}}^{(3)} \mathcal{P}_{n}(\{\mathbf{q}_{i}\}) \right\} \prod_{a=1}^{N} \varphi_{\mathbf{k}_{a}}$$

$$= \int \left(\prod_{\mathbf{p}} d\varphi_{\mathbf{p}} \right) e^{-\int \frac{d^{3}q}{2(2\pi)^{3}} \mathcal{P}_{2}(q)|\varphi_{\mathbf{q}}|^{2}} \exp \left\{ -\sum_{n=3}^{N} \frac{1}{n!} \int \left(\prod_{i} \frac{d^{3}q_{i}}{(2\pi)^{3}} \varphi_{\mathbf{q}_{i}} \right) \tilde{\delta}_{\Sigma \mathbf{q}_{i}}^{(3)} \mathcal{P}_{n}(\{\mathbf{q}_{i}\}) \right\} \prod_{a=1}^{N} \varphi_{\mathbf{k}_{a}}$$

In the above expressions, I used $\tilde{\delta}_{\Sigma \mathbf{q}_i}^{(3)} = (2\pi)^3 \delta^{(3)}(\Sigma \mathbf{q}_i)$. The final line suggests a straightforward algorithm for computing the N-point function $\langle \prod \varphi_{\mathbf{k}_a} \rangle$ from the wavefunction coefficients ψ_n . This procedure can be formulated in terms of Feynman diagrams constructed for an abstract quantum field theory with coupling constants $\left(-\frac{1}{n!}\mathcal{P}_n\right)$. (Let me emphasize that the associated Feynman rules are distinct from the ones we use to compute ψ_n .) In order to calculate the connected part of the correlator $\langle \prod \varphi_{\mathbf{k}_a} \rangle'$, one should use the following rules:

- Draw a connected diagram with N external lines. To each internal line, assign a momentum \mathbf{p}_i and a particle label σ_i .
- For any *n*-point vertex, write down a factor of $-\frac{1}{n!}\mathcal{P}_n \tilde{\delta}^{(3)}(\sum \mathbf{p}_i)$ and sum over all permutations consistent with the diagram structure (sometimes, this will amount to multiplication by a combinatoric factor).
- For any line internal or external with momentum p write down a factor of $(\mathcal{P}_2(p))^{-1}$.
- Integrate over all internal momenta \mathbf{p}_i .
- Sum over all distinct diagrams.

Results for the lowest values of N in the tree level approximation, for the case of a scalar field ϕ , are

$$\left\langle \phi_{\mathbf{k}}\phi_{-\mathbf{k}}\right\rangle' = \frac{1}{\mathcal{P}_2(k)},$$
(3.86)

$$\left\langle \phi_{\mathbf{k}_{1}}\phi_{\mathbf{k}_{2}}\phi_{\mathbf{k}_{3}}\right\rangle' = -\frac{1}{\prod_{a}\mathcal{P}_{2}(k_{a})}\mathcal{P}_{3}(\{\mathbf{k}_{a}\}), \qquad (3.87)$$

$$\left\langle \phi_{\mathbf{k}_{1}}\phi_{\mathbf{k}_{2}}\phi_{\mathbf{k}_{3}}\phi_{\mathbf{k}_{4}}\right\rangle' = -\frac{1}{\prod_{a}\mathcal{P}_{2}(k_{a})} \left(\mathcal{P}_{4}(\{\mathbf{k}_{a}\}) - \left(\sum_{\sigma}\frac{\mathcal{P}_{3}^{\phi\phi\sigma}(\mathbf{k}_{1},\mathbf{k}_{2},-\mathbf{s})\mathcal{P}_{3}^{\phi\phi\sigma}(\mathbf{k}_{3},\mathbf{k}_{4},\mathbf{s})}{\mathcal{P}_{2}^{\sigma}(s)} + (t,u \text{ channels})\right)\right),$$
(3.88)

where $\mathbf{s} = \mathbf{k}_1 + \mathbf{k}_2$.

3.4 The Manifestly Local Test

Implementing the locality condition within the cosmological bootstrap principles has proven to be a difficult task. Despite numerous efforts, we do not know of any general rule that a late-time inflationary correlator or wavefunction coefficient must satisfy to be consistent with the fundamentally local nature of an inflationary action.

One reason for the difficulty mentioned above is that a seemingly non-local lagrangian could nonetheless describe local physics. Two massive scalar fields ϕ , ψ with a mixed quartic interaction $\phi^2 \psi^2$ are often considered as a very simple example [88]. In a flat space, we have

$$S[\phi,\psi] = \int d^4x \left(\frac{1}{2} \left(\dot{\phi}^2 - (\partial_i \phi)^2 - m^2 \phi^2\right) + \frac{1}{2} \left(\dot{\psi}^2 - (\partial_i \psi)^2 - M^2 \psi^2\right) + \frac{1}{4} \lambda \phi^2 \psi^2\right)$$
(3.89)

Let us integrate out the ψ field and consider the effective action for ϕ , written as $S_{\psi}[\phi]$. This new action will be a valid description on energy scales below the mass of ψ , M, and it is defined by

$$e^{iS_{\psi}[\phi]} = \int \mathcal{D}\psi e^{iS[\phi,\psi]}.$$
(3.90)

We calculate it as follows (we use the mostly-minus signature),

$$e^{iS_{\psi}[\phi]} = \left(\prod_{x} \int d\psi(x)\right) \exp\left\{i \int d^{4}x \left(\frac{1}{2} \left((\partial\phi)^{2} - m^{2}\phi^{2}\right) + \frac{1}{2} \left((\partial\psi)^{2} - M^{2}\psi^{2}\right) + \frac{1}{4}\lambda\phi^{2}\psi^{2}\right)\right\}$$

$$= e^{i \int d^{4}x \frac{1}{2} \left((\partial\phi)^{2} - m^{2}\phi^{2}\right)} \left(\prod_{x} \int d\psi(x)\right) \exp\left\{i \int d^{4}x \frac{1}{2}\psi\left(-\partial^{2} - M^{2} + \frac{1}{2}\lambda\phi^{2}\right)\psi\right\}$$

$$= e^{i \int d^{4}x \frac{1}{2} \left((\partial\phi)^{2} - m^{2}\phi^{2}\right)} \left(\prod_{x} \sqrt{i\pi} \det\left(-\partial^{2} - M^{2} + \frac{1}{2}\lambda\phi^{2}\right)^{-1/2}\right)$$
(3.91)

where the *functional determinant* det A [89] satisfies

$$\ln \det A = \operatorname{tr} \ln A. \tag{3.92}$$

Then

$$S_{\psi}[\phi] = \int d^4x \, \frac{1}{2} \left[-(-(\partial\phi)^2 + m^2\phi^2) + \ln\left(\pi \det\left(-\partial^2 - M^2 + \frac{1}{2}\lambda\phi^2\right)^{-1}\right) \right] \\ = \int d^4x \, \frac{1}{2} \left[-(-(\partial\phi)^2 + m^2\phi^2) - \ln\left(\pi^{-1}\det\left(-\partial^2 - M^2 + \frac{1}{2}\lambda\phi^2\right)\right) \right]. (3.93)$$

Dropping constant terms (independent of ϕ) from the action, we get

$$S_{\psi}[\phi] = \int d^4x \, \frac{1}{2} \left[((\partial \phi)^2 - m^2 \phi^2) - \operatorname{tr} \ln \left(1 + \frac{\lambda}{2} (-\partial^2 - M^2)^{-1} \phi^2 \right) \right].$$
(3.94)

The trace is taken with respect to the x coordinate. The inverse $(-\partial^2 - M^2)^{-1}$ should be understood as the corresponding Green's function, which satisfies

$$(-\partial_x^2 - M^2)G(x, x') = \delta(x - x').$$
(3.95)

The presence of this Green's function in (3.94) indicates that the Lagrangian is non-local: it is a function evaluated at multiple distinct points. Even if we restrict attention to perturbation theory in

the ϕ field (for example, to quartic order in ϕ), nontrivial Green's function does arise, and the action is non-local.

Since the original action (3.89) describing the system was a local one, its physics must be consistent with predictions of locality. This means that although (3.94) has non-local terms, it does, in fact, implicitly describe local physics. In other words, the apparently non-local Lagrangian in (3.94) describes local physics since there exists an explicitly local Lagrangian which, on energy scales below M, is equivalent to it.

Within the context of cosmology, a similar issue arises in ADM formalism. We write the metric as [22]

$$ds^{2} = -Ndt^{2} + h_{ij}(dx^{i} + N^{i}dt)(dx^{j} + N^{j}dt), \qquad (3.96)$$

with *lapse* N and *shift* N^i . We then impose the *unitary gauge* conditions,

$$\delta \phi = 0, \quad h_{ij} = a(t)^2 \left[(1 + 2\zeta) \delta_{ij} + \gamma_{ij} \right].$$
 (3.97)

Here, γ_{ij} is a transverse and traceless tensor. Note that N and Nⁱ are not physical degrees of freedom as they appear in the action without time derivatives, so they can be eliminated using the Euler-Lagrange equations,

$$\frac{\delta L}{\delta N} = 0, \quad \frac{\delta L}{\delta N^i} = 0. \tag{3.98}$$

which take the form of the momentum and hamiltonian constraints of General Relativity [22],

$$\nabla_i \left[N^{-1} \left(E^i_{\ j} - \delta^i_{\ j} E \right) \right] = 0, \tag{3.99}$$

$$R^{(3)} - 2V(\phi) - N^{-2} \left(E_{ij} E^{ij} - E^2 \right) - N^{-2} \dot{\phi}^2 = 0.$$
(3.100)

where

$$E_{ij} = \frac{1}{2} \left(\dot{h}_{ij} - \nabla_i N_j - \nabla_j N_i \right) . \tag{3.101}$$

Due to the presence of spatial derivatives acting on N and N^i , solutions for the lapse and shift will be non-local functions of the physical fields ζ and γ_{ij} . Let us see at what order such apparent nonlocalities play a role. To calculate the action up to cubic order in perturbations, it is sufficient to solve the constraints (3.99) - (3.100) to first order in ζ and γ_{ij} (see [22] and Appendix A of [90]). The error introduced by truncating the expansion of N at linear order is equal to the quadratic error in N itself, multiplied by $\frac{\delta L}{\delta N}$ evaluated on the first order solution, which vanishes to linear order. Hence, the error in the perturbative action will be fourth order in the fields.

Since the expression for N^i includes an inverse Laplacian acting on ζ already at linear order in perturbation theory, terms in the cubic action that include the scalar curvature ζ will, in general, contain powers of ∇^{-2} . However, first-order solutions for N and N^i must be independent of γ_{ij} regardless of the details of the theory. Therefore, the cubic action for the tensor perturbation γ_{ij} does not contain inverse Laplacians.

One might ask if there is a general way to ascertain whether a given non-local description admits a local one equivalent to the former in its regime of validity. Unfortunately, I do not know of such a general method, and it is not very likely that an algorithm producing an answer to this question can be easily constructed. It is most likely impossible to verify whether a given action that contains inverse derivatives can be rewritten - for example, using field redefinitions - as a local one or whether it has a local UV completion.

We call an action *manifestly local* if the Lagrangian is expressed in terms of the fields and a finite number of field derivatives evaluated at one spacetime point. An action is *not manifestly local* if the Lagrangian contains inverse derivative operators⁸ such as the inverse Laplacian ∇^{-2} or an infinite tower of derivative terms. For a manifestly local action, there exists a simple way to encode this locality on the wavefunction coefficients, first noticed in [33]. A massless field in de Sitter spacetime has a bulk-to-boundary propagator

$$K(k,\eta) = (1 - ik\eta)e^{ik\eta}, \qquad (3.102)$$

which satisfies

$$\frac{d}{dk}(K(k,\eta))_{k=0} = (k\eta^2 e^{ik\eta})_{k=0} = 0.$$
(3.103)

Since the wavefunction coefficient can always be expressed as an integral over one of more conformal times, such that the integrand contains factors of all the bulk-to-boundary propagators, we have

⁸These are to be interpreted as Green's functions.

$$\frac{\partial}{\partial k_a} (\psi_n(\{k_a\}, \{\mathbf{k}_a\}, \{\mathbf{p}_b\}, \{\mathbf{p}_b\}))_{k_a=0} = 0.$$
(3.104)

This is because the partial derivative with respect to k_a with all other energies and momenta (both external ones $k_{a'}$ and internal ones p_b), as well as \mathbf{k}_a , kept constant, will act nontrivially only on the bulk-to-boundary propagator $K(k, \eta)$ and we have already established that this derivative will yield zero at $k_a = 0$.

The fact that the cubic graviton action is manifestly local - as noted above - provides an excellent opportunity to use the powerful condition (3.104) to bootstrap cubic wavefunction coefficients for gravitons on a de Sitter background. This is done in Chapter 8, published as [4].

3.5 The Cosmological Optical Theorem

The optical theorem is a standard result in wave scattering theory. In its basic form, it relates the scattering amplitude along the line of impact to the total cross-section of the scatterer. In quantum field theory, it is common to invoke a generalized version of the optical theorem, which is essentially a consequence of the unitarity of time evolution. This is, in turn, equivalent to conservation of probability amplitude,

$$\langle \psi; t | \psi; t \rangle = \langle \psi; 0 | \psi; 0 \rangle . \tag{3.105}$$

I follow [85] for the remainder of the derivation. Since $|\psi; t\rangle = S|\psi; 0\rangle$, unitarity implies that

$$\forall \psi \quad \langle \psi; 0 | \psi; 0 \rangle = \langle \psi; 0 | S^{\dagger} S | \psi; 0 \rangle \Rightarrow S^{\dagger} S = \mathbb{1}.$$
(3.106)

In quantum field theory in Minkowski spacetime, the optical theorem is formulated in terms of scattering amplitudes $\mathcal{A}_{i\to f}$. Let $S = \mathbb{1} + iT$, where T is known as the *transfer matrix*. The scattering amplitude is then defined via

$$\langle f|T|i\rangle = (2\pi)^4 \delta^{(4)}(p_i - p_f) \mathcal{A}_{i \to f}.$$
 (3.107)

 $S^{\dagger}S = 1$ entails a series of relations,

$$i(T^{\dagger} - T) = T^{\dagger}T, \qquad (3.108)$$

$$i\langle f|(T^{\dagger}-T)|i\rangle = \langle f|T^{\dagger}T|i\rangle, \qquad (3.109)$$

$$i\langle i|T|f\rangle^* - i\langle f|T|i\rangle = \sum_{I} \int d\Pi_I \langle f|T^{\dagger}|I\rangle \langle I|T|i\rangle, \qquad (3.110)$$

where we used the resolution of identity, integrating over all the elements of the Hilbert space basis with an appropriate measure $d\Pi_I$

$$\mathbb{1} = \sum_{I} \int d\Pi_{I} |I\rangle \langle I|.$$
(3.111)

Then

$$i\mathcal{A}_{f\to i}^* - i\mathcal{A}_{i\to f} = (2\pi)^4 \sum_I \int d\Pi_I \mathcal{A}_{f\to I}^* \mathcal{A}_{i\to I}, \qquad (3.112)$$

and finally, we have the generalized optical theorem,

$$\mathcal{A}_{i \to f} - \mathcal{A}_{f \to i}^* = i(2\pi)^4 \sum_I \int d\Pi_I \mathcal{A}_{f \to I}^* \mathcal{A}_{i \to I}.$$
(3.113)

Since the above flat space result is a consequence of unitarity and we expect de Sitter time evolution to be unitary as well, one could aim to derive an analogous theorem in the cosmological context. Such a result - the *Cosmological Optical Theorem* (COT) - has indeed been obtained in [12–14]. It encodes the unitarity of time evolution directly on the level of wavefunction coefficients introduced in Section 3.3. Let us review the statement and sketch the proof of this theorem.

The cosmological optical theorem holds under the following assumptions:

- The Bunch-Davies vacuum condition in the far past $(\eta \rightarrow -\infty)$.
- Both external and exchanged particles can have an arbitrary spin, mass and effective speed.
- Interactions must be unitary, but aside from that requirement, they are fully general. In particular, they may break the de Sitter boosts.
- Perturbation theory: the COT is valid perturbatively to any loop order.

We define the analytic continuation of ψ_n^9 onto the complex plane by choosing the branch cut to lie on the negative real axis. We also define

⁹Recall that ψ_n does not include the momentum-conserving δ function nor the factor of $(2\pi)^3$.

$$s = |\mathbf{k}_1 + \mathbf{k}_2|, \quad t = |\mathbf{k}_1 + \mathbf{k}_3|, \quad u = |\mathbf{k}_1 + \mathbf{k}_4|$$
 (3.114)

Then, we have

• For contact diagrams:

$$\psi_n(k_a, \hat{k}_a, \hat{k}_b) + \left[\psi_n(-k_a - i\epsilon, \hat{k}_a, \hat{k}_b)\right]^* = 0, \quad (k_a \in \mathbb{R}^+).$$
(3.115)

• For four-point exchange diagrams:

$$\psi_4(k_a, s) + \left[\psi_4(-k_a - i\epsilon, s)\right]^* =$$

$$= P_{\sigma}(s) \left(\psi^{\phi\phi\sigma}(k_1, k_2, s) - \psi^{\phi\phi\sigma}(k_1, k_2, -s)\right) \left(\psi^{\phi\phi\sigma}(k_3, k_4, s) - \psi^{\phi\phi\sigma}(k_3, k_4, -s)\right) + (t, u).$$
(3.116)

where σ is the particle being exchanged and $P_{\sigma}(s)$ is its power spectrum.

In the case of particles with spin, when we formally flip the sign of energies in (3.115) and (3.116), we keep the momenta and, consequently, the polarization tensors unchanged.

Let us first write a non-perturbative analogue of (3.113) for cosmology. Writing the time evolution operator as $\mathcal{U} = \mathbb{1} + \delta \mathcal{U}$, we get

$$\langle \{\mathbf{k}_{a}, \alpha_{a}\} | \delta \mathcal{U} | 0 \rangle + \langle \{\mathbf{k}_{a}, \alpha_{a}\} | \delta \mathcal{U}^{\dagger} | 0 \rangle = -\int d\Pi_{X} \langle \{\mathbf{k}_{a}, \alpha_{a}\} | \delta \mathcal{U} | X \rangle \langle X | \delta \mathcal{U}^{\dagger} | 0 \rangle, \qquad (3.117)$$

where $\int d\Pi_X |X\rangle \langle X| \equiv \mathbb{1}$ is the resolution of the identity.

I now proceed to the proof of the COT for contact diagrams (3.115). The wavefunction coefficient for a contact diagram is of the form

$$\psi_n(k_a) = -iF \int_{-\infty}^{\eta_0} d\eta a(\eta)^{4-m} \prod_{a=1}^n \frac{d^{s_a}}{d\eta^{s_a}} K(k_a, \eta), \qquad (3.118)$$

where I omitted contractions between momenta and polarization tensors, as well as $\mathcal{O}(1)$ factors and the coupling constant, all of which are inconsequential to the argument; these are included in the real factor F. I keep η_0 general to allow for possible late time divergence of the integral. The integral then converges for Im $\sum k_a < 0$, so analytic continuation of $\psi_n(k_a)$ in the region Im $\sum k_a < 0$ is straightforward. To regularize the integral for $k_a \in \mathbb{R}^+$, I use the $i\epsilon$ prescription $k_a \to k_a - i\epsilon$. The key to the proof of (3.115) is to use a simple property of bulk-to-boundary propagator

$$K(k_a, \eta) = (K(-k_a, \eta))^*.$$
(3.119)

Hence,

$$-(\psi_n(-k_a))^* = -iF \int_{-\infty}^{\eta_0} d\eta a(\eta)^{4-m} \prod_{a=1}^n \frac{d^{s_a}}{d\eta^{s_a}} (K(-k_a,\eta))^* = \psi_n(k_a).$$
(3.120)

Reinstating a proper $i\epsilon$ prescription to guarantee the early time convergence for $k_a \in \mathbb{R}$, we get (3.115).

Let us move on to a sketch of proof of (3.116). Rather than presenting the original proof of [12], I use the method of [14]. The bulk-to-bulk propagator of the exchanged σ field is (Eq. (3.83))

$$G(k;\eta,\eta') = i \Big(f_{\sigma}^{-}(k,\eta) f_{\sigma}^{+}(k,\eta') \theta(\eta-\eta') + f_{\sigma}^{+}(k,\eta) f_{\sigma}^{-}(k,\eta') \theta(\eta'-\eta) - f_{\sigma}^{+}(k,\eta) f_{\sigma}^{+}(k,\eta') \frac{f_{\sigma}^{-}(k,\eta_{0})}{f_{\sigma}^{+}(k,\eta_{0})} \Big).$$
(3.121)

Importantly, this propagator satisfies

$$G(p;\eta,\eta') - G^*(p;\eta,\eta') \equiv 2i \operatorname{Im} G(p;\eta,\eta') = 4i P(p) \operatorname{Im} K(p,\eta) \operatorname{Im} K(p,\eta').$$
(3.122)

The above equation can be easily applied in the case of a tree-level exchange contribution to $\psi_n \equiv \psi_n(k_a, p)$. Recalling that $K(k, \eta) = K(-k, \eta)^*$, (3.122) entails

$$\left(\prod_{a} K(k_{a},\eta_{i})\right) G(p;\eta,\eta') - \left(\left(\prod_{a} K(-k_{a},\eta_{i})\right) G(p;\eta,\eta')\right)^{*} =$$

$$= 4iP(p) \left(\prod_{a} K(k_{a},\eta_{i})\right) \operatorname{Im}K(p,\eta) \operatorname{Im}K(p,\eta')$$

$$= -iP(p) \left(\prod_{a} K(k_{a},\eta_{i})\right) (K(p,\eta) - K(-p,\eta)) (K(p,\eta') - K(-p,\eta')).$$
(3.123)

For an exchange diagram with an L-particle vertex to the left and an R-particle vertex to the right of the internal line, (3.123) directly implies

$$\psi_n(k_a, p) + \psi_n(-k_a, p)^* = P(p) \left(\psi_L(k_a, p) - \psi_L(k_a, -p)\right) \left(\psi_R(k_a, p) - \psi_R(k_a, -p)\right).$$
(3.124)

This gives an alternative proof of the Cosmological Optical Theorem for exchange diagrams (3.116).

The Cosmological Optical Theorem is related to a set of more general relations for cosmological wavefunction coefficients, known as *cutting rules*, which can be derived by starting from (3.123) [13, 14, 91, 92].

3.6 Adiabatic modes and soft theorems

In Section 3.2 we saw how simple consistency conditions that are a consequence of symmetries and physical principles could be used to constrain the flat space S-matrix. There is yet another class of consistency relations - known as *soft theorems* - that follow from the invariance of a theory under generalized gauge transformations that also modify the boundary conditions.¹⁰ In this section, I discuss analogous theorems in the context of inflation. Valid under wide assumptions, soft theorems apply to large classes of models, effectively constraining the space of possible primordial correlators.

Generally speaking, a cosmological soft theorem is a functional relation between an (n + 1)-point correlator (or a linear combination of several such objects [93]), in the limit where the momentum of one of the modes is small (*soft*), and the *n*-point correlator of the remaining fields. One can illustrate the principle by considering the simplest of such theorems, which is the relation between the squeezed bispectrum¹¹ and the power spectrum of the scalar curvature perturbation ζ . For later convenience, I use the following notation for the correlator evaluated with one of the modes fixed,

$$\left\langle \prod \zeta_{\mathbf{k}_{i}} \right\rangle_{\zeta_{\mathbf{p}}=v} := \mathbb{E}\left[\prod \zeta_{\mathbf{k}_{i}} | \zeta_{\mathbf{p}}=v \right].$$
 (3.125)

We can usually use the notation $\langle \prod \zeta_{\mathbf{k}_i} \rangle_{\zeta_{\mathbf{p}}}$ without the risk of introducing any ambiguity. Fields written as a lower index of the expectation value operator $\langle . \rangle$ should be understood as actual values (real numbers), not random variables. Note that by Taylor expanding, we get

$$\left\langle \prod \zeta_{\mathbf{k}_{i}} \right\rangle_{\zeta_{\mathbf{p}}} = \left\langle \prod \zeta_{\mathbf{k}_{i}} \right\rangle_{0} + \zeta_{\mathbf{p}} \frac{\delta}{\delta \zeta_{\mathbf{p}}} \left(\left\langle \prod \zeta_{\mathbf{k}_{i}} \right\rangle_{\zeta_{\mathbf{p}}} \right)_{\zeta_{\mathbf{p}}=0} + \mathcal{O}\left(\zeta_{\mathbf{p}}^{2} \left\langle \prod \zeta_{\mathbf{k}_{i}} \right\rangle \right).$$
(3.126)

¹⁰See Chapter 13 of [78] for a basic introduction.

¹¹That is, the bispectrum in the limit where one of the momenta becomes soft.

3.6.1 Maldacena's soft theorem

Let us apply the expansion (3.126) to the specific case of a squeezed bispectrum. I will treat all the Fourier modes whose momentum is of order $q \ll k$ as part of the background. We have

$$\langle \zeta_{\mathbf{q}} \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \rangle = \int \frac{d^3 q'}{(2\pi)^3} \langle \zeta_{\mathbf{q}} \zeta_{\mathbf{q}'} \rangle \frac{\delta}{\delta \zeta_{\mathbf{q}'}} \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \rangle_{\zeta_{\mathbf{q}'}} + \mathcal{O}(\zeta^5).$$
(3.127)

I will now use the fact that for small q', the long mode $\zeta_{\mathbf{q}'}$ resembles the effect of a time-independent¹² coordinate redefinition, i.e. a diffeomorphism. Even though this diffeomorphism does not vanish for large \mathbf{x} , the physics should remain invariant under such a transformation. Placing the fields on a background of the long mode is therefore equivalent to acting on them with the associated diffeomorphism ϵ ,

$$\langle \zeta_{\mathbf{q}} \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \rangle \approx P_{\zeta}(q) \frac{\delta}{\delta \epsilon} \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \rangle_{\epsilon} = P_{\zeta}(q) \left(\left\langle \frac{\delta \zeta_{\mathbf{k}_1}}{\delta \epsilon} \zeta_{\mathbf{k}_2} \right\rangle + \left\langle \zeta_{\mathbf{k}_1} \frac{\delta \zeta_{\mathbf{k}_2}}{\delta \epsilon} \right\rangle \right).$$
(3.128)

The leading error of the above approximation can be identified with the error introduced when approximating the long mode

$$\zeta(t, \mathbf{x}) = \alpha f(t) e^{i\mathbf{q} \cdot \mathbf{x}}, \qquad (3.129)$$

with the effect of the diffeomorphism

$$\epsilon^i(t, \mathbf{x}) = -\alpha f(t_0) x^i \,, \tag{3.130}$$

$$\zeta_{\epsilon}(t, \mathbf{x}) = -\frac{1}{3}\partial_i \epsilon^i = \alpha f(t_0) \,. \tag{3.131}$$

Once we compute the effects of ϵ^i on the short modes $\zeta_{\mathbf{k}_i}$ [94], we obtain a soft theorem of the form

$$\left\langle \zeta_{\mathbf{q}} \zeta_{\mathbf{k}-\mathbf{q}/2} \zeta_{-\mathbf{k}-\mathbf{q}/2} \right\rangle' \sim -P_{\zeta}(q) \left(3+\mathbf{k}.\partial_{\mathbf{k}}\right) P_{\zeta}(k) + \mathcal{O}\left(q^2/k^2, q^2/(a^2(t)H^2(t))\right) \,. \tag{3.132}$$

Note that the error is estimated to be quadratic in q, which is a consequence of taking a symmetric configuration with the large momenta $\mathbf{k} - \mathbf{q}/2$ and $-\mathbf{k} - \mathbf{q}/2$.

¹²The long mode $\zeta_{\mathbf{q}'}$ is indeed time-independent (*frozen*) in the superhorizon limit ($k \ll H$) in common inflationary scenarios such as ordinary slow-roll ($\epsilon, \tilde{\eta} \approx 0$).
In the above, we saw how an approximate equivalence between a physical perturbation $\zeta(t, \mathbf{x})$ in the long wavelength limit and the effect of a diffeomorphism can be used to derive a nontrivial relationship between an (n+1)-point function in the soft limit and the lower order n-point function. Relations like these can be derived systematically, and all possible soft theorems can be constructed if a complete set of large diffeomorphisms is found. In the following section, I discuss the theory of adiabatic modes, which formalizes the conditions under which the diffeomorphisms can be used to derive soft theorems.

3.6.2 Adiabatic modes

To understand adiabatic modes, one should note that not all perturbations of matter fields and the metric are physical, as some are equivalent to, and can be removed by, a coordinate transformation. The equivalence of two descriptions related by a *gauge* transformation is well known in physics. Perhaps less known is the possibility of constructing a generalized gauge transformation that does not vanish at spatial infinity. Even if we use local conditions to fix the gauge of a given field theory, there usually remains a residual diffeomorphism (*diff* for short) that does not converge to zero at infinity and keeps the local gauge conditions unaffected. This diff may also be referred to as a large gauge transformation (LGT).

An LGT given by $x^{\mu} \mapsto x^{\mu} + \epsilon^{\mu}(x)$ generates a metric perturbation $g_{\mu\nu} = \nabla_{\mu}\epsilon_{\nu} + \nabla_{\nu}\epsilon_{\mu}$, which can be decomposed into scalar, vector and tensor parts. These modes, having been generated by a pure diff, are unobservable (unphysical). *Adiabatic modes* are then defined as physical solutions that, for large wavelengths, are locally equivalent to the effect of a residual diff.¹³ Of course, a mode originating from a diffeomorphism cannot carry entropy, so perturbations that resemble LGTs are indeed adiabatic in the usual sense of the word. Adiabatic modes can be seen as universal, long-wavelength solutions since they are derived independently of the details of the matter or field content.

It must be noted that not all LGTs give rise to adiabatic modes. If there exists a family of physical solutions that mimic the effect of an LGT in the large wavelength limit, such LGT is said to satisfy the *adiabaticity condition*. If, in addition, the effect of the LGT resembles that physical mode which is dominant at late times, we say it satisfies the *physicality condition* [95]. To illustrate the relevance

¹³More precisely: A family of physical perturbations parameterized by α is adiabatic in the limit $\alpha \to 0$ if and only if for all x and t they converge pointwise to a residual diff as $\alpha \to 0$.

of the latter, consider the time evolution of a large wavelength perturbation δf . Since the evolution equation is second order, there are two distinct asymptotic solutions for δf ,

$$\delta f(\mathbf{x}, t) \approx A(t)\delta f_1(\mathbf{x}) + B(t)\delta f_2(\mathbf{x}).$$
(3.133)

If A(t) grows faster than B(t), then the LGT satisfies the physicality condition provided its time dependence matches A(t). We are especially interested in LGTs that satisfy the physicality condition because they are the ones that are well approximated by the dominant physical solution.

Let us now make the presentation more concrete and consider the Newtonian gauge, which eliminates all gauge freedom except for the LGTs. The metric in this gauge is given by

$$ds^{2} = -(1+2\Phi)dt^{2} + a(t)G_{i}dtdx^{i} + a(t)^{2}(1-2\Psi)\delta_{ij}dx^{i}dx^{j} + a(t)^{2}\gamma_{ij}dx^{i}dx^{j}.$$
 (3.134)

An LGT given by $x^{\mu} \mapsto x^{\mu} + \epsilon^{\mu}(x)$ preserves the Newtonian gauge if and only if

$$\nabla^2 \epsilon^i = -\frac{1}{3} \partial_i \partial_k \epsilon^k, \qquad (3.135)$$

$$\nabla^2 \epsilon_0 = -a^2 \partial_k \dot{\epsilon}^k. \tag{3.136}$$

Note that with the boundary conditions $\lim_{|\mathbf{x}|\to\infty} \epsilon^{\mu}(x) = 0$, the only solution would be the trivial one. This expresses the fact that (3.134) completely fixes the local gauge. Instead, we are interested in solutions to (3.135)-(3.136) that do not vanish at infinity. The solution should also satisfy the adiabaticity condition, i.e. it should be a large wavelength limit of a physical adiabatic mode. This condition takes the form of a system of equations [96]

$$\Phi = \Psi, \qquad (3.137)$$

$$\dot{H}\delta u = H\Psi + \dot{\Psi}, \qquad (3.138)$$

$$\dot{G}^i = 0,$$
 (3.139)

$$-4\dot{H}a\delta u_j^V = \nabla^2 G_j. \tag{3.140}$$

In fact, the second equation does not impose any condition on ϵ^{μ} , since we are always free to set $\delta u = \frac{1}{\dot{H}}(H\Psi + \dot{\Psi})$. Moreover, for perfect fluids, the last equation is trivially satisfied at leading order [96]. Writing all the nontrivial constraints in terms of ϵ^{μ} , we get

$$\nabla^2 \epsilon^i = -\frac{1}{3} \partial_i \partial_k \epsilon^k , \qquad (3.141)$$

$$\nabla^2 \epsilon_0 = -a^2 \partial_i \dot{\epsilon}^i \,, \tag{3.142}$$

$$\dot{\epsilon}_0 + H\epsilon_0 = \frac{1}{3}\partial_k \epsilon^k , \qquad (3.143)$$

$$(\partial_t + H) \left(a^2 \dot{\epsilon}^i + \partial_i \epsilon_0 \right) = 0.$$
(3.144)

Equations (3.141)-(3.144) are often solved order by order in x, assuming the expansion

$$\epsilon^{\mu}(x) = \sum_{n} a^{\mu}_{\alpha_1 \alpha_2 \dots \alpha_n} x^{\alpha_1}(t) \dots x^{\alpha_n} .$$
(3.145)

Once a particular solution is known, one should check that it satisfies the physicality condition, i.e. that its time dependence matches the time dependence of the dominant superhorizon mode. This generally depends on the background FLRW dynamics. For now, let us assume that the physicality condition is satisfied and equations (3.141)-(3.144) are solved by $\epsilon^{\mu}(x)$. This LGT generates the following curvature perturbations¹⁴ [97],

$$\delta \mathcal{R} = H\epsilon_0 - \frac{1}{3}\partial_i\epsilon^i + \frac{1}{2}\partial_i\epsilon_0\left(-\partial_i\epsilon_0 + 2H\epsilon_i - \dot{\epsilon}_i\right) - \epsilon^\mu\partial_\mu\mathcal{R}, \qquad (3.146)$$

$$\delta\gamma_{ij} = -\partial_i\epsilon_j - \partial_j\epsilon_i + \frac{2}{3}\delta_{ij}\partial_k\epsilon_k - \epsilon^\mu\partial_\mu\gamma_{ij}.$$
(3.147)

Note the first two terms in \mathcal{R} and the first two in γ_{ij} , which correspond to nonlinear shifts, indicating that the symmetry of an LGT is non-linearly realized. The soft theorems can now be derived using the background wave method introduced in the previous subsection. The background wave should now be identified with the nonlinear parts of (3.146)-(3.147), while the linear part can be used to determine the effect of the LGT on the correlator of n fields, $\frac{\delta}{\delta\epsilon} \langle \mathcal{O} \rangle_{\epsilon}$. There are many alternative methods of deriving soft theorems from adiabatic modes, such as Operator Product Expansion [98], calculating the action of an LGT on the wavefunction [99], or writing a Ward identity (see [93] for details),

$$\langle [Q, \mathcal{O}] \rangle = -i \langle \delta \mathcal{O} \rangle ,$$
 (3.148)

where Q is the Noether charge associated with the symmetry transformation (3.146)-(3.147), i.e. $[Q, \zeta] = -i\delta\zeta, [Q, \gamma_{ij}] = -i\delta\gamma_{ij}.$

¹⁴Note that we are now using the comoving curvature perturbation \mathcal{R} , defined as $\mathcal{R} \equiv -\Psi + H\delta u$, where $\delta u = \nabla^{-2}\partial_i u_i$ and u_i is the fluid velocity.

In this section, we saw how adiabatic modes can be used to derive soft theorems such as (3.132). At the classical level, adiabatic modes constitute a general class of linear solutions to the equations of motion and thus describe primordial perturbations in a model-independent way under a set of mild assumptions (single field inflation, matching time dependence of the long wavelength modes and for tensor adiabatic modes - the absence of anisotropic stress). Adiabatic modes produce general solutions even in the presence of multiple fields or fluids, describing those perturbations of energy densities of each fluid *i* that satisfy the conditions [100]

$$\frac{\delta\rho_i}{\bar{\rho}_i} = \frac{\delta P_i}{\bar{P}_i} = \frac{\delta\rho}{\bar{\rho}} = \frac{\delta P}{\bar{P}}, \qquad (3.149)$$

and are therefore locally equivalent to a time shift. However, in the presence of multiple fields, there exist solutions that cannot be represented as adiabatic modes - even on the smallest scales - since they can be always distinguished from a coordinate transformation. Such solutions are called *isocurvature perturbations* and, remarkably, they have not been detected so far [11].

In Chapter 4, I present a derivation of all scalar and tensor adiabatic modes in a $K = 0^{15}$ universe using the decomposition in terms of vector spherical harmonics [101]. This short, unpublished note is an alternative derivation of the main results of [96], which found a complete set of adiabatic modes in the polynomial basis (3.145).

<u>60</u>

¹⁵Recall that K is the background spatial curvature.

Chapter 4

Adiabatic modes in flat FLRW universe

In this chapter, I find explicit decomposition of adiabatic modes in a flat FLRW universe in terms of vector spherical harmonics [101]. I find general solutions that have a much simpler form than those derived in [96], where expansion in powers of x was used.

4.1 Vector spherical harmonics

Consider a large diffeomorphism $\epsilon^{\mu}(t, \mathbf{x})$. The FLRW spacetime is isotropic and homogenous, so it is a good idea to expand $\epsilon^{i}(t, \mathbf{x})$ in vector spherical harmonics, and $\epsilon^{0}(t, \mathbf{x})$ in scalar spherical harmonics. Vector spherical harmonics (VSH) constitute an orthonormal basis for vector fields on S^{2} , just as scalar spherical harmonics are a basis for scalar fields on S^{2} . Thus, we can indeed expand a generic field ϵ^{i} in VSH Y_{lm}^{i} , Ψ_{lm}^{i} and Φ_{lm}^{i} :¹

$$\hat{\mathbf{x}} = \sqrt{\frac{2\pi}{3}} \left(-(\mathbf{Y}_{11} + \Psi_{11}) + (\mathbf{Y}_{1,-1} + \Psi_{1,-1}) \right), \tag{4.1}$$

$$\hat{\mathbf{y}} = i\sqrt{\frac{2\pi}{3}} \left((\mathbf{Y}_{11} + \Psi_{11}) + (\mathbf{Y}_{1,-1} + \Psi_{1,-1}) \right), \tag{4.2}$$

$$\hat{\mathbf{z}} = \sqrt{\frac{4\pi}{3}} (\mathbf{Y}_{10} + \mathbf{\Psi}_{10}).$$
 (4.3)

Thus, a generic uniform vector field can be represented as

$$b^{i}(t) = \sum_{m=0,\pm 1} C_{m}(t) (\mathbf{Y}_{1m} + \boldsymbol{\Psi}_{1m}).$$
(4.4)

¹Let's make a brief detour and decompose a uniform (spatially constant) vector field b^i in terms of VSH. The Cartesian unit vectors can be easily constructed from the vector spherical harmonics Y_{1m}^i and Ψ_{1m}^i ,

$$\epsilon^{i}(t,\mathbf{x}) = \sum_{l,m} \left(E^{0}_{lm}(t,r) Y^{i}_{lm}(\theta,\varphi) + E^{1}_{lm}(t,r) \Psi^{i}_{lm}(\theta,\varphi) + E^{2}_{lm}(t,r) \Phi^{i}_{lm}(\theta,\varphi) \right) , \qquad (4.5)$$

$$\epsilon^{0}(t, \mathbf{x}) = \sum_{l,m} F_{lm}(t, r) Y_{lm}(\theta, \varphi) \,. \tag{4.6}$$

Now we have to impose the constraints satisfied by ϵ^{μ} . These come in three types. The first two are the constraints necessary to preserve the Newtonian gauge (4.7) and the adiabaticity conditions (ensuring that the large gauge transformation can be extended to a physical solution). If we require the solution to be a pure scalar, vector or tensor mode, we may also impose the condition that the remaining components must vanish. However, by necessity, most modes will be mixed, meaning that they contain e.g. both a scalar and a tensor part.

4.2 The constraints

4.2.1 Gauge-preserving constraints

I work in the Newtonian gauge. The metric is given by

$$ds^{2} = -(1+2\Phi)dt^{2} + a(t)G_{i}dtdx^{i} + a(t)^{2}(1-2\Psi)\delta_{ij}dx^{i}dx^{j} + a(t)^{2}\gamma_{ij}dx^{i}dx^{j}, \qquad (4.7)$$

where $\partial_i G_i = 0$ and γ_{ij} is traceless and transverse ($\gamma_{ii} = \partial_i \gamma_{ij} = 0$). If we are to preserve these conditions and remain in the Newtonian gauge after a gauge transformation, it needs to satisfy the *gauge-preserving constraints*

$$\nabla^2 \epsilon^i = -\frac{1}{3} \partial_i \partial_k \epsilon^k , \qquad (4.8)$$

$$\nabla^2 \epsilon_0 = -a^2 \partial_k \dot{\epsilon}^k \,. \tag{4.9}$$

4.2.2 Adiabaticity conditions

The conditions for the large diffeomorphism to be continuously connected to a physical solution are [96]

$$\Phi = \Psi, \qquad (4.10)$$

$$\dot{H}\delta u = H\Psi + \dot{\Psi}, \qquad (4.11)$$

$$\dot{G}^i = 0,$$
 (4.12)

$$-4\dot{H}a\delta u_j^V = \nabla^2 G_j \,. \tag{4.13}$$

In fact, the second equation does not impose any condition on ϵ^{μ} , since we are always free to set δu (the scalar component of the velocity perturbation) to $\frac{1}{\dot{H}}(H\Psi + \dot{\Psi})$. Moreover, the last equation is trivially satisfied at leading order (for perfect fluids) - see [96].

4.2.3 Perturbed quantities

Perturbed quantities can be expressed in terms of ϵ^{μ} as follows:

$$\Psi = -H\epsilon_0 + \frac{1}{3}\partial_k \epsilon^k , \qquad (4.14)$$

$$\Phi = \dot{\epsilon}_0, \qquad (4.15)$$

$$\delta u_i = \partial_i \epsilon_0 , \qquad (4.16)$$

$$\frac{\delta\rho}{\dot{\rho}} = \epsilon_0 , \qquad (4.17)$$

$$G^{i} = a \left(-\partial_{i} \epsilon_{0} - \dot{\epsilon}^{i} \right) .$$

$$(4.18)$$

4.2.4 Summary of the constraints

If we use the equations (4.14) - (4.18) to write the perturbed quantities in terms of the diffeomorphism ϵ^{μ} , the gauge constraints and the adiabaticity conditions become

$$\nabla^2 \epsilon^i = -\frac{1}{3} \partial_i \partial_k \epsilon^k , \qquad (4.19)$$

$$\nabla^2 \epsilon_0 = -a^2 \partial_i \dot{\epsilon}^i \,, \tag{4.20}$$

$$\dot{\epsilon}_0 + H\epsilon_0 = \frac{1}{3}\partial_k \epsilon^k , \qquad (4.21)$$

$$\left(\partial_t + H\right) \left(a^2 \dot{\epsilon}^i + \partial_i \epsilon_0\right) = 0.$$
(4.22)

These equations must be satisfied by **any** adiabatic mode. However, additional equations can be imposed if we demand from our mode that:

• its scalar part vanishes:

$$\epsilon_0 = 0, \tag{4.23}$$

• its vector part vanishes:

$$a^2\dot{\epsilon}^i + \partial_i\epsilon_0 = 0, \quad \nabla^2\dot{\epsilon}^i = 0, \tag{4.24}$$

• its tensor part vanishes:

$$\partial_i \epsilon^j + \partial_j \epsilon^i = \frac{2}{3} \delta_{ij} \partial_k \epsilon^k.$$
(4.25)

Thus, for example, for pure vector modes (4.23) and (4.25) must be satisfied. We will see that most of solutions are not pure scalar/vector/tensor modes, but are *mixed*, e.g. composed of a scalar and a tensor part.

4.3 Solutions

I will start by solving the equation (4.19). Because modes with distinct l, m decouple from each other, it will suffice to look for solutions of the form

$$\epsilon^{i}(t,\mathbf{x}) = E_{lm}^{0}(t,r)Y_{lm}^{i}(\theta,\varphi) + E_{lm}^{1}(t,r)\Psi_{lm}^{i}(\theta,\varphi) + E_{lm}^{2}(t,r)\Phi_{lm}^{i}(\theta,\varphi).$$
(4.26)

The formulae for gradient, divergence and Laplacian in the vector spherical formalism are [101]

$$\nabla(f(r)Y_{lm}) = \frac{df}{dr}\mathbf{Y}_{lm} + \frac{f}{r}\boldsymbol{\Psi}_{lm}, \qquad (4.27)$$

$$\nabla_{k}\epsilon^{k} = \left(\frac{dL_{lm}}{dr} + \frac{2}{r}E_{lm}^{0} - \frac{c(l+1)}{r}E_{lm}^{1}\right)Y_{lm}, \qquad (4.28)$$

$$\nabla^{2}\epsilon^{i} = \left(\frac{1}{r}\frac{d^{2}}{dr^{2}}(rE_{lm}^{0}) - \frac{2+l(l+1)}{r^{2}}E_{lm}^{0} + \frac{2l(l+1)}{r^{2}}E_{lm}^{1}\right)Y_{lm}^{i} + \left(\frac{1}{r}\frac{d^{2}}{dr^{2}}(rE_{lm}^{1}) + \frac{2}{r^{2}}E_{lm}^{0} - \frac{l(l+1)}{r^{2}}E_{l}^{1}m\right)\Psi_{lm}^{i} + \left(\frac{1}{r}\frac{d^{2}}{dr^{2}}(rE_{lm}^{2}) - \frac{l(l+1)}{r^{2}}E_{lm}^{2}\right)\Phi_{lm}^{i}. \qquad (4.29)$$

Using the first two formulae, we can compute $\partial_i \partial_k \epsilon_k$:

$$\partial_i \partial_k \epsilon_k = \frac{d}{dr} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) Y_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} E_{lm}^1 \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} E_{lm}^0 - \frac{l(l+1)}{r} \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac{dE_{lm}^0}{dr} + \frac{2}{r} \right) \Psi_{lm}^i + \frac{1}{r} \left(\frac$$

Equating each of the terms in $\nabla^2 \epsilon_i = -\frac{1}{3} \partial_i \partial_k \epsilon_k$, we get the following system of equations,

$$\frac{1}{r}\frac{d^2}{dr^2}\left(rE_{lm}^0\right) - \frac{2+l(l+1)}{r^2}E_{lm}^0 + \frac{2l(l+1)}{r^2}E_{lm}^1 = -\frac{1}{3}\frac{d^2}{dr^2}\left(E_{lm}^0\right) \\ -\frac{2}{3}\frac{d}{dr}\left(\frac{E_{lm}^0}{r}\right) + \frac{l(l+1)}{3}\frac{d}{dr}\left(\frac{E_{lm}^1}{r}\right),$$
(4.30)

$$\frac{1}{r}\frac{d^2}{dr^2}\left(rE_{lm}^1\right) + \frac{8}{3}\frac{1}{r^2}E_{lm}^0 - \frac{4}{3}\frac{l(l+1)}{r^2}E_{lm}^1 = -\frac{1}{3}\frac{1}{r}\frac{d}{dr}\left(E_{lm}^1\right), \qquad (4.31)$$

$$\frac{1}{r}\frac{d^2}{dr^2}\left(rE_{lm}^2\right) = \frac{l(l+1)}{r^2}E_{lm}^2.$$
(4.32)

The simplest family of solutions to the above equations is $E_{lm}^2 = Zr^l$, $E_{lm}^0 = E_{lm}^1 = 0$. I will call this the **Z branch**. Recall that the functions E_{lm}^k can be time dependent, so Z = Z(t).

The first two equations are slightly more complicated, since they both mix $E_{lm}^{0,1}$. Nevertheless, this is a system of equidimensional differential equations which can be solved by substituting monomials in r for $E_{lm}^{0,1}$. Subject to appropriate regularity conditions at r = 0, the general solution for fixed l, m is

$$E_{lm}^{0} = A_{1}(t)r^{l+1} + C_{1}(t)r^{l-1}, (4.33)$$

$$E_{lm}^{1} = A_{2}(t)r^{l+1} + C_{2}(t)r^{l-1}, (4.34)$$

$$E_{lm}^2 = Z(t)r^l. (4.35)$$

where $C_i \neq 0$ only if $l \ge 1$, and the ratios A_2/A_1 and C_2/C_1 are fixed and independent of time:

$$A_2/A_1 = \frac{l+9}{(l-6)(l+1)},$$
(4.36)

$$C_2/C_1 = \frac{1}{l}.$$
 (4.37)

Note that for l = 6 we must have $A_1 = 0$.

For l = 1 the solution might appear singular at r = 0, because the term $C_1 \mathbf{Y}_{1m}$ represents a radial vector with length independent of r, thus suggesting a discontinuity at the origin. However, for l = 1we have $C_1 = C_2$, which implies that \mathbf{Y}_{1m} and Ψ_{1m} appear only in the combination $\mathbf{Y}_{1m} + \Psi_{1m}$. As I explained in footnote 1 at the beginning of this chapter, such combinations constitute a basis for uniform vector fields which are regular at r = 0 and should be interpreted as translations in the position space.

I conclude that for l = 1, a term with a quadratic dependence on r must accompanied by a

translation. The explicit form of this translation (ie. the value of ϵ^i at r = 0) in terms of the coefficients $C_{1,m} = C_{2,m}$ is

$$b^x = \sqrt{\frac{2\pi}{3}}(-C_{1,1} + C_{1,-1}),$$
 (4.38)

$$b^y = \sqrt{\frac{2\pi}{3}}i(C_{1,1} + C_{1,-1}),$$
 (4.39)

$$b^z = \sqrt{\frac{4\pi}{3}}C_{1,0}.$$
(4.40)

I have thus found a general solution to the first constraint equation, (4.19). The remaining constraints are

$$\nabla^2 \epsilon_0 = -a^2 \partial_k \dot{\epsilon}^k \,, \tag{4.41}$$

$$\dot{\epsilon}_0 + H\epsilon_0 = \frac{1}{3}\partial_k \epsilon^k , \qquad (4.42)$$

$$(\partial_t + H) \left(a^2 \dot{\epsilon}^i + \partial_i \epsilon_0 \right) = 0.$$
(4.43)

It can be shown that the above system of three equations is equivalent to

$$0 = \ddot{\epsilon}^i + 3H\dot{\epsilon}^i - \frac{1}{a^2}\nabla^2\epsilon^i, \qquad (4.44)$$

$$\epsilon_0 = -a^2 \partial_t \nabla^{-2} \partial_k \epsilon^k + \frac{4}{3} (\partial_t + H)^{-1} \partial_k \epsilon^k + \alpha(t, \mathbf{x}), \qquad (4.45)$$

$$\nabla^2 \alpha = 0, \qquad (4.46)$$

$$\dot{\alpha} + H\alpha = 0. \tag{4.47}$$

Equation (4.45) can always be solved for ϵ_0 . We can treat the solution $\epsilon_0 = \alpha$, $\epsilon^i = 0$ as a new family of modes (the **time-shift branch**) that are independent from all the others. I will discuss such solutions later. We can substract a time-shift mode from any other solution, so that if $\epsilon^i \neq 0$, we will assume $\alpha = 0$ without loss of generality.

Thus, any mode given in (4.33)-(4.35) will be an adiabatic mode, provided it has the time dependence specified in (4.44). The associated time shift is given by (4.45) with $\alpha = 0$.

4.3.1 The AC branch

In this subsection, I consider the solutions

$$E_{lm}^{0} = A_{1}r^{l+1} + C_{1}r^{l-1},$$

$$E_{lm}^{1} = A_{2}r^{l+1} + C_{2}r^{l-1},$$

$$E_{lm}^{2} = 0.$$
(4.48)

with $A_2/A_1 = \frac{l+9}{(l-6)(l+1)}$, $C_2/C_1 = 1/l$. Note that different spherical harmonics decouple in (4.19) and (4.44), so we can assume that $E_{lm}^k \propto \delta_{ll'} \delta_{mm'}$. We have

$$\nabla^{2} \epsilon^{i} = A_{1} \left[(l+2)(l+1) - (2+l(l+1)) + 2l(l+1)A_{2}/A_{1} \right] r^{l-1} Y_{lm}^{i} + A_{2} \left[(l+2)(l+1) + 2A_{1}/A_{2} - l(l+1) \right] r^{l-1} \Psi_{lm}^{i} .$$
(4.49)

Hence, (4.44) yields

$$\left(\ddot{A}_{1}+3H\dot{A}_{1}\right)r^{l+1}+\left(\ddot{C}_{1}+3H\dot{C}_{1}-\left(2l+2l\frac{l+9}{l-6}\right)A_{1}a^{-2}\right)r^{l-1}=0$$
(4.50)

$$\frac{A_2}{A_1} \left(\ddot{A}_1 + 3H\dot{A}_1 \right) r^{l+1} + \left(\ddot{C}_2 + 3H\dot{C}_2 - \left(2\frac{l+9}{l-6} + 2 \right) A_1 a^{-2} \right) r^{l-1} = 0.$$
(4.51)

This is equivalent to

$$\ddot{A}_2 + 3H\dot{A}_2 = 0 \tag{4.52}$$

$$\ddot{C}_2 + 3H\dot{C}_2 - 2(l+1)\left(\frac{2l+3}{l+9}\right)A_2(t)a(t)^{-2} = 0.$$
(4.53)

The solutions for l > 0 are

$$A_2(t) = k_0 + \int \frac{k_1}{a(t')^3} dt', \qquad (4.54)$$

$$C_2(t) = k_2 + \int \frac{k_3}{a(t')^3} dt + 2(l+1) \left(\frac{2l+3}{l+9}\right) \int^t \frac{\int^{t'} A_2(t'')a(t'')dt''}{a(t')^3} dt'.$$
(4.55)

Note that the part $\bar{C}_2(t) = k_2 + \int \frac{k_3}{a(t')^3} dt$ is independent of A_2 , and therefore can be interpreted as an independent solution which I will discuss in subsection 4.3.2.

The AC branch, l = 0

Since Y_{00} is constant, the Ψ_{00}^i and Φ_{00}^i vector harmonics vanish identically. We are left with (up to a constant of proportionality):

$$\epsilon^i = A(t)x^i \,. \tag{4.56}$$

The Laplacian vanishes, so the time dependence is given by

$$A(t) = \lambda + \int^{t} \frac{C}{a(t')^{3}} dt'.$$
(4.57)

The corresponding time shift is

$$\epsilon_0 = \frac{\lambda}{a} \int^t a(t')dt' + \frac{C}{a} \left(\int^t dt' a(t') \int^{t'} \frac{dt''}{a(t'')^3} - \frac{1}{2}\mathbf{x}^2 \right) \,. \tag{4.58}$$

The first of these modes ($\lambda \neq 0, C = 0$) is exactly the Weinberg first scalar mode; the second ($\lambda = 0, C \neq 0$) is the time-dependent scalar mode from [96].

The AC branch, l = 1

Recall that for l = 1, an $O(r^2)$ term is accompanied by a translation.

$$A_1(t) = -k_0 - \int \frac{k_1}{a(t')^3} dt', \qquad (4.59)$$

$$A_2(t) = k_0 + \int \frac{k_1}{a(t')^3} dt', \qquad (4.60)$$

$$C_1(t) = C_2(t) = -2 \int^t \frac{\int^{t'} A_1(t'')a(t'')dt''}{a(t')^3} dt'.$$
(4.61)

Here I subtracted the part $\bar{C}_1(t) = k_2 + \int \frac{k_3}{a(t')^3} dt$, since it is an independent solution (a pure translation, discussed in the next subsection).

4.3.2 The C branch

In the solution (4.54) - (4.55), we can take $A_1 = A_2 = 0$ and

$$C_2(t) = k_2 + \int \frac{k_3}{a(t')^3} dt, \qquad (4.62)$$

with $C_1(t) = l C_2(t)$. I will call this the **C branch**. Let's discuss some special cases.

The C branch, l = 1 (the translations)

For l = 1, $A_1 = A_2 = 0$, we get uniform spatial shifts (translations):

$$\epsilon^{i}(t) = \mathcal{C}_{0}^{i} + \int \frac{\mathcal{C}^{i}}{a(t')^{3}} dt'.$$
(4.63)

The time-independent translation C_0^i has no physical relevance, so we can henceforth set it equal to zero. There are then infinitely many consistent choices of ϵ_0 . If we take $\epsilon_0 = 0$, we obtain a pure vector mode - it corresponds to the new vector mode of [96]. If we take $\epsilon_0 = -\frac{1}{a}C^ix^i$, we get a pure scalar mode - in [96], this is called the gradient of Weinberg second scalar mode.

The C branch, l = 2

I conjecture that this solution, when accompanied by an apprioprate dilation that cancels out the trace, corresponds to Weinberg tensor mode (for the time-independent part) or the first mixed mode of [96] (for the time-dependent part).

4.3.3 The Z branch

This branch exists only for l > 0 and is given by

$$E_{lm}^{0} = 0,$$

$$E_{lm}^{1} = 0,$$

$$E_{lm}^{2} = Z(t)r^{l}.$$
(4.64)

On the Z branch, we automatically have $\nabla^2 \epsilon^i = 0$. Thus

$$\ddot{Z}(t) + 3H\dot{Z}(t) = 0 \tag{4.65}$$

$$Z(t) = z_1 + \int^t \frac{z_2}{a(t')^3} dt'$$
(4.66)

For l = 1, these modes correspond to infinitesimal rotations (for example, l = 1, m = 0 corresponds to the generator of rotations around the z axis). Time-independent rotation is a trivial solution, while the time-dependent part corresponds to the gradient vector mode found in [96].

4.3.4 The time shift branch

The final case to consider is $\epsilon^i = 0$, $\epsilon_0 \neq 0$. Then there are infinitely many solutions, which must satisfy

$$\nabla^2 \epsilon_0 = 0, \quad \dot{\epsilon}_0 + H \epsilon_0 = 0. \tag{4.67}$$

A general solution is

$$\epsilon_0 = \frac{\beta(\mathbf{x})}{a(t)}, \quad \nabla^2 \beta = 0.$$
(4.68)

Taking $\epsilon_0 = \frac{c}{a(t)}$, we get a pure scalar mode (Weinberg second scalar mode). If ϵ_0 has spatial dependence, we get a mixed (scalar + vector) mode.

4.4 Comparison with (Pajer and Jazayeri, 2018)

Using the VSH decomposition, I explicitly reproduced the following modes found in [96]:

- Weinberg first scalar mode = the time-independent, l = 0 A mode (dilation).
- Weinberg second scalar mode = the time shift branch, $\epsilon_0 = \frac{c}{a(t)}$.
- Time-dependent scalar mode = the time-dependent, l = 0 A mode (time dependent dilation).
- Gradient of Weinberg second scalar mode = the translation branch (time dependent translation + temporal diffeomorphism).
- Vector mode = the translation branch (time dependent translation with $\epsilon_0 = 0$).
- Gradient vector mode = the Z branch, l = 1.

In addition, I shall make the following conjectures:

- The Weinberg tensor mode = the time-independent, l = 2 C mode, accompanied by l = 0 A mode.
- The first mixed mode = the time-dependent, l = 2 C mode, accompanied by l = 0 A mode.
- Gradient scalars and gradient tensors = the time-independent, l = 2 Z mode, accompanied by a spatial shift (translation).

All of the adiabatic modes derived in this chapter can be used to derive soft theorems following the background wave method presented in Section 3.6, or any other method mentioned therein. Since

our decomposition is related to (3.145) by a linear map, the associated soft theorems will be linear combinations of those derived using the polynomial basis.

As a lesson for the future, let me observe that the VSH decomposition significantly streamlines the calculations. This should be expected, since the system has an intrinsic SO(3) symmetry. Vector spherical harmonics are eigenfunctions of the symmetry generators and are thus more suitable than the polynomial basis (3.145). Yet, VSH seem to be underused in similar applications. I hope that the results of this chapter will convince the readers as to their usefulness.

Chapter 5

Spatial Curvature at Sound Horizon

Abstract

The effect of spatial curvature on primordial perturbations is controlled by $\Omega_{K,0}/c_s^2$, where $\Omega_{K,0}$ is today's fractional density of spatial curvature and c_s is the speed of sound during inflation. Here we study these effects in the limit $c_s \ll 1$. First, we show that the standard cosmological soft theorems in flat universes are violated in curved universes and the soft limits of correlators can have non-universal contributions even in single-clock inflation. This is a consequence of the fact that, in the presence of spatial curvature, there is a gap between the spectrum of residual diffeomorphisms and that of physical modes. Second, there are curvature corrections to primordial correlators, which are not scale invariant. We provide explicit formulae for these corrections to the power spectrum and the bispectrum to linear order in curvature in single-clock inflation. We show that the large-scale CMB anisotropies could provide interesting new constraints on these curvature effects, and therefore on $\Omega_{K,0}/c_s^2$, but it is necessary to go beyond our linear-order treatment.

5.1 Introduction and summary

The corroborated assumption that our universe is homogeneous and isotropic on large scales highly restricts the form of the spacetime metric. The only unknowns are the scale factor, whose evolution is dictated by Einstein equations, and the comoving spatial curvature K, which is fixed by the boundary conditions. General relativity gives us no guidance in choosing a specific value of K and so it is important to derive predictions for observables for generic values of K and confront them with

cosmological data. Current bounds constrain spatial curvature today to be at the per mille level or smaller [39]

$$\Omega_{K,0} \equiv \frac{K}{a_0^2 H_0^2} = 0.0007 \pm 0.0019 \quad (68\%, \text{Planck + BAO}).$$
(5.1)

But our cosmological model gives us also an estimate for a *lower bound* on the absolute value of Ω_K . This comes from the local effect of superHubble curvature perturbations. More specifically, it has been measured that subHubble curvature perturbations have an amplitude of about 2×10^{-9} and an approximate scale-invariant spectrum. It is natural to expect that these perturbations continue to exist on superHubble scales. To leading order, the effect of these superHubble perturbations on our Hubble patch is to induce spatial curvature and a tidal force on Hubble scales (this follows from using Fermi normal coordinates [102, 103] or more conveniently their cosmological generalisation known as conformal Fermi coordinates [104, 105]). An estimate for the local spatial curvature in our Hubble patch due to superHubble fluctuations then is

$$|\Omega_K| \sim \langle \Omega_K^2 \rangle^{1/2} = \frac{2}{3H_0^2} \left[\int_0^{H_0} \frac{d^3k}{(2\pi)^3} \frac{d^3k'}{(2\pi)^3} \langle k^2 \zeta(k) k'^2 \zeta(k') \rangle \right]^{1/2}$$
(5.2)

$$=\frac{2}{3H_0^2} \left[\int_0^{H_0} \frac{dk}{2\pi^2} k^6 P(k) \right]^{1/2}$$
(5.3)

$$= \frac{2}{3H_0^2} \left[\int_0^{H_0} dk k^3 \Delta_{\zeta}^2 \right]^{1/2} = \frac{\Delta_{\zeta}}{3} \simeq 1.5 \times 10^{-5} \,, \tag{5.4}$$

where we used the relation¹ $K = -(2/3)\nabla^2 \zeta$. An anisotropic tidal field is expected at a comparable level. Even though the above explicit estimate assumes a scale invariant power spectrum on all superHubble scales, the integral is dominated by scales that are around the current Hubble radius and so is quite insensitive to changes in the spectral index for ultra-long wavelengths.

The effect of curvature on cosmological observables such as the Cosmic Microwave Background (CMB) or Large Scale Structures (LSS) has been well studied in the literature and many existing numerical Boltzmann codes already allow one to include spatial curvature in solving cosmological perturbation theory. All effects of curvature during the hot big bang are controlled by the parameter $\Omega_K(t)$, evaluated at the relevant time for the given observable. As is well-known, $\Omega_K(t)$ is an

¹This relation is valid only at linear order. But in standard cosmological models $|\Omega_{K,0}| \ll 1$ is actually an upper bound on the value of Ω_K at any time during the hot big bang. We are therefore entitled to account for the effect of curvature during the hot big bang to linear order.



Figure 5.1: The figure shows why primordial perturbations are sensitive to $\Omega_{K,0}/c_s^2$. Since the fractional density of spatial curvature (green line) decreases during inflation, it is larger at sound-horizon crossing (first green dot) than at the crossing of the Hubble radius (second green dot).

increasing function of time in decelerated cosmologies. As a consequence, what controls the effect of curvature in the late universe is bounded by the value of curvature today $\Omega_K(t) \leq \Omega_{K,0}$.

What motivated this work is the observation that the effect of curvature on primordial perturbations from inflation is controlled instead by $\Omega_{K,0}/c_s^2$. Therefore, if primordial perturbations had a small speed of sound, $c_s \ll 1$, then they could provide a very sensitive probe of spatial curvature. This dependence on c_s is easy to understand (see Fig. 6.1). During inflation, perturbations freeze out at the so-called sound horizon, i.e. when the comoving wavenumber satisfies $c_s k = aH$. Consider now a perturbation of size the Hubble radius today $k_{H_0} = a_0H_0 = H_0$. The value of Ω_K at the time during inflation when k_{H_0} froze out is

$$\Omega_K \Big|_{\text{freeze}} = \frac{K}{a^2 H^2} \Big|_{\text{freeze}} = \frac{K}{c_s^2 k_{H_0}^2} = \frac{\Omega_{K,0}}{c_s^2} \,. \tag{5.5}$$

So, if $c_s \ll 1$ primordial perturbations felt a much larger value of Ω_K right before they stopped evolving than any cosmological observable in the late universe. The current bound² on c_s is

²A more detail discussion of the bound on c_s and the other EFT free parameter at the same order will be given in Sec. 5.5.1.

$$c_s \ge 0.021$$
 (95%, Planck T+E). (5.6)

Therefore we expect that the factor c_s^{-2} in (5.5) will give us a boost of up to a thousand in sensitivity to spatial curvature. The explicit calculations in this chapter confirm this rough estimate and furthermore show that the final observable effect also depends on the strength of interactions during inflation. For a fair comparison with other probes of curvature, it should also be mentioned that the effects of curvature on primordial correlators peak at the largest observable scales, where cosmic variance is largest. While this heuristic argument applies also to multifield inflation, in this work we focus exclusively on single-field inflation.

In this chapter, we take the observation that a small speed of sound enhances the sensitivity to spatial curvature in two distinct but related directions. First we study how spatial curvature can affect the soft theorems for cosmological correlators, which in a flat universe provide model-independent consistency relations to test the assumption of a single clock during inflation. While theoretical predictions for correlators are highly model dependent, in recent years it has become clear that symmetries, shared by large classes of models, lead to specific predictions known as soft theorems, which can be tested with current and upcoming data. Soft theorems constrain the squeezed limit of correlators, in which one of the momenta of the correlator is much smaller than any relevant scale in the problem. Cosmological soft theorems take the schematic form

$$\lim_{\mathbf{q}\to 0} \frac{\langle \mathcal{O}(\mathbf{q})\mathcal{O}(\mathbf{k}_1)\dots\mathcal{O}(\mathbf{k}_n)\rangle'}{\langle \mathcal{O}(\mathbf{q})\mathcal{O}(\mathbf{q})\rangle'} = \sum_{a=1}^n L_a \langle \mathcal{O}(\mathbf{k}_1)\dots\mathcal{O}(\mathbf{k}_n)\rangle',$$
(5.7)

where O are some operators, a prime denotes that we have dropped the momentum-conserving delta function and $L = L(k, \partial_k)$ is some linear operator consisting of powers of the momenta and derivatives. The most famous soft theorem has been derived by Maldacena in [106] and fixes the squeezed bispectrum in terms of the power spectrum and it applies to all attractor, single-field models of inflation [107]. This first result has been extended to higher n-point functions for primordial scalar, tensor and vector perturbations [93, 99, 108–117, 96, 95]. Soft theorems are conveniently interpreted as the consequence of residual, non-linearly realized symmetries associated with adiabatic modes [118, 93, 119, 96, 98], namely physical perturbations that are indistinguishable from a change of coordinates in the neighborhood of a point in spacetime. In the presence of additional symmetries beyond diffeomorphism invariance, new adiabatic modes and new soft theorems can be derived. One

example are non-attractor models of inflation such as Ultra-Slow-Roll inflation [120] that are also invariant under a shift symmetry. In this setup, new *generalized adiabatic modes* can be found, which are locally indistinguishable from a change of coordinates *and* a symmetry transformation [98, 121]. The violation of Maldacena's consistency relation in shift-symmetric Ultra-Slow-Roll inflation [122–125], can be attributed to the fact that cosmological perturbations asymptote generalized adiabatic modes, as opposed to the standard adiabatic modes [98, 126]. Adiabatic modes and their associated soft theorems can also be derived [127, 128] in the presence of alternative spacetime symmetry breaking patterns as in solid inflation [129, 130] and other generalizations [131–134]. All results so far have been obtained in spatially-flat FLRW spacetimes³. In this work, we study adiabatic modes and soft theorems in spatially-curved universes. A summary of our result can be found in the next subsection.

A second direction in which we push our investigation is the explicit calculation of corrections to the power spectrum and bispectrum of curvature perturbations that are induced by spatial curvature at linear order in Ω_K . To achieve this result we take advantage of the fact that, at linear order, the local effect of spatial curvature is the same as that of a suitable isotropic long wavelength perturbation on a spatially flat background. This equivalence allows us to use results in the literature for the bispectrum and trispectrum in flat FLRW spacetime to deduce the linear order effect of curvature on the power spectrum and bispectrum. The calculation simply amounts to extract a specific term from the squeezed limit of the higher-point correlator. Our main findings are summarized in the next subsection.

The rest of the chapter is organized as follows. In the next subsection we give a brief summary of our main results. In Sec. 5.2 we outline the procedure to derive adiabatic modes in curved FLRW universes and discuss the difficulty in continuing these modes to physical momentum. In Sec. 5.3, after reviewing the derivation of soft theorems in flat universes, we show that the standard "monochromatic" soft theorems do not exist in the presence of spatial curvature. This holds both for soft scalar and soft tensor modes. Here we also briefly mention some other non-standard and less phenomenologically relevant ways to constrain the soft behavior of correlators. In Sec. 5.4 we calculate the theoretical prediction for the effect of curvature on the power spectrum and bispectrum of curvature perturbations, both for canonical single-field inflation and for the decoupling limit of the Effective Field Theory of Inflation [136]. In Sec. 5.5, we show that, giving current constraints, the correction to the power

³The only exception known to us is [135], where the author studies soft theorems in a "toy" closed universe in 2+1 dimensions. The overall scaling of the violation of the consistency relation observed in that work in around Eq. (20) seems to match our results in 3+1 dimensions.

spectrum can be large enough to be measurable in the CMB, but this happens in a regime in which one needs to go beyond our linear treatment of curvature. We also make some estimates of the how large the non-scale invariant corrections to the bispectrum could be, given current constraints. Finally we conclude in Sec. 6.6 with a discussion and an outlook.

Notation and conventions: We use a mostly positive signature. Greek indices from the middle of the alphabet run over $\mu, \nu = 0, 1, 2, 3$ and latin indices from the middle of the alphabet over i, j = 1, 2, 3. We define symmetrization of indices by $a_{(ij)} = \frac{1}{2}(a_{ij} + a_{ji})$. Our Fourier conventions are

$$f(x) = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\mathbf{x}} f(\mathbf{k}), \qquad f(k) = \int d^3x \, e^{-i\mathbf{k}\mathbf{x}} f(\mathbf{x}). \tag{5.8}$$

Spatial 3-vectors are indicated in boldface, as for example in "x", and a hat denotes a unit norm vector $\hat{\mathbf{q}} \cdot \hat{\mathbf{q}} = 1$.

5.1.1 Summary of the results

In the following we give a short summary of our main results for soft theorems and for the curvature corrections to the power spectrum and bispectrum.

Soft theorems We find that the *standard soft theorems that relate n-point to* (n-1)-*point functions in the squeezed limit are generally violated in curved universes.* In particular the correlation in the squeezed limit is not simply a change of coordinates and therefore does not have the same universal character in a curved universe that it has in a flat universe. In attempting to reproduce the flat universe derivation of soft theorems, we derived all residual diffeomorphisms (diffs) in Newtonian gauge. Residual diffs, for both scalar and tensor modes do exist in curved universes and reduce to the respective flat-universe adiabatic modes in the $K \rightarrow 0$ limit. The main obstacle emerges when one tries to connect residual diffs to physical modes. For both scalars and tensors and in both open and closed universes the spectrum of physical modes (i.e. the eigenvalues of the Laplacian) is separated from the spectrum of residual diffs by a discrete gap. As a consequence of this, the time evolution of physical modes is different from that of residual diffs in some soft limit of a correlator. This introduces an error already at linear order in curvature and so we conclude that soft theorems are violated by curvature corrections. The violation is parameterized by $\Omega_{K,0}/c_s^2$ but depends also on

the strength of the interaction of perturbations, which are dictated by an explicit inflationary model or parameterized by the EFT of inflation. For the EFT of Inflation our main result for the squeezed bispectrum of curvature perturbations is the expression in (5.94), while for canonical single-field inflation we find (5.102).

Power spectrum and bispectrum To linear order, the local effect of spatial curvature can be traded for that of a long wavelength curvature perturbation (see e.g. [103, 112, 105]). This fact was used in [112] to show that the terms at order k_L^2 in the squeezed bispectrum (i.e. for $k_L \rightarrow 0$) are related to the corrections of spatial curvature to the power spectrum. These corrections are not scale invariant and peak on the largest scales, see (5.73) and (5.88)-(5.91). We confront the curvature-dependent power spectrum prediction with data on the CMB temperature anisotropies. The signal-to-noise is saturated by just the first few C_l^{TT} 's. The bounds that we derive (see Figure 5.6) are slightly weaker than the theoretical bounds we have from the validity of our linear treatment of curvature. This means that the CMB can potentially improve current bounds on the $\{\Omega_{K,0}, c_s\}$ plane in the direction $\Omega_{K,0}/c_s^2$, but this requires computing the correction to the power spectrum to all orders in K. We also compute curvature corrections to the bispectrum from the squeezed trispectrum in canonical single-field inflation (from [137, 138]), finding (5.102), and for the so-called P(X)-theories [139] (equivalent to the leading terms in the EFT of inflation), finding (5.88)-(5.90). We show that the signal-to-noise for this leading order curvature correction to the bispectrum is at most of order one for the allowed values of parameters and always smaller than one within the regime of validity of our analysis.

5.2 Residual diffeomorphism

In this section, we derive residual diffs in spatially-curved FLRW universes. Our main finding is that residual diffs do exist, but their momenta are always separated from the spectrum of physical modes by a discrete amount.

5.2.1 Gauge fixing

In this work, we consider curved FLRW universes with the following spacetime metric

$$ds^{2} = -dt^{2} + a^{2}(t)\tilde{g}_{ij}(\mathbf{x})dx^{i}dx^{j} \equiv \bar{g}_{\mu\nu}dx^{\mu}dx^{\nu}, \qquad (5.9)$$

where

$$\tilde{g}_{ij}(\mathbf{x}) = f^2(K\mathbf{x}^2)\delta_{ij} \quad \text{with} \quad f(K\mathbf{x}^2) \equiv \frac{1}{\left(1 + \frac{1}{4}K\mathbf{x}^2\right)}, \quad (5.10)$$

and we have defined $\mathbf{x}^2 \equiv \delta_{ij} x^i x^j$. The universe is spatially flat, open or closed if K = 0, K < 0and K > 0, respectively. In the open case, the radial coordinate satisfies $\mathbf{x}^2 < 4/|K|$, while in the closed case $0 \le x \le +\infty$. The volume contained in an open universe is infinite, while it is finite in a closed universe. Other useful properties of the FLRW metric can be found in Appendix 5.7.1.

For the perturbations around the FLRW background, we choose to use the Newtonian gauge, which is defined through

$$ds^{2} = -(1+2\Phi)dt^{2} + aG_{i}dtdx^{i} + a^{2}\left[(1-2\Psi)\tilde{g}_{ij} + \gamma_{ij}\right]dx^{i}dx^{j}, \qquad (5.11)$$

with

$$\nabla_i G^i = \nabla_i \gamma^{ij} = \gamma^i_i = 0.$$
(5.12)

Here, ∇_i is the covariant derivative with respect to the spatial metric \tilde{g}_{ij} . We raise and lower spatial indices with the \tilde{g}_{ij} metric. Metric perturbations are denoted by $h_{\mu\nu}$. We take the energy momentum tensor to be that of a single perfect fluid, so a universe with a single scalar field is also included in our study. To first order in perturbations, this implies

Generally, there might be residual diffeomorphisms that are compatible with the gauge choice.⁴ Under an infinitesimal change of coordinates $x^{\mu} \rightarrow x^{\mu} + \epsilon^{\mu}$, metric perturbations transform as

$$\Delta h_{00} = 2\dot{\epsilon}^0 \,, \tag{5.14}$$

⁴It should be possible to avoid such residual diffeomorphisms by an appropriate choice of coordinates. In flat space, even in such coordinates, cancelations among scalar, vector and tensor perturbations in the zero momentum limit lead to a set of adiabatic modes. See [128] for a related discussion in the context of Solid Cosmologies.

$$\Delta h_{0i} = \partial_i \epsilon^0 - a^2 f^2 \dot{\epsilon}^i, \tag{5.15}$$

$$\Delta h_{ij} = -2H\epsilon^0 \bar{g}_{ij} + Kf x^k \epsilon^k \bar{g}_{ij} - 2\bar{g}_{k(i}\partial_{j)}\epsilon^k , \qquad (5.16)$$

which in turn yield

$$\Phi = -\dot{\epsilon}^{0}, \qquad (5.17)$$

$$\Psi = H\epsilon^{0} - \frac{1}{2}Kf x^{k}\epsilon^{k} + \frac{1}{3}\partial_{k}\epsilon^{k}, \qquad G_{i} = \nabla_{i}\epsilon^{0} - a^{2}\dot{\epsilon}_{i}, \qquad \gamma_{ij} = 2\nabla_{(i}\epsilon_{j)} - \frac{1}{3}\nabla_{k}\epsilon^{k}\tilde{g}_{ij}.$$

The variables parameterizing the perfect fluid on the other hand change as

$$\Delta \delta \rho = -\dot{\rho} \epsilon^{0}, \qquad (5.18)$$
$$\Delta \delta p = -\dot{p} \epsilon^{0},$$
$$\Delta \partial_{i} \,\delta u + \Delta u_{i}^{V} = \partial_{i} \epsilon^{0}.$$

To maintain the Newtonian gauge, we must impose (5.12), giving

$$\nabla_i G^i = 0 \Rightarrow \quad \nabla^2 \epsilon^0 - a^2 \nabla_i \dot{\epsilon}^i = 0, \qquad (5.19)$$

$$\nabla^{i}\gamma_{ij} = 0 \Rightarrow \quad \nabla^{i}\left(2\nabla_{(i}\epsilon_{j)} - \frac{1}{2}\nabla_{k}\epsilon^{k}\,\tilde{g}_{ij}\right) = 0\,.$$
(5.20)

For diffs that respect (5.19) and (5.20), general covariance guarantees that the perturbations in (5.17) solve the linearized Einstein equations. However, just as for flat FLRW, for these diffs to have a chance to approximate physical perturbations, additional conditions must be satisfied. To see this, recall that physical perturbations in curved universes can be uniquely decomposed into scalars, vectors and tensors (with appropriate fall-off conditions in the open case, see Appendix 5.7.1). Then let us decompose the linearized Einstein equations, $\delta_{\mu\nu} = 0$, in the following way,

$$\delta E_{00} = S^{(1)}, \qquad (5.21)$$

$$\delta E_{0i} = \nabla_i S^{(2)} + V_i^{(1)}, \qquad \delta E_{ij} = S^{(3)} \bar{g}_{ij} + \nabla_i \nabla_j S^{(4)} + \nabla_{(i} V_{j)}^{(2)} + T_{ij},$$

where, $S^{(i)}$, $V^{(i)}$ and T_{ij} are the scalar, transverse vector and transverse traceless tensor components, respectively. For physical perturbations $S^{(1,2,3,4)} = V_i^{(1,2)} = T_{ij} = 0$. These equations are then *necessary* conditions (but as we will see not sufficient) for any residual diffs to be able to approximate physical perturbations. Since for residual diffs we already know that $\delta E_{\mu\nu} = 0$, we need only to further impose⁵

$$S^{(2)} = S^{(4)} = V^{(2)} = 0, (5.22)$$

where when we had the choice we set to zero those components with at most one time derivative. It is easy to verify that the rest of the components, namely $S^{(1)}$, $V^{(1)}$, $S^{(3)}$, and T, must also vanish as result of general covariance and rotational symmetry. In Newtonian gauge, (5.22) becomes

$$S^{(2)}: \dot{\Psi} + H\Phi = \left(\dot{H} - \frac{K}{a^2}\right)\delta u, \qquad S^{(4)}: \Phi = \Psi.$$
 (5.23)

We will refer to these equations as "adiabaticity conditions". Notice that there are no adiabaticity conditions for tensor residual diffs. Since vector modes decay in standard cosmologies, we set them to zero (so that $V^{(2)} = 0$) and ignore them in the rest of the chapter.

5.2.2 Scalar residual diffs

In this subsection, we investigate the existence of scalar residual diffs in an open or closed universe that satisfy the adiabaticity conditions (5.23) and (5.23). Demanding γ_{ij} to vanish restricts ϵ^i to solutions of the following equation

$$\partial_i \epsilon^j + \partial_j \epsilon^j = \frac{2}{3} \delta_{ij} \,\partial_k \epsilon^k \,. \tag{5.24}$$

This is nothing but the conformal Killing equation in Euclidean space. It has the following solutions

(Dilation)
$$\epsilon_{\rm d}^i = \lambda(t) x^i$$
, (5.25)

(Special conformal transformation) $\epsilon_{\text{SCT}}^{i} = \mathbf{b}(t) \cdot \mathbf{x} \, x^{i} - \frac{1}{2} \mathbf{x}^{2} \, b^{i}(t) \,,$ (5.26)

(Translation)
$$\epsilon_{\rm T}^i = c^i(t)$$
, (5.27)

(Rotation) $\epsilon_{\mathbf{R}}^{i} = \omega_{ij}(t) x^{j}$. (5.28)

⁵Notice that after setting $S^{(4)} = 0$, Einstein's equations automatically imply $S^{(3)} = S^{(1)} = 0$.

This is a good point to pause and discuss the spatial profiles of these residual diffs. Translations and rotations are isometries of the background metric and so do not affect perturbations. Dilations and special conformal transformations on the other hand do change perturbations. A common feature that is true regardless of spatial curvature is that the residual diffs in (5.25) and (5.26) cannot be fixed by imposing local conditions on the matter fields and metric components. Let discuss some additional curvature-dependent properties:

- In flat space, these residual diffs do not fall off at spatial infinity (|x| → +∞) and are therefore known as "large" diffs. This behavior should be contrasted with that of physical perturbations that are required to vanish at spatial infinity (so as to justify neglecting total spatial derivatives).
- In open universes, spatial infinity coincides with |x| → 2/√|K|, and residual diffs do not vanish there. In this sense these diffs could also be called "large" diffs. Again this should be contrasted with physical perturbations that should vanish as |x| → 2/√|K|.
- In closed universes, the spatial maximally-symmetric manifold is compact and there is no spatial infinity. In this case the above residual diffs are regular everywhere and they are square integrable. In this sense they are not "large" gauge transformations. They are only residual in the sense that they cannot be fixed by local gauge conditions.

In this work we focus on finding a counterpart to Weinberg's first adiabatic mode [118], which in flat space can be generated by a time-independent dilation, $\epsilon_d^i = \lambda x^i$. Unlike for the flat case, in curved universes the coefficient λ might have nontrivial time dependence, which is fixed by imposing the adiabaticity conditions in (5.23) and (5.23).

The absence of vector modes means that G_i in (5.17) must vanish. This implies

$$\epsilon^0 = -\frac{2a^2}{K}f\dot{\lambda} + \mathcal{D}(t), \qquad (5.29)$$

where \mathcal{D} is an integration "constant". Inserting ϵ^0 and ϵ^i into (5.18) and (5.17), and assuming no shift symmetry on δu^6 , we find the following solution

⁶For a perfect fluid, δu always has a shift symmetry. This simply reflects the equivalence between perfect fluids and superfluids (P(X) theories) in the limit of no-vorticity. We are however interested in a generic single scalar field cosmology, where the relation between δu and the scalar perturbation $\delta \phi$ breaks the shift symmetry [96].

$$\Phi = \frac{1}{K} \frac{2\partial_t (a^2 \dot{\lambda})}{1 + \frac{1}{4} K \mathbf{x}^2} - \dot{\mathcal{D}}(t), \qquad (5.30)$$

$$\Psi = H\mathcal{D} - \lambda + \frac{2(\lambda - \frac{H}{K} a^2 \dot{\lambda})}{1 + \frac{1}{4} K \mathbf{x}^2},$$

$$\delta u = -\frac{2a^2}{K} f \dot{\lambda} + \mathcal{D}(t).$$

So far, λ and \mathcal{D} have been arbitrary time-dependent functions. They are fixed however by imposing the adiabaticity conditions (5.23) and (5.23):

$$\mathcal{D} = \frac{a^2 \lambda}{K} \,, \tag{5.31}$$

$$\ddot{\lambda} + 3H\dot{\lambda} - \frac{K}{a^2}\lambda = 0.$$
(5.32)

In summary, we have found the following seemingly adiabatic solution

$$\Phi = \Psi = \frac{1}{K} \partial_t (a^2 \dot{\lambda}) \left(\frac{2}{1 + K \mathbf{x}^2/4} - 1 \right), \qquad (5.33)$$
$$\frac{\delta \rho}{\dot{\rho}} = \frac{\delta p}{\dot{p}} = -\delta u = \frac{a^2 \dot{\lambda}}{K} \left(\frac{2}{1 + K \mathbf{x}^2/4} - 1 \right), \qquad (5.33)$$
$$\mathcal{R} \equiv -\Psi + H \delta u = -\lambda \left(\frac{2}{1 + K \mathbf{x}^2/4} - 1 \right) = -\Psi - H \frac{\delta \rho}{\dot{\rho}} \equiv \zeta.$$

In the last line, \mathcal{R} and ζ are the curvature perturbations on comoving and constant-density slices, respectively. In the flat-space limit, $K \to 0$, both Weinberg's first and second adiabatic modes are obtained from the two independent $\lambda(t)$ solutions to (5.32).

For future reference, notice that

$$\nabla^2 \zeta = f^{-2} \left(\partial_i \partial_i - \frac{1}{2} K f x^k \partial_k \right) \zeta = -3K\zeta \,. \tag{5.34}$$

That is, our scalar residual diff has $\nabla^2 = -3K$ (see Table 5.1).

As a consistency check, we note that the evolution equation for scalar perturbations \mathcal{R} in curved universe with a scalar inflaton field, as derived for example in [140], takes the form

$$(\mathcal{D}^2 - K\mathcal{E})\ddot{\mathcal{R}} + \left[\left(H + 2\frac{\dot{z}}{z}\right)\mathcal{D}^2 - 3KH\mathcal{E}\right]\dot{\mathcal{R}} + \frac{1}{a^2}\left[K\left(1 + \mathcal{E} - \frac{2}{H}\frac{\dot{z}}{z}\right)\mathcal{D}^2 - \mathcal{D}^4 + K^2\mathcal{E}\right]\mathcal{R} = 0, \qquad (5.35)$$

where

$$\mathcal{D}^2 \equiv \nabla^2 + 3K, \quad z = \frac{a\dot{\phi}}{H}, \quad \mathcal{E} = \frac{\dot{\phi}^2}{2H^2}.$$
(5.36)

The limit $\mathcal{D}^2 \to 0$, which corresponds to the residual diffs, gives

$$\ddot{\mathcal{R}} + 3H\dot{\mathcal{R}} - \frac{K}{a^2}\mathcal{R} = 0, \qquad (5.37)$$

confirming (5.32).

5.2.3 Tensor residual diffs

We switch now to tensor perturbations for which there is no adiabaticity condition. To remove scalar modes we set $\delta u = \Phi = \Psi = 0$. This results in $\epsilon^0 = 0$ and

$$\nabla_i \epsilon^i = 0. \tag{5.38}$$

In addition, setting $G_i = 0$ enforces ϵ^i and, subsequently, γ_{ij} to be time-independent. The spatial diffs must further satisfy the gauge condition (5.20),

$$\nabla^{i} \left(\nabla_{i} \epsilon_{j} + \nabla_{j} \epsilon_{i} \right) = 0.$$
(5.39)

It is straightforward to see that

$$\nabla^2 \epsilon^i = -2K \epsilon^i \quad \Rightarrow \quad \nabla^2 \gamma_{ij} = +2K \gamma_{ij} \quad \text{(tensor residual diffs)}. \tag{5.40}$$

Notice that (5.40) is compatible with Einstein's equations for physical modes, which for tensors lead to

$$\ddot{\gamma}_{ij} + 3H\dot{\gamma}_{ij} - \frac{1}{a^2}(\nabla^2 - 2K)\gamma_{ij} = 0 \quad \text{(physical modes)}.$$
(5.41)

From this we see that a tensor residual diff, which must be constant in time, indeed satisfies

$$\ddot{\gamma}_{ij} = \dot{\gamma}_{ij} = 0 \quad \Rightarrow \quad \nabla^2 \gamma_{ij} = +2K\gamma_{ij} \,.$$
(5.42)

Just like their analogues in flat space, equations (5.40) and (5.38) admit infinitely many solutions [141]. These are easier to be written in spherical coordinates, where the solutions to (5.40) are either of even (+) or odd (-) parity. They are found to be (see [141])

$$\bar{\epsilon}_{r}^{(+)lm} = V_{1}^{l}(r) Y_{lm}(\theta, \phi), \quad \bar{\epsilon}_{\theta}^{(+)lm} = V_{2}^{l}(r) \partial_{\theta} Y_{lm}(\theta, \phi), \quad \bar{\epsilon}_{\phi}^{(+)lm} = V_{2}^{l}(r) \partial_{\phi} Y_{lm}(\theta, \phi), \quad (5.43)$$
$$\bar{\epsilon}_{r}^{(-)lm} = 0, \qquad \qquad \bar{\epsilon}_{\theta}^{(-)lm} = V_{3}^{l}(r) \partial_{\theta} Y_{lm}(\theta, \phi), \quad \bar{\epsilon}_{\phi}^{(-)lm} = V_{3}^{l}(r) \partial_{\phi} Y_{lm}(\theta, \phi).$$

Above, we have pedantically denoted the spatial diffs in the spherical coordinates with $\bar{\epsilon}_a$ $(a = r, \theta, \phi)$ to distinguish them from their Cartesian counterparts, and we have defined

$$V_{1}^{l}(r) = \frac{1}{\sqrt{r^{3} f(Kr^{2})}} P_{-5/2}^{-l-1/2} \left(\frac{4-Kr^{2}}{4+Kr^{2}}\right), \qquad (5.44)$$

$$V_{2}^{l}(r) = \frac{1}{l(l+1)} \frac{1}{f(Kr^{2})} \partial_{r} \left[\sqrt{r f(Kr^{2})} P_{-5/2}^{-l-1/2} \left(\frac{4-Kr^{2}}{4+Kr^{2}}\right)\right], \qquad (5.44)$$

$$V_{3}^{l}(r) = \sqrt{r f(Kr^{2})} P_{-5/2}^{-l-1/2} \left(\frac{4-Kr^{2}}{4+Kr^{2}}\right), \qquad (5.44)$$

where P^{μ}_{λ} are associated Legendre functions. The resulting tensor modes can be computed by inserting any of the above $\bar{\epsilon}_a$'s in $\bar{\gamma}_{ab} = 2\nabla_{(a}\bar{\epsilon}_{b)}$.

5.2.4 From large diffs to physical modes

Now we would like to see if the residual diffs satisfying the adiabaticity conditions that we have found in the previous sections can be smoothly connected to physical modes. In a flat universe this is the case, as in the $k \rightarrow 0$ limit physical perturbations become indistinguishable from (large) residual diffs, for which k = 0. As we will see below, in spatially curved universes we find an obstruction in the form of a discrete gap between the wavenumber of residual diffs and the spectrum of physical modes. This is summarised in Table 5.1.

In both open and closed universes, the scalar and tensor residual diffs discussed in Sec. 5.2.2 and Sec. 5.2.3 are eigenfunctions of the ∇^2 operator, with eigenvalues

$$\nabla^2 S(x) = -3KS(x)$$
 (Scalar residual diffs), (5.45)

$$\nabla^2 T(x) = +2KT(x)$$
 (Tensor residual diffs). (5.46)

Let us compare this with the physical spectrum in open and closed universes. In an open universe

| | | Open ($K < 0$) | Closed ($K > 0$) |
|---------|------------|-------------------------------|--------------------------------------|
| Scalars | Res. diffs | 3 K > 0 | -3K < 0 |
| | Physical | $-(1+p^2) K \leqslant - K $ | $-p(p+2)K < 0, p = 2, 3, \dots$ |
| Tensors | Res. diffs | -2 K < 0 | 2K > 0 |
| | Physical | $-(3+p^2) K \leqslant -3 K $ | $-(p(p+2)-2)K < 0, p = 2, 3, \dots$ |

Table 5.1: This table summarises the eigenvalues of the 3D spatial Laplacian ∇^2 for residual diffs ("res. diffs") and physical modes ("physical") in open and closed universes.

(K < 0), monochromatic⁷ perturbations that provide a complete basis of square integrable functions consist of the so-called *subcurvature* modes, all of which have negative eigenvalues [142, 143]:

$$\nabla^2 S_{plm}(x) = -(1+p^2)|K|S_{plm}(x), \qquad \text{(physical Scalars for } K < 0), \qquad (5.47)$$

$$\nabla^2 T_{plm,ij}(x) = -(3+p^2)|K|T_{plm,ij}(x), \qquad \text{(physical Tensors for } K < 0), \qquad (5.48)$$

where p > 0. Due to the existing gap between the momenta of the physical perturbations and the residual diffs, namely

$$-(1+p^2)|K| < -|K| \text{ vs } + 3|K| \qquad (\text{scalar gap}), \qquad (5.49)$$

$$-(3+p^2)|K| < -3|K| \text{ vs } -2|K| \qquad (\text{tensor gap}), \qquad (5.50)$$

monochromatic physical modes cannot capture the time dependence of the gauge modes in any continuous limit - neither for the scalar nor for the tensor. (This is in contrast with flat space, where eigenfunctions of ∇^2 can have the eigenvalue $-k^2$ arbitrarily close to zero and asymptote to the behavior of the (large) residual diff in the long wavelength limit.)

In closed universes, K > 0, residual diffs have Laplacian eigenvalues again given by (5.45) and (5.46). Normalizable modes on the other hand obey

$$\nabla^2 S_{plm}(x) = -p(p+2)KS_{plm}(x)$$
 with $p = 0, 1, \dots,$ (5.51)

$$\nabla^2 T_{plm,ij}(x) = -(p(p+2)-2)KT_{plm,ij}(x) \qquad \text{with } p = 0, 1, \dots$$
(5.52)

However, all modes with p = 0, 1 are equivalent to diffs and physical modes only start at p = 2, so

⁷By a monochromatic mode, we mean any mode which is an eigenfunction of ∇^2 . Gradients ∇_i do not commute with each other and therefore cannot be simultaneously diagonalized.

the modes we found still cannot be approached in any continuous way by monochromatic modes (see Sec. 5.3 for more discussion).

The discussion so far has focused on monochromatic modes and as such we cannot preclude the possibility of reaching an adiabatic mode as a limit of some other non-monochromatic physical perturbation. We would like to find those perturbations that have the following property:

A family of physical perturbations parameterized by α is *adiabatic* in the limit $\alpha \to 0$ if and only if they converge pointwise in spacetime to a residual diff as $\alpha \to 0$.

How can we take each element of the family to be normalizable yet have them converge to the (non-normalizable) residual diff? The idea is to consider a class of perturbations that give a good approximation of the residual diffeomorphism within some finite region of spacetime and send the size of that region to infinity (in the open case) or to the size of the universe itself (in the closed case). For concreteness, let us concentrate on the dilation residual diff ("Res. Diff") in an open universe. We can take

$$\zeta_{(\alpha)}(t=0,\mathbf{x}) = \zeta_{\text{Res. Diff}}(t=0,\mathbf{x})\exp\left(-\alpha F(x)\right), \qquad (5.53)$$

$$\dot{\zeta}_{(\alpha)}(t=0,\mathbf{x}) = \dot{\zeta}_{\text{Res. Diff}}(t=0,\mathbf{x})\exp\left(-\alpha F(x)\right),$$
(5.54)

where F(x) is a function that increases with x sufficiently fast for each perturbation to be normalizable. Note that we need to specify the field as well as its time derivative, since the evolution equation for scalars, (5.35), is second order in time. It is likely that such a non-monochromatic adiabatic mode would not have the same relation between $\zeta(t = 0, \mathbf{x})$ and $\dot{\zeta}(t = 0, \mathbf{x})$ as the physical pertubation coming from a Bunch-Davies initial state, making these non-monochromatic modes less useful in practice.

We make a technical assumption that the Cauchy problem for linear perturbations is well-posed so that the solution must depend continuously on the initial conditions (although this assumption is not strictly necessary if we take different physical field profiles). The physical modes converge pointwise to the residual diff on the initial time slice, so by continuity of solutions they must also converge pointwise to the gauge mode in the entire spacetime region for which solutions exist (a natural assumption is that solutions do exist in the entire spacetime, ie. linear evolution of smooth initial conditions doesn't lead to singularities). This completes the construction of a family of scalar physical modes that become adiabatic in the limit $\alpha \rightarrow 0$. The argument is similar in the case of tensor modes and is valid in both open and closed universes. We can also conclude that on any finite (but otherwise arbitrary) patch of spacetime, convergence must be uniform.

5.3 Soft theorems

In this section, we start with a brief review of soft theorems in a flat universe to highlight the difficulties in generalizing the usual derivation to a curved universe. We then point out that, due to a gap between the Laplacian eigenvalues of residual diffeomorphisms and physical modes, the soft limit cannot be constrained in the usual way at order $\mathcal{O}(K)$. We also briefly discuss some new non-standard soft theorems of a more formal nature. Later on, in Sec. 5.4 and Sec. 5.5, we will directly compute curvature corrections to correlators, which will confirm the findings of this section.

5.3.1 Flat universe

Residual diffs reflect the underlying symmetries of the gravitational theory and therefore can be used to derive soft theorems for primordial correlators. A generic argument can be constructed as follows. Consider an n + 1-point function where one of the modes is close to a residual diff, such that its dominant time dependence matches that of the residual diff. We can write

$$\langle \zeta_{\mathbf{q}} \zeta_{\mathbf{k}_{1}} \dots \zeta_{\mathbf{k}_{n}} \rangle \sim \int \frac{d^{3}q'}{(2\pi)^{3}} \langle \zeta_{\mathbf{q}} \zeta_{\mathbf{q}'} \rangle \frac{\delta}{\delta \zeta_{\mathbf{q}'}} \langle \zeta_{\mathbf{k}_{1}} \dots \zeta_{\mathbf{k}_{n}} \rangle_{\zeta_{\mathbf{q}'}} + \mathcal{O}(\zeta_{\mathbf{q}}^{3})$$

$$= P_{\zeta}(q) \frac{\delta}{\delta \zeta_{\mathrm{GM}}} \langle \zeta_{\mathbf{k}_{1}} \dots \zeta_{\mathbf{k}_{n}} \rangle_{\zeta_{\mathrm{GM}}} + \mathcal{O}(\zeta_{\mathbf{q}}^{3}, q^{2}/k^{2}, q^{2}/(aH)^{2}).$$
(5.55)

In the final step we used the conservation of momentum and the fact that the soft mode resembles the residual diff (up to corrections of order q^2/k^2 and $q^2/(aH)^2$). The effect of a residual diff on the short modes is precisely a change of coordinates $x^{\mu} \to x^{\mu} + \epsilon^{\mu}$:

$$\delta_{\zeta_{\rm GM}} \langle \zeta_{\mathbf{k}_1} \dots \zeta_{\mathbf{k}_n} \rangle_{\zeta_{GM}} = \sum_{i=1}^n \langle \zeta_{\mathbf{k}_1} \dots (\delta_\epsilon \zeta_{\mathbf{k}_i}) \dots \zeta_{\mathbf{k}_n} \rangle.$$
(5.56)

Then the soft theorem takes the form

$$\lim_{q \to 0} \langle \zeta_{\mathbf{q}} \zeta_{\mathbf{k}_1} \dots \zeta_{\mathbf{k}_n} \rangle \sim P_{\zeta}(q) \sum_{i=1}^n \langle \zeta_{\mathbf{k}_1} \dots (\delta_{\epsilon} \zeta_{\mathbf{k}_i}) \dots \zeta_{\mathbf{k}_n} \rangle + \mathcal{O}(q^2/k_n^2, q^2/(aH)^2) \,. \tag{5.57}$$

5.3.2 Absence of monochromatic soft theorems

In Section 5.2, we have established that in curved universes monochromatic physical modes are always separated from residual diffs by some finite gap that is proportional to K. Thus, a consistency relation in which the soft mode is monochromatic fails to capture the effect of a diff by a discrete amount; a difference of order $O(K/k_s^2)$ is always present between physical modes and residual diffs. Of course this difference can be very small if curvature is very small, and that limit indeed reproduces the flat space results. But already at linear order in K one finds violations of the flat-universe soft theorems. We conclude that soft theorems of the usual form do not exist in curved universes. This conclusion applies to both scalar and tensor soft theorems in both open and closed universes.

We can explicitly show where the $O(K/k_s^2)$ errors originate in the derivation. Focusing (for concreteness) on the open universe, scalar case, we obtain - in terms of open harmonics:

$$\langle \zeta_{q00}\zeta_{k_slm}\zeta_{k'_sl'm'}\rangle = \int dq' \langle \zeta_{q00}\zeta_{q'00}\frac{\delta}{\delta\zeta_{q'00}} \langle \zeta_{k_slm}\zeta_{k'_sl'm'}\rangle_{\zeta} \rangle$$

$$\approx \int dq' \langle \zeta_{q00}\zeta_{q'00}\rangle \frac{\delta}{\delta\epsilon} \langle \zeta_{k_slm}\zeta_{k'_sl'm'}\rangle + \mathcal{O}\left(\frac{q^2 - 3K}{k_s^2}\right) \langle \zeta_{k_slm}\zeta_{k'_sl'm'}\rangle.$$
(5.58)

The error in approximating a physical mode $\zeta_{q'00}$ with a monochromatic residual diff as done in the second line comes from two effects. First, even at some constant time we know that the eigenvalues of the Laplacian must differ at order $\mathcal{O}(q^2 - 3K)$. Second, by the differential equation that each perturbation satisfies, see for example (5.35), this difference in ∇^2 leads to a time dependence that is also different at order $\mathcal{O}(q^2 - 3K)$, since this quantity vanishes for the residual scalar diff but it does not for the physical mode. A short mode with momentum k_s and associated length scale $x_s = k_s^{-1}$ feels this difference as its evolves before freezing out. The relevant dimensionless quantities that estimate this error are then

$$\mathcal{O}\left((q^2 - 3K)x_s^2\right) \simeq \mathcal{O}\left(\frac{q^2 - 3K}{k_s^2}\right) \ge \mathcal{O}\left(\frac{|K|}{k_s^2}\right) \,. \tag{5.59}$$

Note that the error does not contain an $\mathcal{O}(|K|/q_l^2)$ term. Thus the leading order behaviour of a model dependent effect in the squeezed bispectrum will be $\mathcal{O}(|K|/k_s^2)$, as we will show explicitly in Sec. 5.4.

5.3.3 Formal soft theorems and adiabatic modes

We know from the discussion around (5.53) that in curved universes physical perturbations can be approximated arbitrarily well by residual diffs, even though such perturbations cannot be monochromatic and do not come from a Bunch-Davies initial condition. Consider then any family of physical scalar perturbations $\zeta_{(\alpha)}$ that are adiabatic in the limit $\alpha \to 0$. The strategy is to put the $\zeta_{(\alpha)}$ modes on the left hand side of a consistency relation in place of the usual soft mode.

Let us assume that for each α , there exists a basis $\{\zeta_n\}$ (its elements labelled by n) of *pairwise* uncorrelated modes containing $\zeta_{(\alpha)}$ itself. Then we have

$$\langle \zeta_{(\alpha)} \zeta_{klm} \zeta_{k'l'm'} \rangle \sim \int d\mathbf{n} \langle \zeta_{(\alpha)} \zeta_{\mathbf{n}} \rangle \frac{\delta}{\delta \zeta_{\mathbf{n}}} \langle \zeta_{klm} \zeta_{k'l'm'} \rangle_{\zeta_{\mathbf{n}}} \sim \langle \zeta_{(\alpha)}^2 \rangle' \frac{\delta}{\delta \zeta_{\text{Res. diff}}} \langle \zeta_{klm} \zeta_{k'l'm'} \rangle_{\zeta_{\text{Res. diff}}},$$
(5.60)

up to corrections that vanish in the limit $\alpha \to 0$, in which the soft perturbation approaches the residual diff.

A few remarks are in place. The primed 2-pt function $\langle \zeta_{(\alpha)}^2 \rangle'$ depends on the chosen basis $\{\zeta_n\}$ and it will be generally difficult to compute. Similarly, we do not have a closed expression for the action of a diff on the short modes. Finally, O(K) effects in the consistency relation are correctly captured only when the soft mode $\zeta_{(\alpha)}$ resembles the adiabatic mode on scales comparable to the curvature scale, in which case $\langle \zeta_{(\alpha)}^2 \rangle'$ is a superhorizon quantity inaccessible to local observers. Despite all these limitations, the above formulation of a soft theorem is formally valid to an arbitrary accuracy in the limit $\alpha \to 0$, despite the existence of a gap in the momenta of residual diffs and physical modes. We thus demonstrated the possibility of extending soft theorems to universes with nonvanishing spatial curvature, albeit in a formal sense.

5.4 Curvature corrections to the power spectrum and bispectrum

The primary focus of this section is to investigate the dominant effect of curvature on the power spectrum and the bispectrum

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \rangle = (2\pi)^3 P(k_1) \delta^{(3)} \left(\mathbf{k}_1 + \mathbf{k}_2 \right) ,$$
 (5.61)

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle = (2\pi)^3 B(k_1, k_2, k_3) \delta^{(3)} \left(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 \right) \,. \tag{5.62}$$

Whenever the scale associated with curvature is much larger than length scales relevant to observations, the physical effects of the former can be captured by an isotropic perturbation in a flat universe [103]. To see this, let expand the metric around the origin of coordinates, or equivalently to linear order in K:

$$g_{ij} = \frac{a_K(t)^2 \delta_{ij}}{\left(1 + \frac{1}{4} K \mathbf{x}^2\right)^2} \simeq a_K^2 \delta_{ij} \left(1 - \frac{1}{2} K \mathbf{x}^2\right) + \mathcal{O}((K \mathbf{x}^2)^2)$$
(5.63)

$$=a^{2}\delta_{ij}\left[1+2\frac{K}{a}\left(\frac{\partial a_{K}}{\partial K}\right)_{K=0}-\frac{1}{2}K\mathbf{x}^{2}\right]+\mathcal{O}((K\mathbf{x}^{2})^{2},$$
(5.64)

where a_K is the solution of the Friedmann equation in presence of spatial curvature while a is the solution for a flat FLRW metric. We can re-interpret this metric in terms of the flat space metric with a curvature perturbation

$$g_{ij} = a_{flat}^2 \delta_{ij} (1 + 2\tilde{\zeta}_B) \,. \tag{5.65}$$

There are two possible ways to do this⁸:

Option 1:
$$a_{\text{flat}} = a_K$$
 $\tilde{\zeta}_B = -\frac{1}{4}K\mathbf{x}^2$, (5.66)

Option 2:
$$a_{\text{flat}} = a$$
 $\tilde{\zeta}_B = \zeta_K(t) - \frac{1}{4}K\mathbf{x}^2$, (5.67)

where we defined

$$\zeta_K(t) \equiv \frac{K}{a(t)} \left[\frac{\partial a_K(t)}{\partial K} \right]_{K=0} .$$
(5.68)

Option 1 has the advantage that $\tilde{\zeta}_B$ is simpler, but it has the awkward feature that the curvature effects are partly in perturbations and partly in the background quantities, which mixes the perturbative expansion with the K expansion. Option 2 instead has the advantage that $\tilde{\zeta}_B$ is precisely a solution of the Einstein equations in a flat universe, but the drawback is that it contains one more term. In the following we will find it more convenient to work with Option 1. In words, it says that the spherical perturbation in (5.66) is locally indistinguishable from a mean curvature K superimposed with a change in the scale factor or equivalently in the Hubble parameter,⁹

⁸To avoid confusion, *specific field configurations* will be marked by a tilde, as in $\tilde{\zeta}_B$. ⁹*H* is the Hubble parameter in the absence of curvature, while H_K is the Hubble parameter when curvature is present.
$$H_K \to H \left[1 + \frac{K}{2a^2(t)H^2} \right] \,. \tag{5.69}$$

(This is valid for attractor scenarios, where superhorizon modes are approximately frozen.) Thus, if we knew the (n + 1)-point function $\langle \zeta_{\mathbf{q}_{u}} \zeta_{\mathbf{k}_{1}} \dots \zeta_{\mathbf{k}_{n}} \rangle$ in the regime $q_{u} \ll k_{i}$, we could recover the leading order effect of curvature on the *n*-point function. We expect that order *K* corrections to the power spectrum and to the bispectrum obtained through this method are in no way constrained by soft theorems. This is because $\mathcal{O}(K)$ in the *n*-point function corresponds to $\mathcal{O}(q_{u}^{2})$ in the (n + 1)-pt function, which is already a model-dependent effect. Hence, $\mathcal{O}(K)$ corrections will depend on the details of inflationary theory, and might be enhanced even in single field scenarios.

In this section, we quote the result of [112] for the curvature correction to the *power spectrum* in the framework of the EFT of inflation with small speed of sound. Then we study $\mathcal{O}(K)$ corrections to the *bispectrum* in two cases: canonical, single field, slow roll inflation as well as the EFT of inflation. Our argument follows closely the method of [112], and we compare our results to the $\mathcal{O}(K)$ correction to the power spectrum derived therein. It must be noted that we work with Fourier modes defined with respect to the coordinates and metric of the *flat* background, that is, when the curvature is treated as a separate perturbation introduced on top of the flat reference space.

One caveat of our analysis is that we will be assuming a Bunch-Davies initial state in calculating correlators. In a curved universe, the Euclidean vacuum instead gives a modified initial state (see for example [144–146, 141]). The difference between these two initial states is non-perturbative in *K* and therefore cannot be captured by the arguments in this section. We have performed a preliminary investigation of the relative importance of the initial state modification in canonical inflation as compared with the perturbative corrections using the analytical results of [143]. We found that the non-perturbative terms give an effect that is numerically negligible in the final primordial power spectrum for the parameters relevant to this work. This is to be expected as the deviation from Bunch-Davies is non-perturbative in curvature, and consistency forces us to remain in the perturbative regime. Because of this we will systematically neglect these corrections.

5.4.1 The power spectrum

The effect of spatial curvature to the power spectrum can be calculated at linear order in K by summing up two physical effects: the presence of the ultra-long mode $\tilde{\zeta}_B$ and a change in the Hubble parameter.

The effect of the ultra-long mode was computed to leading order in K in [112] by considering the 3-pt function (in the EFT of inflation) in the squeezed limit, as explained in the previous subsection. The result is

$$\Delta_{\tilde{\zeta}_B} P_K(k) = -P_{\text{flat}}(k) \frac{19 + 6c_3}{8c_s^2} \frac{K}{k^2}, \qquad (5.70)$$

The effect of the change in Hubble parameter can be found from the familiar relation $P_{\text{flat}}(k) = H_*^2/(4\epsilon k^3)$. To leading order,

$$\Delta_H P_K(k) = P_{\text{flat}}(k) \frac{K}{k^2}, \qquad (5.71)$$

so this effect is subdominant for small c_s .

In addition to the above there is yet another effect of a more geometrical nature. When observations of the sky are performed, some assumptions have to be made regarding the connection between position space and Fourier space power spectra. While this correspondence is unambiguous in flat space (up to constants of proportionality), in a spatially curved universe there exist multiple conventions that could lead to slightly different Fourier space results.¹⁰ All the power spectra must give the same flat space limit and their ratio is a purely geometrical quantity that can only depend on K and the momentum, not on the physical quantities such as the EFT parameters and in particular c_s . Thus, in general the "geometrical" effects contribute

$$\Delta_{\text{geom.}} P_K(k) = P_{\text{flat}}(k) \mathcal{O}(K/k^2), \qquad (5.72)$$

and are again subdominant for small c_s . We will neglect both (5.71) and (5.72) in the analysis in Sec. 5.5.

In conclusion, to leading order in $c_s \to 0$ the only contributing term is $\delta_{\tilde{\zeta}_G}$. So to linear order in K the power spectrum becomes

$$\zeta_{\mathbf{k}}^{\text{flat}} = \int d^3 x e^{-i\mathbf{k}\cdot\mathbf{x}} \zeta^{\text{flat}}(x)$$

and compare this with the Fourier transform of perturbations on top of a curved background, with respect to the coordinate system defined from the apparent distance to objects in a curved universe,

$$\zeta_{\mathbf{k}} = \int d^3 r e^{-i\mathbf{k}\cdot\mathbf{r}} \zeta(r).$$

Because $\zeta(r) \neq \zeta^{\text{flat}}(x=r)$ (rather, there is a discrepancy of order Kx^2), the two Fourier transforms are not equivalent.

¹⁰As an example, consider *flat Fourier modes*, defined as the Fourier transform of perturbations on a flat reference background:



Figure 5.2: Three-point function diagrams for EFT of inflation. The ultra-long mode k_U mocks the effect of spatial curvature.

$$P_K(k) = A_s k^{(n_s - 1) - 3} \left(1 - \frac{19 + 6c_3}{8c_s^2} \frac{K}{k^2} \right) + \mathcal{O}(Kc_s^0).$$
(5.73)

5.4.2 Background curvature argument

Let us assume that K > 0, and later we can analytically continue our results to K < 0. Recall from the discussion in the beginning of this section that we can trade curvature for the following spherically-symmetric perturbation around flat space,

$$\tilde{\zeta}_B(t, \mathbf{x}) = -\frac{1}{4} K \mathbf{x}^2 + \mathcal{O}((K \mathbf{x}^2)^2), \qquad (5.74)$$

plus a modification in the scale factor, i.e. $a_{\text{flat}} = a_K$. This implies that, for example, a three-point function of ζ in a curved universe equals the same three-point function in the background of $\tilde{\zeta}_B$ in a flat universe (with a modified scale factor $a_K(t)$). ¹¹, i.e.

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle_{K,H_K} = \langle \zeta_{\mathbf{q}} \zeta_{\mathbf{k}} \zeta_{-\mathbf{k}} \rangle_{\tilde{\zeta}_B,H_K} + \mathcal{O}((K\mathbf{x}^2)^2), \tag{5.75}$$

where correlators are assumed to be taken in flat space unless otherwise specified by the label "K". In the following we will keep implicit in all formulae that the scale factor should be a_K and we will come back to this issue at the end of this section. The configuration given by $\tilde{\zeta}_B$ can be mimicked by the long mode limit of the superposition of three orthonormal plane waves (all having the same momentum magnitude $|p_a| = \sqrt{K/2}$ for a = 1, 2, 3, and with $\hat{p}_a \cdot \hat{p}_b = \delta_{ab}$) subtracted with an inconsequential constant,

¹¹Here we assume that the geometric effects stemming from the curvature of spatial slices are negligible with respect to the enhanced effects originating from non-gaussianity, e.g. we still expand ζ in terms of plane waves and ignore that the sum of spatial momenta does not vanish.

$$\tilde{\zeta}_B(t, \mathbf{x}) = \sum_{a=1}^3 \left[\cos(\mathbf{p}_a \cdot \mathbf{x}) - 1 \right] + \mathcal{O}((K\mathbf{x}^2)^2) \,. \tag{5.76}$$

It can be easily seen that the expression on the RHS starts at order p^2 and that it coincides with (5.74). The three-point function on a slowly varying background ζ_B (inclusive of $\tilde{\zeta}_B$) is given by

$$\langle \zeta_{\mathbf{k}_{1}} \zeta_{\mathbf{k}_{2}} \zeta_{\mathbf{k}_{3}} \rangle_{\zeta_{B}} \cong \langle \zeta_{\mathbf{k}_{1}} \zeta_{\mathbf{k}_{2}} \zeta_{\mathbf{k}_{3}} \rangle + \zeta_{B} \left[\frac{\partial}{\partial \zeta_{B}} \langle \zeta_{\mathbf{k}_{1}} \zeta_{\mathbf{k}_{2}} \zeta_{\mathbf{k}_{3}} \rangle_{\tilde{\zeta}_{B}} \right]_{\zeta_{B}=0}$$

$$+ \partial_{i} \zeta_{B} \left[\frac{\partial}{\partial (\partial_{i} \zeta_{B})} \langle \zeta_{\mathbf{k}_{1}} \zeta_{\mathbf{k}_{2}} \zeta_{\mathbf{k}_{3}} \rangle_{\zeta_{B}} \right]_{\zeta_{B}=0}$$

$$+ \partial_{i} \partial_{j} \zeta_{B} \left[\frac{\partial}{\partial (\partial_{i} \partial_{j} \zeta_{B})} \langle \zeta_{\mathbf{k}_{1}} \zeta_{\mathbf{k}_{2}} \zeta_{\mathbf{k}_{3}} \rangle_{\zeta_{B}} \right]_{\zeta_{B}=0} + \dots$$

$$(5.77)$$

Therefore,

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle_{\tilde{\zeta}_B} = \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle - \frac{3}{2} K \delta_{ij} \left[\frac{\delta}{\delta(\partial_i \partial_j \zeta_B)} \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle_{\zeta_B} \right]_{\zeta_B = 0} .$$
(5.78)

For simplicity we drop $[...]_{\zeta_B=0}$ in the remainder. In order to find the last term in the expression above, we consider the correlation between an ultra-long monochromatic mode, $\tilde{\zeta}_{\mathbf{p}}(t, \mathbf{x}) = \zeta_{\mathbf{p}}(t) \exp(i\mathbf{p} \cdot \mathbf{x})$, and three short modes. Up to leading order in the gradients of the ultra-long mode, this trispectrum can be simplified into

$$\begin{aligned} \langle \zeta_{\mathbf{p}} \zeta^{3} \rangle &= P_{\zeta}(p) \left[\frac{\partial}{\partial \zeta_{B}} \langle \zeta^{3} \rangle_{\zeta_{B}} + ip^{i} \cdot \frac{\partial}{\partial (\partial_{i} \zeta_{B})} \langle \zeta^{3} \rangle_{\zeta_{B}} - p^{i} p^{j} \frac{\partial}{\partial (\partial_{i} \partial_{j} \zeta_{B})} \langle \zeta^{3} \rangle_{\zeta_{B}} + \mathcal{O}(p^{3}) \right] \\ &= \mathbf{S}.\mathbf{T}. + P_{\zeta}(p) \left[-p^{i} p^{j} \frac{\partial}{\partial (\partial_{i} \partial_{j} \zeta_{B})} \langle \zeta^{3} \rangle_{\zeta_{B}} + \mathcal{O}(p^{3}) \right], \end{aligned}$$

where "S.T." stands for the $\mathcal{O}(p^0)$ and $\mathcal{O}(p)$ parts of the correlator, which are fixed by soft theorems [108]. Summing over three orthogonal **p**'s one finds ¹²

$$\sum_{i=1}^{3} \langle \zeta_{\mathbf{p}_{i}} \zeta^{3} \rangle = \mathbf{S}.\mathbf{T}. + P_{\zeta}(p) \delta_{ij} \left[-p^{2} \frac{\delta}{\delta(\partial_{i}\partial_{j} \zeta_{B})} \langle \zeta^{3} \rangle_{\zeta_{B}} + \mathcal{O}(p^{3}) \right].$$
(5.80)

Equivalently, we can write

$$\delta_{ij} \frac{\delta}{\delta(\partial_i \partial_j \zeta_B)} \langle \zeta^3 \rangle_{\zeta_B} = -\lim_{p \to 0} \frac{1}{2} \frac{\partial^2}{\partial p^2} \sum_{i=1}^3 \left[P(p)^{-1} \langle \zeta_{\mathbf{p}_i} \zeta^3 \rangle \right].$$
(5.81)

¹²Since we are interested in $\mathcal{O}(K)$ corrections, it is allowed to send $p \to 0$ although we had assumed $p = \sqrt{K/2}$.

Finally, by putting together (5.75), (5.77) and (5.81) we find

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle_K \cong \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle_0 + \frac{3}{2} K \lim_{p \to 0} \frac{1}{2} \frac{\partial^2}{\partial p^2} \sum_{i=1}^3 P(p)^{-1} \left[\langle \zeta_{\mathbf{p}_i} \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle \right] \,. \tag{5.82}$$

5.4.3 Effects of curvature in the bispectrum: the EFT of inflation

The effective field theory (EFT) of inflation [147, 136] is the general theory of single-field fluctuations around any FLRW spacetime. In place of the covariant scalar field ϕ , the action is expressed in terms of the Goldstone mode of the broken time translation, defined through $\phi(t, \mathbf{x}) = \overline{\phi}(t + \pi)$ (where $\overline{\phi}$ is the background of the scalar field). Asking π to non-linearly realize the temporal diffs and linearly realize the spatial ones fixes the dynamics up to some arbitrary time dependent functions.

At sufficiently short distances, the physics of π decouples from the metric perturbations. In this chapter, we will assume that the time dependence of all background quantities is sufficiently slow that we can approximate them as constant¹³. The EFT action in the decoupling limit up to cubic order in slow-roll corrections and up to quartic order in the field becomes

$$S_{\pi} = \int d^{4}x \sqrt{-g} \, \frac{\epsilon H^{2} M_{p}^{2}}{c_{s}^{2}} \qquad \left[\left(\dot{\pi}^{2} - c_{s}^{2} \frac{(\nabla \pi)^{2}}{a^{2}} \right) + C_{\dot{\pi}(\nabla \pi)^{2}} \dot{\pi} \frac{(\nabla \pi)^{2}}{a^{2}} + C_{\dot{\pi}^{3}} \dot{\pi}^{3} + C_{(\nabla \pi)^{4}} \frac{1}{a^{4}} (\nabla \pi)^{4} + C_{\dot{\pi}^{2}(\nabla \pi)^{2}} \frac{1}{a^{2}} \dot{\pi}^{2} (\nabla \pi)^{2} + C_{\dot{\pi}^{4}} \dot{\pi}^{4} \right].$$
(5.83)

As for the cubic operators, the coefficient $C_{\pi(\nabla \pi)^2}$ is entirely determined by demanding the non-linear realization of Lorentz boosts

$$C_{\dot{\pi}(\nabla\pi)^2} = c_s^2 - 1\,,\tag{5.84}$$

while C_{π^3} is a free time-dependent function and is conventionally parametrised as

$$C_{\pi^3} = (1 - c_s^2) \left(1 + \frac{2}{3} \frac{c_3}{c_s^2} \right).$$
(5.85)

As for the quartic interactions, invariance under boosts relates two of the coefficients to the cubic operators, namely

$$C_{(\nabla\pi)^4} = -C_{\dot{\pi}(\nabla\pi)^2}, \qquad C_{\dot{\pi}^2(\nabla\pi)^2} = -\frac{3}{2} (C_{\dot{\pi}(\nabla\pi)^2} - C_{\dot{\pi}^3}), \qquad (5.86)$$

¹³Notice that this is not the same as assuming a shift symmetry, as discussed in details in [121].

whereas, $C_{\dot{\pi}^4}$ is a free coefficient. Inasmuch as the $\dot{\pi}^4$ operator is unconstrained by the constraints on the bispectrum, it is allowed to be much bigger than $\dot{\pi}^2(\partial_i \pi)^2$ and $(\partial_i \pi)^4$ [148]. For this reason, in the remainder we only keep the contribution of $\dot{\pi}^4$ to the trispectrum.

Now we turn to computing the bispectrum of ζ in the EFT of inflation to leading order in K and slow-roll corrections by using the method explained in 5.4.2. The trispectrum generated by operators in (5.84) was calculated in [139]. The cubic operators contribute to the trispectrum via four types of exchange diagrams depicted in Figure 5.3. Below, we separately give the bispectrum generated by individual Feynmann diagrams in Figure 5.3, namely $B_{\pi^3\pi^3}$, corresponding to the exchange diagram with two $\dot{\pi}^3$ vertices, $B_{\pi^3\pi(\nabla\pi)^2}$ for the diagram with one $\dot{\pi}^3$ and one $\dot{\pi}(\nabla\pi)^2$ vertex, and finally $B_{\pi(\nabla\pi)^2\pi(\nabla\pi)^2}$ representing the diagram with two $\dot{\pi}(\nabla\pi)^2$ vertices. Since the final answer is symmetric under permutations of momenta, we give the expressions in terms of the elementary symmetric polynomials, defined by

$$e_1 = \sum_{i=1}^3 k_i, \qquad e_2 = \sum_{i < j} k_i k_j, \qquad e_3 = \prod_{i=1}^3 k_i.$$
 (5.87)

We find

$$B_{\dot{\pi}^{3}\dot{\pi}^{3}} = 18 \left(\frac{H^{2}}{4\epsilon c_{s}M_{p}^{2}}\right)^{2} \left(\prod_{i=1}^{3} \frac{1}{2k_{i}^{3}}\right) \frac{K}{c_{s}^{4}e_{1}^{5}} (-1+c_{s}^{2})^{2} (2c_{3}+3c_{s}^{2})^{2} \times \left(-2e_{3}e_{1}^{3}+e_{2}^{2}e_{1}^{2}+3e_{2}e_{3}e_{1}+76e_{3}^{2}\right),$$
(5.88)

$$B_{\pi^{3}\pi(\nabla\pi)^{2}} = \frac{3}{2} \left(\frac{H^{2}}{4\epsilon c_{s} M_{p}^{2}} \right)^{2} \left(\prod_{i=1}^{3} \frac{1}{2k_{i}^{3}} \right) \frac{K}{c_{s}^{4} e_{3}^{2} e_{1}^{5}} (-1+c_{s}^{2})^{2} (2c_{3}+3c_{s}^{2}) \times (5.89)$$

$$\left[6e_{3}e_{1}^{9} - 3e_{2}^{2}e_{1}^{8} - 18e_{2}e_{3}e_{1}^{7} + 9 \left(e_{2}^{3} + 5e_{3}^{2} \right) e_{1}^{6} - 57e_{2}^{2}e_{3}e_{1}^{5} - 1184e_{3}^{4} + 12e_{2} \left(e_{2}^{3} + e_{3}^{2} \right) e_{1}^{4} + 4 \left(3e_{2}^{3}e_{3} - 107e_{3}^{3} \right) e_{1}^{3} + 232e_{2}^{2}e_{3}^{2}e_{1}^{2} + 672e_{2}e_{3}^{3}e_{1} \right] ,$$

$$B_{\pi(\nabla\pi)^{2} \pi(\nabla\pi)^{2}} = \frac{3}{2} \left(\frac{H^{2}}{4\epsilon c_{s} M_{p}^{2}} \right)^{2} \left(\prod_{i=1}^{3} \frac{1}{2k_{i}^{3}} \right) \frac{K}{c_{s}^{4} e_{1}^{5} e_{3}^{2}} \times$$

$$\left[19 \left(2e_{3}e_{1}^{4} - e_{2}^{2}e_{1}^{3} - 6e_{2}e_{3}e_{1}^{2} + 3e_{2}^{3}e_{1} - 19e_{2}^{2}e_{3} \right) e_{1}^{5} + 405e_{3}^{2}e_{1}^{6} - (44e_{2}e_{3}^{2} - 76e_{2}^{4})e_{1}^{4} + 4(19e_{3}e_{2}^{3} + 253e_{3}^{3})e_{1}^{3} - 452e_{2}^{2}e_{3}^{2}e_{1}^{2} - 1508e_{2}e_{3}^{3}e_{1} + 2256e_{3}^{4} \right] .$$

$$(5.90)$$

The only contact term that we consider is generated via the operator $\dot{\pi}^4$, and the resulting bispectrum

is ¹⁴

$$B_{\dot{\pi}^4} = 3 \times 36 \times 96 \left(\frac{H^2}{4\epsilon c_s M_p^2}\right)^2 \left(\prod_{i=1}^3 \frac{1}{2k_i^3}\right) K \left[C_{\dot{\pi}^4} - \frac{9}{4} \left(1 + \frac{2}{3}\frac{c_3}{c_s^2}\right)^2\right] \frac{e_3^2}{e_1^5}.$$
 (5.91)

Squeezed limit in the EFT of inflation

Let us compute the squeezed limit of the curvature corrections to the bispectrum in a curved universe, in the framework of the EFT of inflation. In particular, we want to find the dominant $\mathcal{O}(K)$ corrections to the bispectrum in the regime $c_s \ll 1$ in terms of the two EFT quantities c_s and c_3 that parameterize the 3-point vertices. This can be done by taking the squeezed limit of (5.88)-(5.90). For a more explicit derivation we can evaluate the trispectrum

$$\langle \zeta_{\mathbf{q}_{\mathbf{u}}} \zeta_{\mathbf{q}_{\mathbf{l}}} \zeta_{\mathbf{k}_{\mathbf{s}}} \zeta_{\mathbf{k}_{\mathbf{s}'}} \rangle$$
 (5.92)

in the double squeezed limit

$$q_u \ll q_l \ll k_s \sim k_{s'} \,, \tag{5.93}$$

and use the background curvature argument to find the $\mathcal{O}(K)$ term in the bispectrum. We leave the details of the calculation to Appendix 5.7.3. In conclusion, we find the following curvature corrections to the squeezed EFT bispectrum:

$$\frac{B(q_l, |\mathbf{k}_s - \frac{1}{2}\mathbf{q}_l|, |\mathbf{k}_s + \frac{1}{2}\mathbf{q}_l|)}{P_{\zeta}(q_l)P_{\zeta}(k_s)} \sim (1 - n_s) + \frac{c_s^2 - 1}{c_s^2} \left[\left(2 + \frac{1}{2}c_3 + \frac{3}{4}c_s^2 \right) - \frac{5}{4}(\hat{\mathbf{q}}_l \cdot \hat{\mathbf{k}}_s)^2 \right] \frac{q_l^2}{k_s^2}$$

$$(5.94)$$

$$+ \frac{3}{2}c_s^{-4} \left[\left(\frac{3}{4}c_3^2 + \frac{43}{8}c_3 + \frac{19}{2} \right) - \left(\frac{15}{8}c_3 + \frac{95}{16} \right) (\hat{\mathbf{q}}_l \cdot \hat{\mathbf{k}}_s)^2 \right] \frac{K}{k_s^2},$$

where the first line is the flat-space result obeying Maldacena's consistency relation and the second line is the curvature correction. Notice that there is an interesting relation between the leading-order curvature correction in the squeezed bispectrum and in the power spectrum. If we average the second line of (5.94) over the angle θ between $\hat{\mathbf{q}}_l$ and $\hat{\mathbf{k}}_s$, the term $(\hat{\mathbf{q}}_l \cdot \hat{\mathbf{k}}_s)^2$ reduces to a factor of 1/3 and we find

¹⁴Notice that in the interaction Hamiltonian the coefficient of the $\dot{\pi}^4$ term differs from the one in the original Lagrangian due to the correction that the conjugate momentum of π receives from the cubic term $\dot{\pi}^3$.

$$\int d\cos\theta \,\Delta_{\tilde{\zeta}_B} B = 2 \left(\Delta_{\tilde{\zeta}_B} P\right)^2 \,, \tag{5.95}$$

where $\Delta_{\tilde{\zeta}_B} P$ is the leading-order curvature correction to the power spectrum discussed around (5.70) and $\Delta_{\tilde{\zeta}_B} B$ is the second line of (5.94). This can be understood in various ways. For example, recall that from the wavefunction of the universe ψ , which takes the form

$$\psi = \exp\left[-\sum_{n=2}^{\infty} \int \frac{1}{n!} \psi_n \zeta^n\right] , \qquad (5.96)$$

we can derive the following expressions for the correlators

$$P(k) = \frac{1}{2\text{Re}\psi_2(k)},$$
(5.97)

$$B(q_u, k_s, k_{s'}) = -\frac{1}{4} \frac{\text{Re}\psi_3(q_u, k_s, k_{s'})}{\text{Re}\psi_2(q_u)\text{Re}\psi_2(k_s)\text{Re}\psi_2(k_{s'})},$$
(5.98)

$$T(\vec{q}_{u}, \vec{q}_{l}, \vec{k}_{s}, \vec{k}_{s'}) = -\frac{1}{8\text{Re}\psi_{2}(q_{u})\text{Re}\psi_{2}(q_{l})\text{Re}\psi_{2}(k_{s})\text{Re}\psi_{2}(k_{s'})} \times \left[\text{Re}\psi_{4}(\vec{q}_{u}, \vec{q}_{l}, \vec{k}_{s}, \vec{k}_{s'}) + (5.99) -\frac{\text{Re}\psi_{3}(q_{u}, q_{l}, k_{I})\text{Re}\psi_{3}(k_{s}, k_{s'}, k_{I})}{(5.100)} + 2\text{ perm's}\right].$$

$$\frac{\mathrm{e}\psi_3(q_u, q_l, k_I)\mathrm{Re}\psi_3(k_s, k_{s'}, k_I)}{\mathrm{Re}\psi_2(k_I)} + 2 \text{ perm's} \right], \qquad (5.100)$$

where $\vec{k}_I = \vec{q}_u + \vec{q}_l$. These expressions are valid for any momenta, but we have chosen the momenta to match the derivation of curvature effects from flat-universe correlators. Because we found that the flat-universe four-point interaction does not contribute to curvature effects in the squeezed bispectrum, we can neglect ψ_4 above and we see that T is related to B^2 . Also, only the permutation displayed contributes in the relevant limit $q_u \ll q_l \ll k_s \sim k'_s$ and $k_I \simeq q_l$. Following the strategy outlined earlier in this section, we can extract from the flat universe T the curvature correction to B in a curved universe by averaging over the direction of \vec{q}_u . Upon doing this, we see that one of the ψ_3 on the right-hand side of (5.100), which can be traded for B, gets also angle-averaged and becomes $\Delta_{\vec{\zeta}_B} P$. To get to (5.95) we need to also angle average over \vec{q}_l , which transforms the second ψ_3 into a second factor of $\Delta_{\vec{\zeta}_B} P$.

5.4.4 Effects of curvature in the bispectrum: canonical, single-field inflation

We now study Einstein gravity coupled to a single scalar inflaton field, which gives a simple action

$$S = \frac{1}{2} \int d^4x \sqrt{-g} \left(R - (\nabla \phi)^2 - 2V(\phi) \right).$$
 (5.101)

We would like to compute $\mathcal{O}(K)$ contributions to the bispectrum $\langle \zeta_{\mathbf{q}_l} \zeta_{\mathbf{k}_s} \zeta_{-\mathbf{k}_s-\mathbf{q}_l} \rangle'$ in the soft limit $q_l \ll k_s$, working to leading order in the slow-roll parameters. According to (5.82), we need to find the $\mathcal{O}(p^2)$ term¹⁵ in $P(p)^{-1} \langle \zeta_{\mathbf{p}} \zeta_{\mathbf{q}} \zeta_{\mathbf{k}} \zeta_{\mathbf{k}'} \rangle$ and sum over directions of \mathbf{p} .

For the theory with the action (5.101), the following diagrams contribute to the scalar trispectrum:

- the contact interaction [137],
- the scalar-exchange diagram,
- the graviton-exchange diagram [138].

The scalar-exchange diagram is subleading in the slow-roll parameters [138]. In Appendix 5.7.2, we show that the graviton exchange contribution vanishes identically after summing over the directions of p. In the same appendix we show that the contact contribution starts at order p^2 - hence we avoid the ambiguities described in footnote 15 - and we compute the relevant coefficient.

The final result for small K, in the regime $q_l \ll k_s$, to leading order in slow-roll parameters, is

$$\langle \zeta_{\mathbf{q}_l} \zeta_{\mathbf{k}_s} \zeta_{-\mathbf{k}_s} \rangle_K' \cong P(q_l) P(k_s) \left[(1 - n_s) + \mathcal{O}(q_l/k_s) + \frac{27}{16} \epsilon \frac{K}{k_s^2} \left(14(\hat{\mathbf{k}}_s \cdot \hat{\mathbf{q}}_l)^2 - 13 \right) \right].$$
(5.102)

The power spectra in the above expression are the *flat* power spectra, i.e. those evaluated in the absence of curvature. The $O(q_l/k_s)$ terms are fixed by the flat-space soft theorem and are not directly affected by K. Note that in canonical single-field inflation, the O(K) correction to the squeezed bispectrum is strongly suppressed by the slow-roll parameter ϵ .

It is arguably more elegant to express the right hand side in terms of curved universe quantities, so rather than use the flat universe power spectrum, we should use $P_K(k)$ (the curved-universe power spectrum evaluated for the canonical single-field scenario). The slow-roll parameter ϵ can be neglected relative to the scalar tilt $(1 - n_s)$ because current data already imposes the small hierarchy $\epsilon/(1 - n_s) < 1/6$ (which in turn implies conformal invariance of all correlators [90]), we have

¹⁵Note that in general, Taylor expansion in p might be ambiguous, at least at first sight. On one hand, we can impose a constraint $\mathbf{k}' = -\mathbf{k} - \mathbf{q} - \mathbf{p}$ and then write expansion coefficients as functions of \mathbf{q} and \mathbf{k} . On the other hand - just as an example - we can define $\mathbf{s} := \mathbf{k} + \frac{1}{2}\mathbf{p}$, impose $\mathbf{k}' = -\mathbf{s} - \mathbf{q} - \frac{1}{2}\mathbf{p}$ and then write expansion coefficients as functions of \mathbf{q} and \mathbf{k} . On the other hand - just as an example - we can define $\mathbf{s} := \mathbf{k} + \frac{1}{2}\mathbf{p}$, impose $\mathbf{k}' = -\mathbf{s} - \mathbf{q} - \frac{1}{2}\mathbf{p}$ and then write expansion coefficients as functions of \mathbf{q} and \mathbf{s} . The two results will generically differ at order p^2 provided that the lower order Taylor coefficients (namely, $\mathcal{O}(p^0)$ and $\mathcal{O}(p^1)$) do not vanish. Nonetheless, it turns out that in the case under consideration ambiguities can be neglected, as we will show in Appendix 5.7.2.

 $P_K(k) = P(k) \left(1 - \frac{K}{k^2}\right)$ and

$$\langle \zeta_{\mathbf{q}_l} \zeta_{\mathbf{k}_s} \zeta_{-\mathbf{k}_s} \rangle_K' \cong P_K(q_l) P_K(k_s) \left[(1 - n_s) + \mathcal{O}(q_l/k_s) + (1 - n_s) \frac{K}{q_l^2} \right].$$
(5.103)

5.4.5 An explanation of the scaling in the squeezed bispectrum

In this section, we give a heuristic derivation of the c_s dependence in the squeezed bispectrum $B(q_l, k_s, k'_s)$, $q_l \ll k_s \sim k'_s$. We wish to compare the behaviour of the $\mathcal{O}(q_l^2/k_s^2)$ terms to that of the leading curvature terms $\mathcal{O}(K/k_s^2)$, mainly to demonstrate how the latter effect is enhanced (relative to the former) by the small speed of sound.

We begin by summarizing the size of the standard, flat-universe corrections of order q_l^2/k_s^2 in the EFT of inflation [112]:

$$\left\langle \zeta_{\mathbf{q}_{l}} \zeta_{\mathbf{k}_{s}-\mathbf{q}_{l}/2} \zeta_{\mathbf{k}_{s}+\mathbf{q}_{l}/2} \right\rangle \sim P(q_{l}) P(k_{s}) \frac{q_{l}^{2}}{k_{s}^{2}} \frac{c_{s}^{2}-1}{c_{s}^{2}} \left[\left(2 + \frac{1}{2}c_{3} + \frac{3}{4}c_{s}^{2}\right) - \frac{5}{4} (\mathbf{\hat{q}}_{l} \cdot \mathbf{\hat{k}}_{s})^{2} \right] + \dots \quad (5.104)$$

We note the c_s^{-2} enhancement for $c_s \to 0$. This scaling can actually be understood without performing the full calculation by tracking the powers of c_s .

After re-introducing the term enforced by Maldacena's consistency relation, we schematically have for $q \ll k$

$$\langle \zeta_{\mathbf{q}_l} \zeta_{\mathbf{k}_s - \mathbf{q}_l/2} \zeta_{\mathbf{k}_s + \mathbf{q}_l/2} \rangle \sim P_u P_l \left[(1 - n_s) + \mathcal{O}(1) c_s^{-2} \frac{q_u^2}{q_l^2} \right].$$
(5.105)

Let us now discuss the $\mathcal{O}(K)$ correction to the squeezed bispectrum, which might arise from the double soft limit of the trispectrum according to formula (5.82). We are only concerned with the double soft limit at 0^{th} , 1^{st} and 2^{nd} order in the ultra-long momentum q_u . We expect that for a small speed of sound c_s , the $\mathcal{O}(q_u^2/k_s^2)$ contribution to the double-squeezed trispectrum will be enhanced by some negative powers of c_s .

Here we give a transparent argument for the scaling of the $O(q_u^2/k_s^2)$ term and reproduce that scaling by tracing the origin of the term to concrete diagrams. At tree-level, the trispectrum receives three contributions: scalar exchange, graviton exchange and contact interaction. Let us discuss them in turn.

Scalar exchange This diagram can have poles in the total "energy" $k_t = \sum_{a=1}^{4} k_a$ and in the momentum k_I of the exchanged scalar. The largest contribution comes from exchanging the softest



Figure 5.3: The scalar exchange and contact diagrams for the four-point function in the EFT of inflation. k_U corresponds to the ultra-long mode that mimics the effect of spatial curvature.

possible particles, corresponding to the s-channel with q_u , q_l coming in, merging into a scalar propagator $k_I \sim q_l$ and coming out as $\vec{k}_s \sim -\vec{k}'_s$. There are two cubic interactions within the diagram, and for each one we get

$$q_u + q_l \to k_I \sim q_l \quad \Rightarrow \quad P_u P_l \left[(1 - n_s) + \frac{1 - c_s^2 q_u^2}{c_s^2 q_l^2} \right],$$
 (5.106)

$$k_s + k_s \to k_I \sim q_l \quad \Rightarrow \quad P_s P_l \left[(1 - n_s) + \frac{1 - c_s^2}{c_s^2} \frac{q_l^2}{k_s^2} \right].$$
 (5.107)

The resulting trispectrum is (notice that P_l comes only once as it is in the internal "propagator" that connects the two vertices)

$$T \sim P_u P_l P_s \left[(1 - n_s)^2 + \frac{\left(1 - c_s^2\right)\left(1 - n_s\right)}{c_s^2} \left(\frac{q_u^2}{q_l^2} + \frac{q_l^2}{k_s^2}\right) + \frac{1 - c_s^2}{c_s^4} \frac{q_u^2}{k_s^2} \right].$$
(5.108)

Graviton exchange While in general graviton exchange does contribute to the squeezed bispectrum, we will show that the contribution always vanishes if we sum over the directions of \mathbf{k}_u , as we need to do if we want to interpret the ultra-long mode as a spatial curvature.

We have already seen that the graviton exchange contribution to the trispectrum vanishes after summing over directions of the ultra-long mode in the case of canonical single-field inflation, but this fact holds more generally. Consider the cubic vertex in the graviton-exchange diagram. This vertex is of the form $\pi^2 \gamma$, possibly with some derivatives. Now, π is a scalar, but $\gamma \equiv \gamma_{ij}$ is a transverse traceless tensor. We need to contract the *i*, *j* indices with derivatives ∂_i , ∂_j , and since γ_{ij} is transverse, the only lowest order self-consistent operator, up to total derivatives, is $(\partial_i \pi)(\partial_j \pi)\gamma_{ij}$. (Higher-order operators can be constructed by acting with additional time derivatives or pairs of spatial derivatives.)

The γ_{ij} field operator gives rise to a polarization tensor $\epsilon_{ij}^{s}(\mathbf{k})$ in the correlator and ∂_{i} gives rise to momentum k_{i} . The vertex factor of our lowest-order operator when one of the π legs is the ultra-long mode will be thus proportional to

$$q_i k_j \epsilon_{ij}^s (\mathbf{q} + \mathbf{k}) = -q_i q_j \epsilon_{ij}^s (\mathbf{q} + \mathbf{k}), \qquad (5.109)$$

which can be shown to always vanish after summing over three orthogonal directions of q.

Contact interaction The contact interaction has only poles in k_t , which goes as $k_t \sim k_s + k'_s$ in the double squeezed limit. One cannot have any $1/q_l$ enhancement, i.e. the contact interaction can never give any $\mathcal{O}(K/q_l^2)$ contribution to the squeezed bispectrum.

This interaction might be universal or model dependent, as in the EFT of inflation. The universal part must obey a soft theorem

$$T^{\text{grav}} \sim P_u P_l P_s (1 - n_s) \left[1 + \frac{q_u^2}{k_s^2} \right] \dots$$
 (5.110)

where $(1 - n_s)$ is a proxy for slow-roll suppressed terms that arise when performing a scaling transformation on the bispectrum. Note the absence of q_u^2/q_l^2 terms, due to the fact that there can be no poles in the long momentum q_l .

The model-dependent part on the other hand, can contribute to the squeezed limit

$$T^{EFT} \sim CP_u P_l P_s \frac{q_u^2}{k_s^2},\tag{5.111}$$

with some overall amplitude C that can be large $(C \gg (1 - n_s))$. In the squeezed bispectrum, this leads to

$$B \sim CP_l P_s \frac{q_u^2 \zeta_u}{k_s^2} \sim CP_l P_s \frac{K}{k_s^2} \,. \tag{5.112}$$

While this is the general expectation, single field inflation is an exception. In this case, as we mentioned before, it is only the operator $\dot{\pi}^4$ that can give a large trispectrum, and the result is

$$T \cong \frac{1}{q_1 q_2 q_3 q_4 k_t^5},\tag{5.113}$$

which does not have the form in (5.111), but rather in the double squeezed limit it behaves as

$$T^{EFT} \sim CP_u P_l P_s \frac{q_u^2}{k_s^2} \frac{q_l^2}{k_s^2} \,. \tag{5.114}$$

We quote a more explicit result [139] in Appendix 5.7.3. The effect on the bispectrum is then to give a very small correction in the soft limit:

$$B \sim CP_l P_s \frac{K}{k_s^2} \frac{q_l^2}{k_s^2},$$
 (5.115)

which does not violate Maldacena's consistency relation.

5.5 Observational constraints on curvature

In this section, we discuss the possibility of constraining or detecting spatial curvature through measurements of the $O(K/c_s^2)$ corrections to the primordial power spectrum and bispectrum. In the case of slow-roll, canonical single-field inflation, the corrections are $O(K/k_s^2)$ and suppressed by a factor of $(1 - n_s)$; it is very unlikely that we could detect this signal in the conceivable future. However, in the EFT of inflation with a small speed of sound, $c_s \ll 1$, curvature effects can potentially be observable. We will estimate how large curvature corrections can be given the current separate constraints on curvature and primordial non-Gaussianity.

Our treatment of curvature corrections is valid at linear order in K. When curvature corrections become large, we can no longer neglect higher-order effects in K/c_s^2 . We will find that current bounds allow for a large magnitude of the $\mathcal{O}(K)$ corrections. This means that the curvature effects *might* be significant and their comparison with observation could provide further constraints on curvature and the speed of sound. On the other hand, we will find that for the power spectrum one needs to go beyond the analysis in this chapter and perform a non-perturbative calculation in K/c_s^2 .

5.5.1 Constraints from the power spectrum

As we saw in Sec. 5.4.1, the curvature corrections to the power spectrum are enhanced in the presence of large non-Gaussianity. Before putting bounds on these corrections, we derive here the current bounds coming from the CMB bispectrum on non-Gaussian parameters and the related coefficients in the EFT of inflation. The EFT of inflation (in a flat universe) predicts the following equilateral and



Figure 5.4: 68%, 95%, and 99.7% confidence regions in the parameter space $(f_{NL}^{equil}, f_{NL}^{orth})$.



Figure 5.5: 68%, 95%, and 99.7% confidence regions in the single-field inflation parameter space (c_s, c_3), obtained from Figure 5.4 via the change of variables in Eq. (5.116) - (5.117), while allowing $c_s > 1$.

orthogonal non-Gaussianities in the bispectrum [149]:

$$f_{NL}^{equil} = (c_s^{-2} - 1) \left[-0.275 - 0.0780 \left(c_s^2 + 2/3c_3 \right) \right],$$
(5.116)

$$f_{NL}^{orth} = \left(c_s^{-2} - 1\right) \left[0.0159 + 0.0167 \left(c_s^2 + 2/3c_3\right)\right].$$
(5.117)

The Planck collaboration [149] found (for T+E channels, with lensing not substracted):

$$f_{NL}^{equil} = -25 \pm 47,$$
 $f_{NL}^{orth} = -47 \pm 24.$ (5.118)

We use this observational data to find and plot the 68%, 95% and 99.7% confidence regions for f_{NL}^{equil} and f_{NL}^{orth} under a simplifying assumption that the f_{NL} covariance matrix is diagonal, which is a good approximation (Figure 5.4). Next, we use (5.116) - (5.117) to map the confidence regions to the $\mathbf{c} \equiv (c_s, c_3)$ parameter space (Figure 5.5), by noting that for any region A in the parameter space, $\mathbb{P}(\mathbf{c}(\mathbf{f}_{NL}) \in \mathbf{c}(A)) = \mathbb{P}(\mathbf{f}_{NL} \in A)$ (we allow $c_s > 1$ for sake of simplicity; in fact, most of the relevant confidence regions do lie within $c_s < 1$). We see that the bispectrum likelihood peaks in the region $0.02 < c_s < 0.1, c_3 \sim \mathcal{O}(1)$.

Then we would like to constrain the curvature-induced modification of the power spectrum, given



Figure 5.6: The red, circular areas represent the 68%, 95% and 99.7% probability regions of the joint PDF of Ω_K and the non-gaussianity parameter $(19 + 6c_3)/(8c_s^2)$, derived from the Planck 2018 data. The green, hyperbolic areas show the 68%, 95% and 99.7% probability regions constrained by the measurements of $C_{l=2,3,4}$. Finally, the black contour in the centre of the plot shows the approximate region of validity of the linear theory used in this work.

in (5.73), based on the measurements of the CMB temperature angular power spectrum. Let us begin by estimating the signal-to-noise ratio in C_l^{TT} 's,

$$(S/N)_2 \equiv \sqrt{\sum_l \frac{(C_l^{\text{th}} - C_l^{\text{fid}})^2}{(\Delta C_l)^2}},$$
 (5.119)

where C_l^{th} and C_l^{fid} are the temperature angular power spectrum derived from (5.73) for $K \neq 0$ and K = 0, respectively, and ΔC_l is the cosmic variance in C_l .

As could be anticipated from the scaling of (5.73), $(S/N)_2$ is dominated by the low- ℓ multipoles, where we can use the Sachs-Wolfe transfer function to estimate $C_l^{\text{th},16}$ We find

$$(S/N)_2 \sim 3.2 \times \Omega_K \left(\frac{19+6c_s}{8c_s}\right) \sqrt{\sum_{\ell>1} \frac{4l+2}{9(l^2+l-2)^2}} \sim \Omega_K \left(\frac{19+6c_s}{8c_s}\right) \,. \tag{5.120}$$

¹⁶Since the dipole ($\ell = 1$) is degenerate with the Earth peculiar motion, we discard it. We also neglect slow-roll suppressed terms in our estimation of the signal-to-noise ratio.

Since more than 90% of signal-to-noise ratio is contained in l = 2, 3 and 4, we use the latest measurements¹⁷ of $C_{2,3,4}$ [150] and find

$$\frac{\Omega_K(19+6c_3)}{8c_8^2} = -0.78 \,{}^{+1.9}_{-0.6}\,. \tag{5.121}$$

In Figure 5.6 we compare this result (green shaded regions) with Planck's latest bounds on Ω_K and non-Gaussianities¹⁸ (red shaded region, where we combined the two independent constraints). Naively, this plot seems to show that within the red confidence region allowed by Planck, there are regions that are excluded by the power spectrum constraints on curvature corrections. However, we should notice that the curvature corrections to the power spectrum in (5.73) are derived only at linear order in K and so they should be trusted as long as

$$\frac{|\Omega_K(19+6c_3)|}{8c_s^2} \ll 1.$$
(5.122)

To guide the eye, in Figure 5.6 we plot a black line where this parameter takes the value 1/2. From the plot it is clear that the validity of our theoretical calculation is slightly more constraining than the CMB temperature power spectrum. In other words, current power spectrum data allows for a curvature correction that goes beyond the linear regime. In order to improve upon Planck's limits on the combination of parameters in (5.121), one would need to compute the power spectrum to higher order in K/c_s^2 . This is beyond the scope of this work but it is certainly interesting for future research.

5.5.2 Forecast of constraints from the bispectrum

In this subsection, we revisit the results of Section 5.4 and estimate the magnitude of the dominant $\mathcal{O}(K)$ corrections to the bispectrum given the current constraints on c_s and c_3 .

Squeezed limit

Recall the leading-order behaviour of the bispectrum expressed in terms of flat Fourier modes, (5.94), in the regime $q_l \ll k_s$, $c_s \ll 1$ and to linear order in K:

¹⁷Notice that we use the flat universe prediction for the transfer functions used to compute C_l as opposed to the transfer functions in a curved universe. The difference is only of order Ω_K , much smaller than the Ω_K/c_s^2 effect that we are after here.

¹⁸For $c_s \ll 1$, the linear combination $\alpha f_{NL}^{eq} + \beta f_{NL}^{orth}$ is an unbiased estimator for $z = \frac{3.2(19+6c_3)}{8c_s^2}$ (for $\alpha = -20.78, \beta = 118.5$).

$$\frac{B(q_l, |\mathbf{k}_s - \frac{1}{2}\mathbf{q}_l|, |\mathbf{k}_s + \frac{1}{2}\mathbf{q}_l|)}{P_{\zeta}(q_l)P_{\zeta}(k_s)} \sim (1 - n_s) + \frac{c_s^2 - 1}{c_s^2} \left[(2 + \frac{1}{2}c_3 + \frac{3}{4}c_s^2) - \frac{5}{4}(\mathbf{\hat{q}}_l \cdot \mathbf{\hat{k}}_s)^2 \right] \frac{q_l^2}{k_s^2} + \left[B_{K1} + B_{K2}(\mathbf{\hat{q}}_l \cdot \mathbf{\hat{k}}_s)^2 \right] \frac{K}{k_s^2},$$
(5.123)

where

$$B_{K1} \equiv \frac{3}{16} \left(6c_3^2 + 43c_3 + 76 \right) c_s^{-4} = \frac{3}{16} \left(6c_3 + 19 \right) \left(c_3 + 4 \right) c_s^{-4}, \tag{5.124}$$

$$B_{K2} \equiv -\frac{15}{32} \left(6c_3 + 19\right) c_s^{-4}.$$
(5.125)

to leading order in slow-roll coefficients and in c_s . The presence of an overall factor of $(6c_3 + 19)$ in each of the terms is expected, since it arises from the left-hand vertices (those connected to the ultralong mode) in Figure 5.3. In the above formula, we neglected contributions due to the contact interactions, originating from terms of the form $\dot{\pi}^4$ in the EFT action, because they contribute at subleading order, namely $\mathcal{O}(Kq_l^2/k_s^4)$.

The magnitude of the dominant $\mathcal{O}(K)$ correction depends only on the values of the EFT coefficients c_s and c_3 , which have been partially constrained (see Figure 5.5 and 5.6). We would like to answer two questions:

- Are the coefficients B_{K1} and B_{K2} sufficiently large for the associated $\mathcal{O}(K)$ effect to have a significant signal-to-noise ratio?
- Are the coefficients B_{K1} and B_{K2} sufficiently large for the curvature term $(\mathcal{O}(K/k_s^2))$ to be at least comparable to the flat space correction $(\mathcal{O}(q_l^2/k_s^2))$?

To answer the first question, we consider the ratios between the curvature signal and the flat space $O(q_l^2/k_s^2)$ contribution to the squeezed limit:

$$R_1 := \frac{K}{q_l^2} \frac{B_{K1}}{\left(1 - c_s^{-2}\right) \left(2 + \frac{1}{2}c_3 + \frac{3}{4}c_s^2\right)},$$
(5.126)

$$R_2 := -\frac{K}{q_l^2} \frac{B_{K2}}{\frac{5}{4} \left(1 - c_s^{-2}\right)}.$$
(5.127)

In fact, to leading order in c_s^{-1} we have

$$R_1 = R_2 = \frac{3}{8} \frac{6c_3 + 19}{c_s^2} \frac{K}{q_l^2},$$
(5.128)



Figure 5.7: 68%, 95%, and 99.7% confidence regions for (B_{K1}, B_{K2}) , obtained from Figure 5.5 via change of variables (log scale). Both axes have been snipped so as to show the entire region of interest in a single plot.

so the ratios are identical to the power spectrum effect. We conclude that, within the validity of linear-order treatment of curvature, the curvature corrections cannot be larger than the flat-space corrections ($O(q_l^2/k_s^2)$) to the squeezed limit. It should also be noted that these two corrections have different scaling and so they are in principle distinguishable.

To answer the second question, we map the constraints (confidence regions) for c_s and c_3 obtained from Planck bispectrum data to the constraints (confidence regions) for B_{K1} and B_{K2} .¹⁹ Figure 5.7 shows the approximate 68%, 95% and 99.7% confidence regions for (B_{K1}, B_{K2}) , obtained by mapping the confidence region from the (c_s, c_3) parameter space. We see that the magnitude of the coefficients is bounded, and an order-of-magnitude estimate is $|B_{K1,2}| \leq 10^6 \sim 10^7$. We can then give an upper bound to the signal-to-noise ratio of the squeezed component of the $\mathcal{O}(K)$ correction to the bispectrum. For the estimate, we will take the maximal value consistent with the constraints, $B_{K1,2} = 10^7$, and $|\Omega_K| = 10^{-3}$. We obtain

$$\left(\frac{S}{N}\right)_{\text{squeezed}} \simeq \sqrt{\int_{H_0}^{\infty} dq_l \, q_l \, \int_{k_s^{\min}}^{\infty} dk_s \int_{-1}^{1} d\cos\theta \, \frac{B_{K1,K2}^2 P_{\zeta}^2(q_l) P_{\zeta}^2(k_s)}{P_{\zeta}(q_l) P_{\zeta}^2(k_s)} \left(\frac{|K|}{k_s^2}\right)^2 \tag{5.129}$$

¹⁹It is convenient that the map is one-to-one.



Figure 5.8: The posterior probability distribution function (unnormalized) of the sign $(b_{K,equil}) \log_{10}(|b_{K,equil}|)$ parameter, derived from the Planck bispectrum data, assuming a flat prior on $(f_{NL}^{equil}, f_{NL}^{orth})$. Blue line shows the pdf for $b_{K,equil} > 0$ while the red line shows the pdf for $b_{K,equil} < 0$.

$$\sim |B_{K1,K2}\Omega_{K,0}|\Delta_{\zeta} \left(\frac{H_0}{k_s^{\min}}\right)^{3/2} \ll 10^7 \cdot 10^{-3} \cdot 10^{-5} \sim 0.1,$$
 (5.130)

where k_s^{\min} is the smallest k_s we want to allow in the squeezed limit $k_s^{\min} \gg q_l \ge H_0$. Notice that we didn't impose the requirement that the linear term in K is smaller than the zeroth order term. We conclude that, accounting for current bounds, the curvature corrections to the squeezed bispectrum are too small to be detected, even if we were able to extrapolate the leading-order corrections beyond its regime of validity.

Equilateral configuration

Linear curvature correction to the bispectrum is suppressed by k_s^{-2} in the squeezed limit, so we may hope to obtain a larger signature if we consider an equilateral configuration of the momenta. We use the same background-curvature argument as in Section 5.4. We need to consider the trispectrum in the limit where one of the modes becomes very long, while the other three form an equilateral shape, $q_u \ll k_2 = k_3 = k_4 \equiv k$:

$$\langle \zeta_{\mathbf{q}_{u}} \zeta_{\mathbf{k}_{2}} \zeta_{\mathbf{k}_{3}} \zeta_{\mathbf{k}_{4}} \rangle = \langle \zeta_{\mathbf{q}_{u}} \zeta_{\mathbf{k}_{2}} \zeta_{\mathbf{k}_{3}} \zeta_{\mathbf{k}_{4}} \rangle_{SE} + \langle \zeta_{\mathbf{q}_{u}} \zeta_{\mathbf{k}_{2}} \zeta_{\mathbf{k}_{3}} \zeta_{\mathbf{k}_{4}} \rangle_{GE} + \langle \zeta_{\mathbf{q}_{u}} \zeta_{\mathbf{k}_{2}} \zeta_{\mathbf{k}_{3}} \zeta_{\mathbf{k}_{4}} \rangle_{CI}.$$
(5.131)

The graviton exchange does not contribute to the bispectrum correction. The scalar exchange and contact interaction diagrams will generically both lead to a significant $O(K/k^2)$ effect.

The scalar-exchange diagrams give, to leading order in c_s :

$$B(k,k,k)_{K,SE} = P_{\zeta}(k)^2 \left[\frac{65}{6} \left(1 + \frac{2}{3} \frac{c_3}{c_s^2} \right)^2 + \frac{5321}{432} \left(1 + \frac{2}{3} \frac{c_3}{c_s^2} \right) c_s^{-2} + \frac{4747}{216} c_s^{-4} \right] \times \frac{K}{k^2}.$$
(5.132)

The contact interactions give, to leading order in c_s :

$$B(k,k,k)_{K,CI} = \frac{16K}{3k^2} P_{\zeta}(k)^2 \left[C_{\dot{\pi}^4} - \frac{9}{4} \left(1 + \frac{2}{3} \frac{c_3}{c_s^2} \right)^2 \right].$$
 (5.133)

We first focus on the enhancement that cubic interactions induce by neglecting the $C_{\dot{\pi}^4}$ coefficient. In this case, the curvature correction to the bispectrum in the equilateral configuration becomes

$$B(k,k,k)_K \approx P_{\zeta}(k)^2 \left(-0.519c_3^2 + 16.42c_3 + 21.98\right) c_s^{-4} \times \frac{K}{k^2}.$$
 (5.134)

Let

$$b_{K,\text{equil}} := \left(-0.519c_3^2 + 16.42c_3 + 21.98\right)c_s^{-4}.$$
(5.135)

Assuming a uniform prior for the equilateral and orthogonal non-Gaussianities f_{NL}^{equil} and f_{NL}^{orth} in the flat-universe approximation and working in the limit $c_s \ll 1$, we find the posterior pdf for $b_{K,\text{equil}}$ by mapping the corresponding pdf for the non-Gaussianities and marginalizing over c_3 . The posterior pdf for $b_{K,\text{equil}}$ is shown on Figure 5.8. We see that $b_{K,\text{equil}} \sim 10^8$ is still allowed by current constraints.

Let us switch off the cubic terms and consider the contribution of the quartic term $\dot{\pi}^4$ to the equilateral bispectrum,

$$\frac{B_{K,CI}(k,k,k)}{P_{\zeta}^{2}(k)} \sim \frac{16C_{\dot{\pi}^{4}}}{3} \frac{K}{k^{2}}.$$
(5.136)

The Planck observational constraint on the trispectrum [149] implies the following rough upper bound on the coefficient in front of K/k_s^2 ,

$$\frac{16}{3} |C_{\pi^4}| < 10^7 \,. \tag{5.137}$$

This is slightly weaker than the maximum size allowed for $b_{K,equil}$ and so it can be neglected for an order of magnitude estimate. Let us now derive a rough upper bound on the signal-to-noise ratio for measuring spatial curvature in the 3D bispectrum of ζ as follows:

$$\left(\frac{S}{N}\right)_{\text{equil.}} \simeq \sqrt{\int_{H_0}^{\infty} dk \, k^2 \, \frac{b_{K,\text{equil}}^2 P_{\zeta}^4(k)}{P_{\zeta}^3(k)} \left(\frac{|K|}{k^2}\right)^2} \tag{5.138}$$

$$\sim |b_{K,\text{equil}}\Omega_{K,0}|\Delta_{\zeta}$$
 (5.139)

Notice that the integral in dk is strongly supported on the largest observable scales, $k = H_0$, and so it is insensitive to the UV cutoff in k, which we have taken to infinity for simplicity. If we naively used $|b_{K,equil}| = 10^8$, a proxy for the largest value allowed by the current constraints depicted in Figure 5.8, and $|\Omega_{K,0}| = 10^{-3}$, we would get that the signal is barely detectable $S/N \sim 10^{8-3-5} = 1$. From this estimate we conclude that the corrections to the equilateral configurations are unlikely to be detectable even if we extrapolate beyond the validity of the linear-order treatment in K. On the other hand, if we insist that the curvature correction is smaller than the K^0 term, then we can at most take $|b_{K,equil}\Omega_{K,0}| \sim f_{NL}^{eq,ort} \sim 10^2$. We then find a very small signal-to-noise ratio, $S/N \sim 10^{-3}$. The intuitive reason why the curvature corrections are so hard to detect is that, although they can be very large for the largest observable scales, the non-scale invariant signal drops very quickly on shorter scales and the signal-to-noise ratio saturates with just a few multipoles. Conversely, the traditional non-Gaussianity such as equilateral and orthogonal shapes are scale invariant and they can be constrained by all B_{l_1,l_2,l_3} .

5.6 Discussion and conclusion

In this work, we have discussed the effects that spatial curvature induces in primordial correlators, at linear order. As explained in Figure 6.1, these effects are parameterised by Ω_K/c_s^2 and so could be large if c_s was small during inflation or equivalently if perturbations interacted more strongly. We have shown that in the presence of curvature, the soft limit of the bispectrum acquires model-dependent corrections that deviate from Maldacena's consistency relation in flat space. More generally, we have argued that residual diffeomorphisms are separated from the spectrum of physical perturbations by a gap of order |K| and so standard soft theorems should be violated at linear order in curvature. We have furthermore studied how large these corrections can be both in the power spectrum and in the bispectrum, also going beyond the squeezed limit. We have found that in the power spectrum, constraints from the CMB are close to but slightly weaker than the validity of our linear order treatment of curvature. For the bispectrum on the other hand, the signal-to-noise for these corrections is always smaller than one, even assuming we had access to a full 3D map of the primordial correlators to the cosmic variance limit.

There are a few avenues for future research. First, as discussed in Sec. 5.5.1, we could not harvest the full constraining power of the CMB because our theoretical prediction was limited to linear-order in K/c_s^2 , in which regime the corrections to the temperature angular power spectrum are slightly smaller than the experimental bound. It would therefore be very interesting to compute the power spectrum of primordial perturbations with small c_s in a curved universe to all orders in K and then compare the prediction again with the lowest CMB multipoles. Such calculation was performed for $c_s = 1$ in [145, 146, 143]. In that work, the authors also accounted for the initial state dictated by the Euclidean continuation for a bubble nucleation event, that is different from the Bunch-Davies state we have used in this work. The difference is non perturbative in K and our preliminary analysis shows that it might be safe to neglect this effect, but a more detailed study should be performed. Second, it would also be interesting to use the polarisation of the CMB and analyse C_l^{EE} and C_l^{TE} to improve the constraints on curvature corrections to the power spectrum. Because polarisation is so small on Hubble scales, we do not expect a large improvement from this additional data. In fact, it seems likely that C_l^{EE} will give a negligible improvement because the error bars on the first few l's are so large, while C_{i}^{TE} might improve the bounds by a few tens of percent, after the covariance has been taken into account. Third, one might extend our analysis of the CMB power spectrum to the bispectrum, including both temperature and polarisation but one should be aware that our rough estimate for the signal-to-noise ratio in the bispectrum is much smaller than one within the regime of validity of the linear theory; even going beyond this regime, the ratio is at most $\mathcal{O}(1)$ for the largest allowed values of parameters. Finally, in this work we have focussed on single-field inflation because of our interest in discussing the squeezed limit consistency relation. But one could investigate curvature effects in multifield inflation, where there is more room for producing a large signal.

5.7 Appendices

5.7.1 Perturbations around FLRW

Here we collect useful formulas on the kinematics of FLRW spacetimes. The non-vanishing Christoffel symbols in the coordinates used in (5.9) are

$$\Gamma_{ij}^{0} = a\dot{a}\tilde{g}_{ij}, \qquad (5.140)$$

$$\Gamma_{0i}^{i} = H \,\delta_{ij}, \qquad (5.140)$$

$$\Gamma_{ij}^{k} = \frac{1}{2}Kf\left(x^{k}\delta_{ij} - 2x^{(i}\delta_{j)k}\right).$$

On spatial sections, the Laplacian operator acting on scalars is

$$\nabla^2 S = \frac{1}{f^2} \left(\partial_i \partial_i S - \frac{1}{2} K f x^i \partial_i S \right) .$$
 (5.141)

The isometry group of a curved FLRW spacetime is SO(3,1) if it is open, and SO(4) if it is closed. The Killing vectors are

$$T_{i} = f(2f - 1)c_{i} + \frac{K}{2}f^{2}c_{j}x^{j}x^{i}, \qquad (5.142)$$
$$R_{i} = f^{2}\omega_{ij}x^{j}, \qquad \omega_{(ij)} = 0.$$

Sometimes the T_i 's are called *quasi-translations*, i.e. they become ordinary spatial translations in flat space, while the R_i 's form rotations.

The Scalar-Vector-Tensor Decomposition

The scalar-vector-tensor decomposition of a tensor field living on a constant curvature manifold with the metric

$$ds^{2} = f^{2}(K\mathbf{x}^{2}) \, d\mathbf{x} \cdot d\mathbf{x} = f^{2}(Kr^{2})(dr^{2} + r^{2}d\Omega^{2}), \tag{5.143}$$

is an old topic of interest in differential geometry. Here we briefly review the decomposition of a tensor on a sphere or a hyperboloid, along the lines of [151]. We would like to determine under what condition the Poisson equation,

$$\nabla^2 \Phi = J(x) \,, \tag{5.144}$$

admits a unique solution Φ for any given J. If two solutions existed, their difference would solve the Laplace equation, namely

$$\nabla^2 \Phi = 0. \tag{5.145}$$

Multiplying both sides with Φ and integrating over the whole space, one finds

$$-\int_{\mathcal{M}} \sqrt{g} \, d^3 x \, (\nabla_i \, \Phi)^2 + \int_{\partial \mathcal{M}} \sqrt{h} \, d\theta \, d\phi \, \Phi \, n^i \, \nabla^i \Phi = 0 \,, \tag{5.146}$$

where g_{ij} is the metric of the curved space, g is its determinant, $\partial \mathcal{M}$ is empty for a sphere and is a 2-sphere for a hyperboloid. For the latter, n^i is the unit vector normal to the boundary 2-sphere and h_{ab} $(a, b = \theta, \phi)$ is the induced metric on the boundary 2-sphere. Therefore, on a 3-sphere, (5.146) implies that $\nabla_i \Phi = 0$ everywhere, giving as only solution $\Phi = \text{const.}$ Let us move to the hyperboloid. If Φ decays rapidly enough towards the boundary, namely

$$f \Phi \partial_r \Phi \to 0 \quad \text{for} \quad r \to \frac{2}{\sqrt{|K|}},$$
(5.147)

then the only solution of the Laplace equation is $\Phi = \text{const}$, and therefore the solution of the Poisson equation is unique up to a constant. This in turn implies that the splitting of a vector into a longitudinal and a transverse part, i.e.

$$A_i = \nabla_i \phi + A_i^T, \qquad \nabla^i A_i^T = 0, \qquad (5.148)$$

is unique $i\!f\!f \int \sqrt{g} d^3x \, \nabla_i A^i$ is finite.

Rank-2 objects should be dealt with more carefully because the non-commutation of covariant derivatives brings about some complication. Consider the following decomposition

$$H_{ij} = H_S^{(1)} g_{ij} + \nabla_i \nabla_j H_S^{(2)} + 2\nabla_{(i} H_{j)}^V + H_{ij}^T, \qquad (5.149)$$

in which

$$g^{ij}H_{ij}^T = \nabla^i H_{ij}^T = \nabla^i H_i^V = 0.$$
(5.150)

It is easy to check that

$$H_i^i = 3H_S^{(1)} + \nabla^2 H_S^{(2)}, \qquad (5.151)$$
$$\nabla^{-2} \nabla^i \nabla^j H_{ij} = H_S^{(1)} + (\nabla^2 + 2K) H_S^{(2)}.$$

Assuming a proper asymptotic decay of H_{ij} , $H_S^{(1,2)}$ can be uniquely fixed up to

$$H_S^{(2)} \to H_S^{(2)} + \chi, \quad H_S^{(1)} \to H_S^{(1)} + K\chi,$$
 (5.152)

where χ is any solution of

$$(\nabla^2 + 3K)\chi = 0. (5.153)$$

It remains to solve for H_i^V by taking the divergence of H_{ij} ,

$$2\nabla^{i} \nabla_{(i} H_{j)}^{V} = \nabla^{i} H_{ij} - \nabla_{j} \left(\nabla^{-2} \nabla^{k} \nabla^{l} H_{kl} \right).$$
(5.154)

However, the homogeneous equation, i.e. $\nabla^i \nabla_{(i} H_{j)}^V = 0$, could admit non-trivial solutions. Multiplying the latter with H_j^V and integrating over the space yields

$$\int \sqrt{g} \, d^3x \, H^{V\,i} \, \nabla^j \nabla_{(i} H^V_{j)} = \int_{\partial M} \sqrt{h} d\theta d\phi \, H^{V\,i} \, \nabla_{(i} H^V_{j)} - \frac{1}{2} \int \sqrt{g} d^3x \, (\nabla_{(i} H^V_{j)})^2 = 0 \, .$$

Therefore, for both a hyperboloid—assuming that H_i^V vanishes quickly enough near the boundary— -and a sphere, the boundary term vanishes, and as a result the solutions consist only of the Killing vectors. However, Killing vectors of a hyperboloid are not bounded, hence we must discard them. In conclusion, the SVT decomposition of H_{ij} is unique up to (5.152) and separately

$$H_i^V \to H_i^V + \xi_i \,, \quad \nabla_{(i}\,\xi_{j)} = 0 \,.$$
 (5.155)

Notice that for all ξ_i 's we have $\nabla^2 \xi_i = -2K \xi_i$. Thus, as long as one works with eigenfunctions of the ∇^2 operator with eigenvalues unequal to -3K (for scalars) and -2K (for vectors), the SVT

decomposition is unique and well defined.

The Spectrum of The Laplacian Operator

For K < 0 the eigenfunctions of the Laplacian operator for scalars, defined through

$$\nabla^2 Y_{plm} = (1+p^2) K Y_{plm} , \qquad (5.156)$$

are given by (see e.g.[144])

$$Y_{plm} = \frac{\Gamma(ip+l+1)}{\Gamma(ip+1)} \frac{p}{\sqrt{|K|r f(Kr^2)}} P_{ip-1/2}^{-l-1/2} \left(\sqrt{1+K^2r^2f^2(Kr^2)}\right) Y_{lm}(\theta,\phi),$$

where $P^{\mu}_{\nu}(x)$ are Associated Legendre functions of the first kind, and $\Gamma(x)$ is the Euler Gamma function. Only mode functions with p > 0 ($-\nabla^2 > |K|$) are square integrable and in this sense physical. They are also normalized and orthogonal, i.e.

$$\int r^2 f^3 dr d\Omega Y_{plm} Y^*_{p'l'm'} = \delta(p - p') \delta_{mm'} \delta_{ll'} .$$
(5.157)

For K > 0 the spectrum of scalar harmonics is ([152])

$$\nabla^2 Y_{plm} = -p(p+2)KY_{plm}, \quad p = 0, 1, ..., \text{ and } l = 0, ..., p.$$
 (5.158)

and they are given by

$$Y_{plm} = \sqrt{\frac{(p+1)\Gamma(p+l+2)}{\Gamma(p-l+1)}} \frac{p}{\sqrt{Kr f(|K|r^2)}} P_{p-1/2}^{-l-1/2} \left(\sqrt{1-K^2r^2f^2(Kr^2)}\right) Y_{lm}(\theta,\phi),$$

with the same property as (5.157) except that $\delta_{pp'}$ replaces $\delta(p - p')$.

5.7.2 Canonical Trispectrum in the soft limit

In this appendix, our goal is to compute the ${\cal O}(q_u^2)$ term in the pure scalar 4-point function

$$\left\langle \zeta_{\mathbf{q}_{\mathbf{u}}} \zeta_{\mathbf{q}_{\mathbf{l}}} \zeta_{\mathbf{k}_{\mathbf{s}}} \zeta_{\mathbf{k}_{\mathbf{s}}'} \right\rangle. \tag{5.159}$$

[137, 138] give explicit formulas for this 4-point function derived under the following assumptions:

- Inflation is described by a single field with Lagrangian $\mathcal{L} = -\frac{1}{2}(\nabla \phi)^2 V(\phi)$.
- All interactions are that of the inflaton minimally coupled to Einstein gravity.
- Terms that are suppressed by higher powers of slow-roll parameters can be neglected, ie. we work to leading order in ϵ and η .²⁰

The 4-point function has a contribution due to the contact interaction [137] as well as due to graviton exchange [138]:

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \zeta_{\mathbf{k}_4} \rangle = \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \zeta_{\mathbf{k}_4} \rangle_{CI} + \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \zeta_{\mathbf{k}_4} \rangle_{GE}.$$
(5.160)

We are interested in the double squeezed limit, and we will take $\mathbf{q}_u \equiv \mathbf{k}_1$, $\mathbf{q}_u \equiv \mathbf{k}_2$, $\mathbf{k}_s \equiv \mathbf{k}_3$, $\mathbf{k}'_s \equiv \mathbf{k}_4$. We have to compute the term proportional to q_u^2 .

The graviton exchange

The contribution to the trispectrum from the graviton exchange is given by

$$\langle \zeta_{\mathbf{k}_{1}} \zeta_{\mathbf{k}_{2}} \zeta_{\mathbf{k}_{3}} \zeta_{\mathbf{k}_{4}} \rangle^{GE} = (2\pi)^{3} \delta(\sum_{a} \mathbf{k}_{a}) \frac{H_{*}^{6}}{\epsilon^{2} \prod_{a} (2k_{a}^{3})} k_{1}^{2}$$
(5.161)

$$\times \left[\frac{k_{3}^{2}}{k_{12}^{3}} [1 - (\hat{\mathbf{k}}_{1} \cdot \hat{\mathbf{k}}_{12})^{2}] [1 - (\hat{\mathbf{k}}_{3} \cdot \hat{\mathbf{k}}_{12})^{2}] \cos 2\chi_{12,34} \cdot (\mathcal{I}_{1234} + \mathcal{I}_{3412})$$
$$+ \frac{k_{2}^{2}}{k_{13}^{3}} [1 - (\hat{\mathbf{k}}_{1} \cdot \hat{\mathbf{k}}_{13})^{2}] [1 - (\hat{\mathbf{k}}_{2} \cdot \hat{\mathbf{k}}_{13})^{2}] \cos 2\chi_{13,24} \cdot (\mathcal{I}_{1324} + \mathcal{I}_{2413})$$
$$+ \frac{k_{2}^{2}}{k_{14}^{3}} [1 - (\hat{\mathbf{k}}_{1} \cdot \hat{\mathbf{k}}_{14})^{2}] [1 - (\hat{\mathbf{k}}_{2} \cdot \hat{\mathbf{k}}_{14})^{2}] \cos 2\chi_{14,23} \cdot (\mathcal{I}_{1423} + \mathcal{I}_{2314}) \right]$$

where $\mathcal{I}_{abcd} + \mathcal{I}_{cdab}$ are given below, and $\chi_{12,34}$ is the angle between the plane defined by the vectors \mathbf{k}_1 , \mathbf{k}_2 and the plane defined by the vectors \mathbf{k}_3 and \mathbf{k}_4 . Although the above formula may not look *manifestly* invariant under relabelling of the momenta, we verified it to be unaffected by permutations.

²⁰In particular, the tree diagram involving an exchange of a scalar is suppressed by additional power of slow-roll parameters. Thus, at tree level we only need to consider the contact interaction and the graviton-exchange diagram.

$$\mathcal{I}_{1234} + \mathcal{I}_{3412} = \frac{k_1 + k_2}{a_{34}^2} \left[\frac{1}{2} (a_{34} + k_{12}) (a_{34}^2 - 2b_{34}) + k_{12}^2 (k_3 + k_4) \right] + (1, 2 \leftrightarrow 3, 4) \quad (5.162)$$
$$+ \frac{k_1 k_2}{k_t} \left[\frac{b_{34}}{a_{34}} - k_{12} + \frac{k_{12}}{a_{12}} \left(k_t^{-1} + a_{12}^{-1} \right) \right] + (1, 2 \leftrightarrow 3, 4)$$
$$- \frac{k_{12}}{a_{12} a_{34} k_t} \left[b_{12} b_{34} + 2k_{12}^2 (\prod_a k_a) \left(\frac{1}{k_t^2} + \frac{1}{a_{12} a_{34}} + \frac{k_{12}}{k_t a_{12} a_{34}} \right) \right],$$

where $a_{ab} \equiv k_a + k_b + k_{ab}$ and $b_{ab} \equiv (k_a + k_b)k_{ab} + k_ak_b$.

The presence of an overall factor $k_1^2 = q_u^2$ in each of the terms in (5.161) is very convenient, allowing us to divide the entire expression by q_u^2 and then evaluate the remaining part in the limit $\mathbf{q}_u \to 0$. The result will in general depend on the direction from which $\mathbf{q}_u = q_u \hat{\mathbf{q}}_u$ approaches zero.

Let's assume that $\hat{\mathbf{q}}_0$ is fixed as $q_u \to 0$. As explained in Section 5.4, we need to take the sum of the GE contribution over any three²¹ mutually orthogonal directions of $\hat{\mathbf{q}}_u$. Since $\lim_{q_u\to 0} (\mathcal{I}_{abcd} + \mathcal{I}_{cdab})$ does not depend on $\hat{\mathbf{q}}_u$ and the same holds true for k_{1i}^2 , $\hat{\mathbf{k}}_3 \cdot \hat{\mathbf{k}}_{12}$, $\hat{\mathbf{k}}_2 \cdot \hat{\mathbf{k}}_{13}$ and $\hat{\mathbf{k}}_2 \cdot \hat{\mathbf{k}}_{14}$ (since in each of these instances we can use $\mathbf{k}_{1i} \to \mathbf{k}_i$), all the dependence on $\hat{\mathbf{q}}_u$ is in the following factor:

$$\lim_{q_u \to 0} \left[\left(1 - \left(\hat{\mathbf{q}}_u \cdot \hat{\mathbf{k}}_{12} \right)^2 \right) \cos 2\chi_{12,34} \right]$$
(5.163)

and its permutations. This is equal to

$$\left(1 - (\hat{\mathbf{q}}_u \cdot \hat{\mathbf{k}}_2)^2\right) \cos 2\chi_{12,34}, \qquad (5.164)$$

which can be easily shown to vanish when averaged over any three orthogonal directions of q_u . The same applies to the other two terms in the correlation function.

In conclusion, the contribution of the graviton exchange to the $O(q_u^2)$ term vanishes completely due to averaging over angles.

²¹The final answer would better not depend on which three directions we choose! Otherwise we cannot avoid a major inconsistency.

The contact interaction

The contribution from the contact (4-vertex) interaction reads

$$\langle \zeta_{\mathbf{k}_{1}} \zeta_{\mathbf{k}_{2}} \zeta_{\mathbf{k}_{3}} \zeta_{\mathbf{k}_{4}} \rangle_{CI}' = \frac{H_{*}^{6}}{4\epsilon^{2} \prod_{a} (2k_{a}^{3})} \sum_{24 \text{ perms}} \mathcal{M}_{4}(\mathbf{k}_{1}, \mathbf{k}_{2}, \mathbf{k}_{3}, \mathbf{k}_{4}).$$
(5.165)

where

$$\mathcal{M}_{4}(\mathbf{k}_{1}, \mathbf{k}_{2}, \mathbf{k}_{3}, \mathbf{k}_{4}) = -2\frac{k_{1}^{2}k_{3}^{2}}{k_{12}^{2}k_{34}^{2}} \frac{W_{24}}{k_{t}} \left(-(\mathbf{k}_{1} \cdot \mathbf{k}_{4})(\mathbf{k}_{2} \cdot \mathbf{k}_{3}) + (\mathbf{k}_{1} \cdot \mathbf{k}_{3})(\mathbf{k}_{2} \cdot \mathbf{k}_{4}) + \frac{3}{4}\sigma_{12}\sigma_{34} \right) \\ -\frac{1}{2}\frac{k_{3}^{2}}{k_{34}^{2}}\sigma_{34} \left(\frac{\mathbf{k}_{1} \cdot \mathbf{k}_{2}}{k_{t}} W_{124} + 2\frac{k_{1}^{2}k_{2}^{2}}{k_{t}^{3}} + 6\frac{k_{1}^{2}k_{2}^{2}k_{4}}{k_{t}^{4}} \right)$$

and

$$\begin{aligned} \sigma_{ab} &= \mathbf{k}_{a} \cdot \mathbf{k}_{b} + k_{b}^{2}, \\ W_{ab} &= 1 + \frac{k_{a} + k_{b}}{k_{t}} + \frac{2k_{a}k_{b}}{k_{t}^{2}}, \\ W_{abc} &= 1 + \frac{k_{a} + k_{b} + k_{c}}{k_{t}} + \frac{2(k_{a}k_{b} + k_{b}k_{c} + k_{c}k_{a})}{k_{t}^{2}} + \frac{6k_{a}k_{b}k_{c}}{k_{t}^{3}} \end{aligned}$$

Using $P(k) = \frac{H_*^2}{4\epsilon k^3}$, (5.165) can be written as

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \zeta_{\mathbf{k}_4} \rangle_{CI}' = \epsilon P(k_1) P(k_2) P(k_3) k_4^{-3} \sum_{24 \text{ perms}} \mathcal{M}_4(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4).$$
(5.166)

Our goal is to compute $\frac{1}{2} \frac{\partial^2}{\partial q_u^2} P(q_u)^{-1} \langle \zeta_{\mathbf{q}_u} \zeta_{\mathbf{k}_s} \zeta_{\mathbf{k}'_s} \rangle'_{CI}$, so $P(q_u)$ is cancelled out, and the only factor dependent on \mathbf{q}_u is now $\sum_{24 \text{ perms}} \mathcal{M}_4(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)$. We have thus reduced the problem to finding

$$\left(\frac{\partial^2}{\partial q_u^2} \sum_{24 \text{ perms}} \mathcal{M}_4(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)\right)_{q_u=0}.$$
(5.167)

We will find it easier to deal with the $\mathcal{M}_4(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)$ permutation only, and instead consider all the 24 different bijections between $(\mathbf{q}_u, \mathbf{q}_l, \mathbf{k}_s, \mathbf{k}'_s)$ and $(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)$.

Terms $\mathcal{O}(q_u^0)$, $\mathcal{O}(q_u^1)$ Using symbolic manipulation in Mathematica, we verified that $O(q_u^0)$ and $O(q_u^1)$ terms in \mathcal{M}_4 vanish.

The contribution from $q_u \equiv k_1$ Let us first consider the six terms in which q_u is the first entry.

Note that in the limit $q_u \rightarrow 0$ we can use $\mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 = 0$.

$$\frac{\partial}{\partial q_u^2} \sum_{6 \text{ perms}} \mathcal{M}_4(\mathbf{q}_u, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = \sum_{6 \text{ perms}} \left[-2 \frac{k_3^2}{k_2^2 k_{34}^2} \frac{W_{24}}{k_t} \cdot \frac{3}{4} k_2^2 (\mathbf{k}_3 + \mathbf{k}_4) \cdot \mathbf{k}_4 - \frac{1}{2} \frac{k_3^2}{k_{34}^2} (\mathbf{k}_3 + \mathbf{k}_4) \cdot \mathbf{k}_4 \left(\frac{\partial}{\partial q_u^2} (\frac{\mathbf{q}_u \cdot \mathbf{k}_2}{k_t} W_{q_u, 24})_{q_u = 0} + 2 \frac{k_2^2}{k_t^3} (1 + 3 \frac{k_4}{k_t}) \right) \right]$$
$$= -\sum_{6 \text{ perms}} \frac{k_3^2}{k_{34}^2} (\mathbf{k}_3 + \mathbf{k}_4) \cdot \mathbf{k}_4 \left[\frac{3}{2} \frac{W_{24}}{k_t} + \frac{k_2^2}{k_t^3} (1 + 3 \frac{k_4}{k_t}) + \frac{1}{4} \left(\frac{\partial^2}{\partial q_u^2} (\frac{\mathbf{q}_u \cdot \mathbf{k}_2}{k_t} W_{q_u, 24}) \right)_{q_u = 0} \right].$$
(5.168)

Now, we have $\frac{\partial}{\partial q_u} (W_{q_u,24}/k_t)_{q_u=0} = 0$, so the last term vanishes. We get

$$\frac{\partial}{\partial q_u^2} \sum_{\text{6 perms}} \mathcal{M}_4(\mathbf{q}_u, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = \sum_{\text{6 perms}} \frac{k_3^2}{k_{34}^2} \mathbf{k}_2 \cdot \mathbf{k}_4 \left[\frac{3}{2} \frac{W_{24}}{k_t} + \frac{k_2^2}{k_t^3} \left(1 + 3\frac{k_4}{k_t} \right) \right].$$
(5.169)

After some more transformations,

$$\frac{\partial}{\partial q_u^2} \sum_{\text{6 perms}} \mathcal{M}_4(\mathbf{q}_u, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = \sum_{\text{6 perms}} \frac{k_3^2}{k_t^3} \mathbf{k}_2 \cdot \mathbf{k}_4 \left[1 + 3\frac{k_4}{k_t} + 3k_2^{-2}(k_t^2 - \frac{1}{2}k_3k_t) \right].$$
(5.170)

Let's evaluate the dominant term of the above expression for $q_l \ll k_s \sim k'_s$. If it is nonvanishing in the limit $q_l/k_s \rightarrow 0$ (we will shortly see that it is), then due to the presence of q_l in the prefactor, the $1 + 3\frac{k_4}{k_t}$ part gives a zero contribution in this limit. The other term survives as $q_l/k_s \rightarrow 0$ only if $\mathbf{q}_l \equiv \mathbf{k}_2$. We are therefore left with

$$3q_l^{-2}k_t^{-2}\left(k_s^2(\mathbf{q}_l\cdot\mathbf{k}_s')(k_t-\frac{1}{2}k_s)+k_s'^2\mathbf{q}_l\cdot\mathbf{k}_s(k_t-\frac{1}{2}k_s')\right)=-\frac{3}{8}k_s(\hat{\mathbf{k}}_s\cdot\hat{\mathbf{q}}_l)^2+\mathcal{O}(q_l).$$
 (5.171)

The contribution from $q_u \equiv k_3$ Derivation is analogous to that of the previous section, only much simpler. The contribution is

$$\frac{3}{8}k_s[5(\hat{\mathbf{k}}_s \cdot \hat{\mathbf{q}}_l)^2 - 7].$$
(5.172)

The contribution from $q_u \equiv k_2$ Using symbolic manipulation in Mathematica, we found

$$\frac{\partial^2}{\partial q_u^2} \mathcal{M}_4(\mathbf{q}, \mathbf{q}_u, \mathbf{k}, \mathbf{K})_{q_u=0} = \frac{k_s^2}{q_l^2 k_t^4} \mathbf{q}_l \cdot \mathbf{k}_s' \left[3k_t^2 (k_t + k_s') + 2q_l^2 (k_t + 3k_s') \right].$$
(5.173)

The only other permutation giving a comparable contribution is the one in which \mathbf{k}_s and \mathbf{k}'_s are

swapped. The remaining 4 permutations of q_l, k_s, k'_s lead to subdominant contributions, which we can ignore. We have

$$q_l^{-2}k_t^{-4} \left(k_s^2(\mathbf{q}_l \cdot \mathbf{k}_s') 3k_t^2(k_t + k_s') + 2q_l^2(k_t + 3k_s') + k_s^2\mathbf{q}_l \cdot \mathbf{k}_s 3k_t^2(k_t + k_s) + 2q_l^2(k_t + 3k_s)\right) = \frac{3}{4}k_s(5(\hat{\mathbf{k}}_s \cdot \hat{\mathbf{q}}_l)^2 - 3) + O(q_l).$$
(5.174)

The contribution from $q_u \equiv k_4$ Again, using symbolic manipulation in Mathematica, we found

$$\frac{\partial^2}{\partial q_u^2} \mathcal{M}_4(\mathbf{q}_l, \mathbf{k}_s, \mathbf{k}'_s, \mathbf{q}_u)_{q_u=0} = 3 \frac{q_l^2}{k_s^{\prime 2} k_t^2} (k_t + k_s) \mathbf{k}_s \cdot \mathbf{k}'_s.$$
(5.175)

The dominant contributions when $\mathbf{q}_u \equiv \mathbf{k}_4$ will actually arise from another two permutations: (k_s, k'_s, q_l) and (k'_s, k_s, q_l) . We have

$$3k_t^{-2}q_l^{-2}(k_s'^2(k_t+k_s)\mathbf{k}_s\cdot\mathbf{q}_l+k_s^2(k_t+k_s')\mathbf{k}_s'\cdot\mathbf{q}_l) = \frac{3}{4}k_s(5(\hat{\mathbf{k}}_s\cdot\hat{\mathbf{q}}_l)^2-3) + O(q_l).$$
(5.176)

Summary

After summing up all the permutations, we get

$$\left(\frac{\partial}{\partial q_u^2} \sum_{24 \text{ perms}} \mathcal{M}_4(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)\right)_{q_u=0} \sim \frac{3}{8} k_s (14(\hat{\mathbf{k}}_s \cdot \hat{\mathbf{q}}_l)^2 - 13)$$
(5.177)

for $q_l \ll k_s, k'_s$.

Result - the trispectrum contribution

Using (5.82), we get

$$\langle \zeta_{\mathbf{q}_l} \zeta_{\mathbf{k}_s} \zeta_{-\mathbf{k}_s} \rangle_K \sim P(q_l) P(k_s) \left[(1 - n_s) + \mathcal{O}(q_l/k_s) + \frac{27}{16} \epsilon \frac{K}{k_s^2} \left(14(\hat{\mathbf{k}}_s \cdot \hat{\mathbf{q}}_l)^2 - 13 \right) \right].$$
(5.178)

5.7.3 EFT Trispectrum in the soft limit

Following [139], we define the trispectrum form factor \mathcal{T} as

$$\langle \zeta^4 \rangle = 8 \times (2\pi)^3 \delta^{(3)} \left(\sum_{i=1}^4 \mathbf{k}_i \right) P_{\zeta}(k_1) P_{\zeta}(k_2) P_{\zeta}(k_3) \frac{1}{k_4^3} \mathcal{T}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) , \qquad (5.179)$$

where $P_{\zeta}(k) = \langle \zeta_{\mathbf{k}} \zeta_{-\mathbf{k}} \rangle'$. We will use the results of [139]; the parameters λ and Σ from [139] are related to c_s and c_3 in the following way:

$$\lambda = \frac{1}{2} \Sigma \left(1 + \frac{2}{3} \frac{c_3}{c_s^2} \right), \qquad (5.180)$$

$$\Sigma = \frac{\epsilon H^2}{c_s^2}.$$
(5.181)

Throughout, we assume that $c_s \ll 1$, while $c_3 \sim \mathcal{O}(1)$, so that $\lambda \sim \mathcal{O}(c_s^{-4})$, $\Sigma \sim \mathcal{O}(c_s^{-2})$.

The EFT power spectrum

For models with small speed of sound, we have

$$P_{\zeta}(k) = \frac{H_*^2}{4\epsilon M_{\eta l}^2 c_s k^3}.$$
(5.182)

Scalar-exchange diagram

The scalar exchange contributes to the trispectrum at order $c_s^{-4}q_u^2k_sP_{\zeta}^4$. In particular, there are no terms at 0^{th} or 1^{st} order in the ultralong momentum q_u , as has been verified by our Mathematica scripts. In the computations we outline in this subsection, all contributions that are subleading in the regime $q_u \ll k_l \ll k_s$ are neglected.

The dominant contributions come from the diagrams in which the exchanged momentum is $\mathbf{q}_u + \mathbf{k}_l$, corresponding to the choices $(\mathbf{q}_u, \mathbf{k}_l) \equiv (\mathbf{k}_3, \mathbf{k}_4)$ or $(\mathbf{q}_u, \mathbf{k}_l) \equiv (\mathbf{k}_1, \mathbf{k}_2)$, where (a, b) stands for an unordered pair. Hence, we only have to sum over 8, rather than 24, permutations of the momenta.

We also perform a summation over three²² directions of q_u that are mutually orthogonal, but otherwise arbitrary.

There are three types of scalar-exchange diagrams that differ by the type of vertices, shown on Fig. 5.3. For the sake of transparency, we consider each of the three cases in a separate subsection, writing out the partial contributions before presenting the final result.

The $\dot{\pi}^3 imes \dot{\pi}^3$ diagrams

These diagrams are given by (B.3) - (B.4) in [139]. It is straighforward to show that the leadingorder contribution to \mathcal{T} is

²²Of course, the $O(q_u^2)$ terms are even in \mathbf{k}_u , so we do not need to average over two opposite directions.

$$\mathcal{T}_1 = \frac{9}{32} \left(\frac{\lambda}{\Sigma}\right)^2 q_u^2 k_s = \frac{9}{128} \left(1 + \frac{2}{3} \frac{c_3}{c_s^2}\right)^2 q_u^2 k_s.$$
(5.183)

After summing over three orthogonal directions of q_u ,

$$\sum_{\hat{\mathbf{q}}_u} \mathcal{T}_1 = \frac{27}{128} \left(1 + \frac{2}{3} \frac{c_3}{c_s^2} \right)^2 q_u^2 k_s.$$
(5.184)

The $\dot{\pi}^3 \times \dot{\pi} (\partial_i \pi)^2$ and $\dot{\pi} (\partial_i \pi)^2 \times \dot{\pi}^3$ diagrams

These are given by (B.5) - (B.10) in [139]. By computing $\partial^2/\partial q_u^2$ of the sum of the eight relevant permutations, we can deduce the q_u^2 term. We find, to leading order in k_s/k_l ,

$$\sum_{\hat{\mathbf{q}}_u} \mathcal{T}_2 = \frac{3}{128} \left(1 + \frac{2}{3} \frac{c_3}{c_s^2} \right) \left(\frac{1}{c_s^2} - 1 \right) q_u^2 k_s \left(43 - 15 (\hat{\mathbf{k}}_l \cdot \hat{\mathbf{k}}_s)^2 \right).$$
(5.185)

The $\dot{\pi}(\partial_i \pi)^2 \times \dot{\pi}(\partial_i \pi)^2$ diagrams

These diagrams correspond to equations (B.11) - (B.17) from [139]. Again, after summing over the eight permutations, computing the $\partial^2/\partial q_u^2$ derivative, summing over the three directions and keeping only the terms that are leading order in k_s/k_l , we get

$$\sum_{\hat{\mathbf{q}}_u} \mathcal{T}_3 = \frac{19}{128} \left(\frac{1}{c_s^2} - 1 \right)^2 q_u^2 k_s \left(8 - 5(\hat{\mathbf{k}}_l \cdot \hat{\mathbf{k}}_s)^2 \right).$$
(5.186)

Summary for the scalar exchange The total form factor due to scalar exchange is, after summing over angles,²³

$$\mathcal{T}_{SE} = \frac{1}{8} q_u^2 k_s \left(\alpha_1 + \alpha_2 (\hat{\mathbf{k}}_l \cdot \hat{\mathbf{k}}_s)^2 \right), \qquad (5.187)$$

with α_1 and α_2 that can be expressed in terms of c_s, c_3 . If we keep only the c_s^{-4} terms, we have

$$\alpha_1 = \left(\frac{3}{4}c_3^2 + \frac{43}{8}c_3 + \frac{19}{2}\right)c_s^{-4}, \tag{5.188}$$

$$\alpha_2 = -\left(\frac{15}{8}c_3 + \frac{95}{16}\right)c_s^{-4}.$$
(5.189)

Then

²³The factor of 1/8 is introduced in order to cancel out the factor of 8 in front of T in (5.179).

$$\sum_{\hat{\mathbf{q}}_u} \langle \zeta^4 \rangle_{SE} \sim (2\pi)^3 \delta^{(3)} \left(\sum_{i=1}^4 \mathbf{k}_i \right) P_{\zeta}(q_u) P_{\zeta}(k_l) P_{\zeta}(k_s) \frac{q_u^2}{k_s^2} \left(\alpha_1 + \alpha_2 (\hat{\mathbf{k}}_l \cdot \hat{\mathbf{k}}_s)^2 \right).$$
(5.190)

The corresponding contribution to the bispectrum on a curved background is

$$\langle \zeta_{\mathbf{k}_l} \zeta_{\mathbf{k}_s} \zeta_{-\mathbf{k}_s} \rangle_{K,SE}' \sim \frac{3}{2} P_{\zeta}(k_l) P_{\zeta}(k_s) \left(\alpha_1 + \alpha_2 (\hat{\mathbf{k}}_l \cdot \hat{\mathbf{k}}_s)^2 \right) \frac{K}{k_s^2}.$$
(5.191)

The above power spectra are evaluated on a flat background. But α_i scale as c_s^{-4} while the correction to the power spectrum scales as c_s^{-2} , so it might be neglected in the regime $c_s \ll 1$. Then in (5.191) we are allowed to use the curved power spectrum.

Graviton exchange

Another tree-level contribution to the scalar 4-pt function is the graviton exchange. However, as we have shown in Section 5.4, the contribution of the lowest-order operators to the final O(K) terms in the bispectrum is always exactly zero.

Contact interaction

There is yet another contribution to the trispectrum that cannot be accounted for in the cubic action, since it originates from the contact diagram corresponding to the 4-vertex scalar interaction. The contact diagram is evaluated in [139]. In the limit $c_s \rightarrow 0$, the result is dominated by

$$\mathcal{T}_{c1} = 36 \left(C_{\dot{\pi}^4} - 9 \left(\frac{\lambda}{\Sigma} \right)^2 \right) \frac{\prod_{i=1}^4 k_i^2}{k_t^5}.$$
(5.192)

In the double squeezed limit, this gives

$$\langle \zeta^4 \rangle_{c1} \propto \left(C_{\dot{\pi}^4} - 9 \left(\frac{\lambda}{\Sigma} \right)^2 \right) P_{\zeta}(q_u) P_{\zeta}(k_l) P_{\zeta}(k_s) \frac{q_u^2 k_l^2}{k_s^4}.$$
 (5.193)

Recall that $(\lambda/\Sigma)^2 \sim \mathcal{O}(c_s^{-4})$. For the particular case of DBI inflation, C_{π^4} also scales as c_s^{-4} ; we assume, for simplicity, that the scaling of the above contribution is always c_s^{-4} . There are also other terms that scale as c_s^{-2} and are subdominant in the limit $c_s \to 0$. These are the only contributions to the trispectrum that are linear in q_u for small q_u (it has been verified in [111] that these terms reproduce the conformal consistency relation for the 4-pt function).

Summing up everything, we have (schematically)

$$\langle \zeta^4 \rangle_c \sim P_{\zeta}^3 \left[c_s^{-2} \left(\frac{q_u k_l}{k_s^2} + \frac{q_u^2}{k_s^2} \right) + c_s^{-4} \frac{q_u^2 k_l^2}{k_s^4} \right].$$
 (5.194)

Due to the nonvanishing $O(q_u)$ terms, redefinitions of the momenta will influence the $O(q_u^2)$ terms; but only at order c_s^{-2} , not c_s^{-4} :

$$\langle \zeta^4 \rangle_c^{(q_u^2)} \sim P_{\zeta}^3 \left[c_s^{-2} \frac{q_u^2}{k_s^2} + c_s^{-4} \frac{q_u^2 k_l^2}{k_s^4} \right].$$
 (5.195)

The first term in the brackets is subdominant in the limit $c_s \rightarrow 0$ relative to the scalar-exchange diagram. The second term is also subdominant, having a different momentum dependence than the leading-order scalar exchange contribution.

In conclusion, the contact interaction gives a negligible contribution to the $\mathcal{O}(K)$ correction to the squeezed bispectrum. The final result for the squeezed limit of the bispectrum in a curved universe is given by (5.191).
Chapter 6

The Boostless Bootstrap: Amplitudes without Lorentz boosts

Abstract

Poincaré invariance is a well-tested symmetry of nature and sits at the core of our description of relativistic particles and gravity. At the same time, in most systems Poincaré invariance is not a symmetry of the ground state and is hence broken spontaneously. This phenomenon is ubiquitous in cosmology where Lorentz boosts are spontaneously broken by the existence of a preferred reference frame in which the universe is homogeneous and isotropic. This motivates us to study scattering amplitudes without requiring invariance of the interactions under Lorentz boosts. In particular, using on-shell methods we show that the allowed interactions around Minkowski spacetime are severely constrained by unitarity and locality in the form of consistent factorization. Our analysis assumes massless, relativistic and luminal particles of any spin, and a restricted ansatz for the four-particle amplitude, which can be shown to be equivalent to having Lorentz covariant fields in the Lagrangian description. We find that the existence of an interacting massless spin-2 particle enforces (analytically continued) three-particle amplitudes to be Lorentz invariant, even those that do not involve a graviton, such as cubic scalar couplings. We conjecture this to be true for all *n*-particle amplitudes. Also, particles of spin S > 2 cannot self-interact nor can be minimally coupled to gravity, while particles of spin S > 1 cannot have electric charge. Given the growing evidence that free gravitons are well described by massless, luminal relativistic particles, our results imply that cubic graviton interactions in Minkowski must be those of general relativity up to a unique Lorentz-invariant higher-derivative

correction of mass dimension 9. Finally, we point out that consistent factorization for massless particles is *highly IR sensitive* and therefore our powerful flat-space results do not straightforwardly apply to curved spacetime.

6.1 Introduction and summary

Symmetry is a physicist's compass and Poincaré invariance is perhaps the most precisely tested symmetry in nature [153–156]. Empirically, we observe it everywhere: from electromagnetism to the reign of subatomic particles and the expanse of the cosmos. But just as importantly, Poincaré invariance sits at the heart of our description of the laws of nature. On the one hand, it provides us with the organizing principle to model the interactions of subatomic particles through Quantum Field Theory (QFT), and constitutes one of the pillars of the standard model of particle physics. On the other hand, Poincaré symmetry is so powerful and rigid that it makes our theoretical description inevitable. We can appreciate this from two complementary points of view.

Weinberg argues in [157] that Poincaré invariance, combined with quantum mechanics and locality (in the form of cluster decomposition), uniquely selects QFT as the necessary language of nature, at least at low energies. Moreover, from this standpoint, microscopic causality and the analyticity of the S-matrix follow from the above assumptions rather than being invoked as general principles. But fields come at a cost: the spectrum of massless particles cannot fit inside a set of Poincaré covariant fields and we are obliged to invoke unobservable "gauge" symmetries. Also, the scattering of particles cannot be uniquely mapped into the interactions of fields, as is evident in perturbative field redefinitions. These observations have motivated physicists to look for an alternative description of scattering that does not invoke fields or gauge redundancies. Modern on-shell methods for amplitudes, an intellectual descendant of the S-matrix program of the 60's (see e.g. [158]), have made tremendous progress towards precisely this goal (reviews include [30, 31, 28]). It is from this complementary point of view that the rigidity imposed by Poincaré invariance becomes once again manifest. All (analytically continued) non-perturbative three-particle amplitudes for massless fields of any spin are uniquely fixed by symmetry, and in theories such as Yang-Mills [159, 160] and general relativity [161] all higher tree-level amplitudes are uniquely determined in terms of these building blocks.

In the discussion so far we have implicitly assumed that Poincaré invariance is a symmetry of the ground state of the theory. While this is a good approximation for some particle physics applications,

the vast majority of physical systems are not Poincaré invariant in their ground state. Indeed, the specific way in which Poincaré is thus spontaneously broken determines much of the behavior of a given system. While all possibilities have been classified [162], a particularly simple and interesting case arises when the "vacuum" consists of a static, homogeneous and isotropic medium that permeates spacetime. Observers at rest with respect to this medium are special, as they observe a more symmetric configuration, hence Lorentz boosts are spontaneously broken. This is the case for many condensed matter systems but also for cosmological models as we will discuss in detail shortly. Some even go a step further and speculate about possible explicit breaking of Poincaré invariance, perhaps arising in a UV-complete theory of gravity.

The above considerations beg the question of what happens to the rigidity of the laws of nature when Poincaré invariance is not respected by the ground state, as it is for example the case in our universe at cosmological distances. If the free theory is Poincaré invariant, what can we say about interactions? In particular, we will focus on the following formulation of this question:

What boost-breaking interactions are allowed for massless, relativistic spinning particles?

This question is not just academic. Rather it's motivated by practical considerations. For example, we have recently observed that the free propagation of gravitational waves is extremely well described by the relativistic theory of a (classical) massless spin-2 particle [163]. What does this imply for the interactions that gravitons can have in a consistent theory? More precisely, in this work we will derive all possible on-shell three-particle amplitudes, and the allowed singularities of four-particle amplitudes, for relativistic, massless, luminal particles, while allowing for boost-breaking interactions. Whether Lorentz boosts are broken explicitly, or more likely only spontaneously, will be irrelevant for our discussion (see [164] for a recent discussion of Goldstone theorem for boosts). Our assumption that the free theory is Poincaré invariant leads us to a particular ansatz for four-particle amplitudes, which can be shown to be equivalent to assuming that the underlying Lagrangian is constructed out of Lorentz covariant fields with the breaking of boosts due to the freedom to add time derivatives at will. Although this does not capture the most general set of boost-breaking theories, it provides us with an excellent testing ground and already produces some surprising results. Indeed, we will find that internal consistency severely restricts the allowed set of interactions, especially in the presence of a massless spin-2 particle. We summarise our results in Section 6.1.2.

6.1.1 Motivations

Because of the very general methodology that we adopt, our results can be approached and interpreted from a variety of perspectives. In the following, we motivate our analysis from three points of view.

Cosmology The expansion of the universe spontaneously breaks time translations and boosts¹. Both breakings are manifest in many cosmological phenomena. For example, the breaking of time translations can be thought of as the root cause of the redshift of light as it travels freely across the cosmos: in the absence of time translation invariance, energy is not conserved and the energy of a free photon can change with time. The breaking of boost invariance is evident in the existence of the Cosmic Microwave Background (CMB) or the cosmic neutrino background. The CMB picks out a preferred reference frame in which the universe looks homogeneous and isotropic. The Earth moves with respect to this preferred frame and so we observe the CMB to be anisotropic to one part in a thousand. Measurements of this CMB dipole by the Planck satellite are shown in Figure 6.1 [165].

A priori, it is impossible to compare the breaking of time translations with that of boosts because the respective parameters have different dimensions²: the breaking of time translations is characterized by a certain time scale t_b , while that of boosts by a certain velocity v_b . Since in this work we will study the time-translation invariant dynamics of massless particles with broken boosts, it is important to understand under what conditions our results have a chance to be relevant for cosmology.

First, we notice that for the scattering of particles at energy E, the breaking of time translation should be parameterized by $1/(Et_b)$, which is negligible at sufficiently high energies. So in cosmology, where the characteristic time scale is the Hubble parameter, $t_b^{-1} \sim H$, time-translation invariance is often a good approximate symmetry at energies $E \gg H$. Conversely, for the scattering of massless luminal particles, which are the focus of our study, the typical center of mass velocity is always of order the speed of light. Hence, in cosmology, where the speed of light is often the characteristic speed $v_b \sim c$, the breaking of boosts can be a large effect.

Second, in many models of the very early universe and of dark energy, additional symmetries are invoked to suppress the breaking of time translations. The archetypal example is that of a so-called

¹Everywhere in this chapter we assume invariance under spacetime translations and rotations, but for conciseness we will avoid stating this repeatedly.

²This is evident in the examples above. In observing the CMB, we see the breaking of boosts in the presence of a dipole, but we can safely neglect the breaking of time translations because observations are conducted over tens of years while the CMB changes in time over 10^5 years. Conversely, the redshift of photons from distant sources is mostly caused by the breaking of time translations, while the effect of peculiar motion, which is evident in redshift space distortions, is much smaller.



Figure 6.1: The figure shows CMB dipole at the level of 3 mK align with the $\pm \beta_{\parallel}$ direction. The two perpendicular directions $\pm \beta_{\times}$ and $\pm \beta_{\perp}$ are also shown for reference. This observation highlights the existence of a preferred frame in our universe and hence implies the spontaneous breaking of boost invariance.

superfluid or *P*-of-*X* theory, namely a shift-symmetric scalar field whose evolution is assumed to be approximately linear in time³. In this case, while time-translations, which are generated by $T^{0\mu}$, and shifts, which are generated by j^{μ} , are separately broken spontaneously, an (approximate) unbroken diagonal linear combination t^{μ} exists

$$t^{\mu} = T^{0\mu} + j^{\mu} \quad \Rightarrow \quad \nabla_{\mu} t^{\mu} = 0.$$
(6.1)

In inflationary models this unbroken diagonal symmetry is eventually responsible for the (approximate) scale invariance of primordial perturbations that we have observed in the data. One might ask whether a similar mechanism can be developed to suppress or eliminate the breaking of boosts. As pointed out recently in [168] (see also [169]), this is problematic because one would need to invoke a higher-spin symmetry, which in flat space is forbidden by the Coleman-Mandula theorem [170]. Indeed, it was proven in [168] that if one insists on having unbroken boost invariance for cosmological correlators in single-clock inflation, all interactions are forbidden and the theory must be free. Thus, the breaking of boosts cannot be eliminated and in principle it could always affect the interactions.

The discussion above highlights the importance for cosmology of time-translation invariant

³In general, the existence of a shift symmetry is not sufficient to ensure time-translation invariance. Rather, its general consequences are new cosmological soft theorems [98] and recursive relations for the time-dependence of the low-energy coupling constants [121]. It is only when one further assumes a linear evolution for the shift-symmetric scalar that a diagonal symmetry emerges, which plays the role of time-translation invariance, a general mechanism that goes under the name of spontaneous symmetry probing [166]. See [167] for a recent discussion on using a constant shift symmetry, and other symmetries, to realise a diagonal form of unbroken translations in the presence of additional non-linearly realised symmetries.

theories that (spontaneously) break boosts. In this work we study some of these theories in the context of scattering amplitudes. It will turn out that the application of our results to cosmology shows an unexpected and very interesting twist. We will discuss this in Section 6.5.

Cosmological correlators The calculation of primordial initial conditions from models of the early universe provides a major motivation for the study of boost-breaking amplitudes. The key observation is that the correlators of n fields of momenta \vec{k}_a with a = 1, ..., n in an expanding universe encode the information of n-particle scattering amplitudes in Minkowski in the residue of the highest k_T pole (see [171, 172]), where $k_T = \sum |\vec{k}_a|$ is sometimes called the "total energy". Schematically, the relation takes the form⁴

$$\lim_{k_T \to 0} \langle \prod_{a=1}^n \phi_a \rangle' \sim \frac{\operatorname{Re}A_n}{\left(\prod_{a=1}^n k_a\right)^2 k_T^p} + \dots$$
(6.2)

where the dots represent subleading terms in $k_T \rightarrow 0$, ϕ_a are fields (not necessary scalars), A_n is the flat space amplitude for the scattering of the particles created by the ϕ_a 's, and a prime denotes that we are dropping the momentum conserving delta function. The value of the positive exponent p depends on the interactions included in the theory, with larger p's corresponding to the inclusion of operators of higher and higher dimension [173]. This relation gives us a handle to leverage our knowledge of amplitudes to better understand cosmological correlators.

The idea to constrain cosmological correlators from symmetries has been pursued from various angles over the years. In [171] it was shown that the graviton bispectrum is completely fixed non-perturbatively by the isometries of de Sitter to be a linear combination of only two shapes, one corresponding to the Einstein-Hilbert term and the other to a higher-derivative term. In [174], de Sitter isometries were used to fixed the bispectrum of a spectator scalar. In [175], it was shown how an approximate version of de Sitter isometries constrains the leading-order scalar-scalar-tensor bispectrum. In [176, 115, 116] the study was extended to the scalar bispectrum and trispectrum. In [90], it was shown that the ζ bispectrum in the de Sitter-invariant limit of single-field inflation is fully fixed by approximate de Sitter isometries. More recently, in [177–181] an ambitious program has been proposed to systematically use not only symmetries but also general principles such as unitarity

⁴There are many exceptions to this result. For example, when the amplitude vanishes, this relation should be modified since the leading pole disappears. This is what happens in the DBI theory, due to the increased symmetry in the flat space-limit, as recently noticed in [167].

and locality to "bootstrap" correlators, in analogy with the on-shell methods for amplitudes. In the current incarnation of this *cosmological bootstrap*, the isometries of de Sitter spacetime still play an essential role, analogously to the role Poincaré invariance plays for amplitudes. On the one hand, it is clear from the above literature that de Sitter isometries are so constraining that many correlators are uniquely specified by them. On the other hand, we know that most observationally interesting correlators, such as for example equilateral and orthogonal non-Gaussianity, are not de Sitter invariant, and so cannot be studied directly with these methods. More generally, in [168] it was proven that in single-field inflation, the only theory whose ζ correlators are invariant under de Sitter isometries is the free theory. It is therefore very important to extend the cosmological bootstrap to less symmetric cases. In particular, it is the invariance under de Sitter boosts that should be relaxed, as this has not been observed in the data and indeed is not present in many models, for example those with a reduced speed of sound, $c_s < 1$. Much insight can already be gained by perturbative calculations [182–187].

The amplitudes that emerge on the total energy pole in (6.2) when de Sitter boosts are broken are not Lorentz invariant, rather they break Lorentz boosts. So one crucial step to extend the cosmological bootstrap to correlators with broken de Sitter boosts is to understand boost-breaking amplitudes. This is one of our primary motivations for this work.

Gravitational waves The recent detection of gravitational waves has ushered a new era in astronomy. But the detection of this 100 year old prediction of general relativity (GR) has implications well beyond the study of binary compact objects. It provides strong constraints on modified gravity (see e.g. [188–191]) and on the properties of the graviton. In particular, the concurrent observation of GW170817 [192] and the gamma-ray burst GRB170817A [193] has put extremely strong constraints on the difference Δv between the speed of gravity and the speed of light [163]

$$-3 \times 10^{-15} < \Delta v/c < 7 \times 10^{-16} \,. \tag{6.3}$$

More general Lorentz-breaking modifications of the graviton dispersion relation were classified and severely constrained in [194] using gravitational Cerenkov radiation by cosmic rays, and the constraints are even stronger when the GW170817 and GRB170817A data is included [163]. In particular, Lorentz-breaking deviations from a relativistic dispersion relation $E^2 = c^2 \mathbf{p}^2$ have to be smaller than a part in 10^{-13} , and some specific modifications must be as small as a part in 10^{-45} . The mass of the graviton is also strongly constrained by a variety of measurements. Largely model-independent bounds on the graviton mass m_g can be as strong as $m_g < 10^{-22}$ eV from observations such as Yukawa-like corrections to Newton's law [195] or gravitational waves from binary mergers [196] (see [197] for a recent summary and more details). More model-dependent bounds can be as strong as $m_g < 10^{-32}$ eV from observations of gravitational lensing [198] or of the earth-moon precession [199]. All of these bounds strengthen our confidence that GR provides a good description of free gravitons.⁵

It is then natural to ask: *what gravitational interactions are compatible with the observation that the graviton is a relativistic, massless spin two particle*? Any theoretical guidance in answering this question is of particular relevance also because it is much harder to directly probe the non-linear dynamics of gravitons, due to the weakness of gravity. It has been known for half a century that Lorentz invariance forces the self-interaction of a massless spin-2 particle, as well as the interactions with any other particle, to be universal in the infra-red around Minkowski spacetime and to correspond to the interactions of GR [201, 202]. More generally, from a purely on-shell perspective, there are only three possible cubic (analytically continued) amplitudes for three gravitons, which reduce to two if one assumes parity [29]. These are the interactions of GR, coming from the Ricci scalar *R*, and higher derivative interactions from the (dimension 9) Riemann cubed terms, which are highly suppressed at low energies. Self-interactions with broken Lorentz boosts have received less attention. In [203], it is argued that the explicit breaking of Lorentz symmetry is inconsistent with dynamical gravity, while this obstruction may be absent if the breaking is spontaneous. In [204], the authors show that assuming only spatial covariance, the leading order couplings of the graviton must display Lorentz invariance, which from this perspective appears as an emergent symmetry.

In this work we will take a complementary approach. We will only discuss physical on-shell (massless) particles, thus avoiding any mention of gauge symmetries such as general covariance. General principles such as unitarity and locality will then enforce Lorentz invariance and agreement with GR, within the assumptions that we make about the form of our four-particle amplitudes. Our results are summarized below in Section 6.1.2.

⁵Finally, from a more theoretical perspective, [200] argues that the special relativistic energy-momentum relation is a consequence of locality and of the existence of massless gravitons mediating long range forces.

6.1.2 Summary of the main results

The main body of this chapter consists of a detailed derivation of our results. We attempted to make our derivation pedagogical and the presentation self-contained, so that this chapter can be approached without much familiarity with on-shell methods and the spinor helicity formalism. While many of our derivations are technical in nature, our final results can be stated in simple terms. For the reader who is not interested in the details, we therefore outline our main findings here. All the statements below are valid under the following assumptions:

- The spacetime is Minkowski.
- All particles are relativistic, massless and luminal, i.e. they all propagate at *the same speed*, which we set to one and call the "speed of light", even when no photons are present in the spectrum.
- All interactions respect spacetime translations and rotations, but we allow for interactions that are not invariant under Lorentz boosts. Whether Lorentz boosts are non-linearly realized or explicitly broken plays no role in our analysis.
- While our results for three-particle amplitudes are non-perturbative in nature, our factorization constraints on the four-particle amplitudes ignore loop contributions.
- The helicity scaling of four-particle amplitudes is fixed in terms of "angle" and "square" spinor helicity brackets only. This assumption amounts to assuming that the underlying Lagrangian is a function of Lorentz covariant fields with the breaking of Lorentz boost induced by time derivatives, which can appear at will. This assumption means that our results do not apply to theories that are written in terms of *SO*(3) covariant fields such as the Framid and the Solid of [162]. We will explain in Section 6.4 why the amplitudes of these theories are not captured by our ansatz.

From these assumptions and demanding unitarity and locality through the consistent factorizations of four-particle amplitudes, we are able to show that the set of consistent interactions is severely restricted. In more detail:

• We derive all possible boost-breaking cubic amplitudes for relativistic massless particles of any spin. Unlike in the Lorentz-invariant case, there are always infinitely many possibilities, which are characterized by a generic function of the particles' energies (see (6.48)). This result is completely non-perturbative.

- If interactions with a massless spin-2 particle are allowed, three-particle amplitudes must be Lorentz invariant, even those that do not involve a graviton (see Section 6.4.3). For example, amplitudes corresponding to boost-breaking cubic scalar interactions such as $\dot{\phi}^3$, $\dot{\phi}(\partial\phi)^2$ and all other higher-derivative ones are forbidden. We conjecture this to be true for all other higher-particle amplitudes. This is strong evidence that Lorentz invariance follows from having consistent interactions involving a massless spin-2 particle, at least as long as the Lagrangian is written in terms of covariant fields as we stated in the above assumptions.
- The cubic graviton amplitudes must be those of GR at low energies (corresponding to dimension-5 operators). As for the Lorentz-invariant case, the only other graviton amplitudes correspond to the two possible Riemann³ couplings (dimension-9 operators).
- Particles with spin S > 1 cannot have an electric charge (see Section 6.4.2). Particles with spin S > 2 cannot have cubic self-interactions of dimensionality lower than 3S. They also cannot interact gravitationally via the GR vertex (see Section 6.4.1 and 6.4.3). Lower spin particles (S < 2) can indeed be minimally coupled to the graviton and these couplings are fixed by the coupling of the GR vertex. This is the on-shell manifestation of the equivalence principle.
- Unlike for the Lorentz-invariant case, cubic self-interactions of a single massless spin-1 particle do exist (dimension-6 operators) when boosts are broken (see Section 6.4.1). All lower dimension operators are forbidden.
- We find large classes of self-consistent, boost-breaking interactions among scalars, photons and spin-1/2 fermions, already at leading order in spatial derivatives. In other words, QED, scalar QED and scalar theories allow for the breaking of boosts at the cubic level (see Section 6.4.2).
- We point out that the four-particle test for massless particles is *highly IR sensitive* (see Section 6.5). As a consequence, the results that follow from it cannot be straightforwardly applied to cosmology, where the Hubble parameter that characterizes the curvature of spacetime constitutes an IR modification of Minkowski spacetime. Conversely, all those results that are exclusively based on symmetries, such as for example the form of the three-particle amplitude (see Section 6.3) are robust and do apply to curved spacetime as well.

Notation and conventions Since we will be dealing with boost-breaking theories, for dimensional analysis we will have to separate units of length from units of time. Working with $\hbar = 1 = c$, we will indicate by "dim..." the scaling of an object with *spatial* momentum, which has units of inverse length, *excluding the dimension of all coupling constants*. For Lorentz-invariant theories this gives to

the standard energy/mass dimension, as e.g. in [28]. For example,

$$\dim(ii) = 0, \qquad \qquad \dim[ij] = \dim\langle ij\rangle = 1. \tag{6.4}$$

We will work with the mostly minus metric signature $\eta_{\mu\nu} = \text{diag}(+1, -1, -1, -1)$ and follow [205] for spinor conventions. We use the beginning of the Greek alphabet for SU(2) indices $(\alpha, \beta, \gamma, ...)$, and the middle of the alphabet for SO(1,3) indices $(\mu, \nu, \rho, \sigma, ...)$. Our basis for the Pauli matrices $\sigma^{\mu}_{\alpha\dot{\alpha}}$ and $(\bar{\sigma}^{\mu})^{\dot{\alpha}\alpha}$ is

$$(\sigma^{0})_{\alpha\dot{\alpha}} = (\bar{\sigma}^{0})^{\alpha\dot{\alpha}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad (\sigma^{1})_{\alpha\dot{\alpha}} = -(\bar{\sigma}^{1})^{\alpha\dot{\alpha}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad (6.5)$$

$$(\sigma^2)_{\alpha\dot{\alpha}} = -(\bar{\sigma}^2)^{\alpha\dot{\alpha}} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad (\sigma^3)_{\alpha\dot{\alpha}} = -(\bar{\sigma}^3)^{\alpha\dot{\alpha}} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad (6.6)$$

and amongst the many useful identities these matrices satisfy

$$\sigma^{\mu}_{\alpha\dot{\alpha}}\bar{\sigma}^{\dot{\beta}\beta}_{\mu} = 2\delta_{\alpha}{}^{\beta}\delta^{\dot{\beta}}{}_{\dot{\alpha}}\,,\tag{6.7}$$

$$\sigma^{\mu}_{\alpha\dot{\alpha}}(\sigma_{\mu})_{\beta\dot{\beta}} = 2\epsilon_{\alpha\beta}\epsilon_{\dot{\alpha}\dot{\beta}}\,,\tag{6.8}$$

$$(\bar{\sigma}^{\mu})^{\dot{\alpha}\alpha}\bar{\sigma}^{\dot{\beta}\beta}_{\mu} = 2\epsilon^{\alpha\beta}\epsilon^{\dot{\alpha}\dot{\beta}}, \qquad (6.9)$$

where the components of the epsilon and delta tensors are

$$\epsilon^{12} = -\epsilon^{21} = \epsilon_{21} = -\epsilon_{12} = 1, \qquad \qquad \delta_{\alpha}{}^{\beta} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$
 (6.10)

We use these epsilon tensors to raise and lower the dotted and undotted SU(2) indices as

$$\psi_{\alpha} = \epsilon_{\alpha\beta}\psi^{\beta}, \qquad \psi^{\alpha} = \epsilon^{\alpha\beta}\psi_{\beta}, \qquad \bar{\psi}_{\dot{\alpha}} = \epsilon_{\dot{\alpha}\dot{\beta}}\bar{\psi}^{\bar{\beta}}, \qquad \bar{\psi}^{\dot{\alpha}} = \epsilon^{\dot{\alpha}\dot{\beta}}\bar{\psi}_{\dot{\beta}}. \tag{6.11}$$

Note added During the completion of this work a paper appeared [206] that argues that the consistent description of a massless spin-2 particle requires certain tree-exchange diagram to be Lorentz invariant. One of our main results in this work is in complete agreement with this finding, while other results for gravitons are new. In a similar vein, [207] recovers the central tenets of electromagnetism, such

as charge conservation, without imposing boost invariance. Our point of view and methodology are complementary to that in [206], [207] since we only use on-shell methods and make no use of the field theory apparatus.

6.2 On-shell methods: symmetries and bootstrap techniques

The aim of the S-matrix bootstrap program is to construct, directly at the level of the S-matrix, consistent scattering amplitudes exhibiting a given set of (linearly realised) symmetries. This on-shell technique bypasses the usual Lagrangian formalism of effective field theories, thereby avoiding redundancies such as field redefinitions and gauge transformations. In this section we introduce the basic principles of this bootstrap program.

6.2.1 Symmetries and on-shell conditions for free particles

We begin by discussing the symmetries we are assuming so that we can clearly compare and contrast our results with those in the literature [29, 31, 28, 208–210]. Up to now, on-shell methods and the four-particle test of [29] have been applied to theories for which the vacuum is assumed to be invariant under the full Poincaré group ISO(1,3), consisting of spacetime translations, spatial rotations and Lorentz boosts. In this work we relax the assumption that Lorentz boosts leave the vacuum unchanged, while assuming that spacetime translations and spatial rotations remain good linearly realised symmetries. We will be agnostic about whether boosts are explicitly broken or spontaneously broken and non-linearly realized. In four spacetime dimensions our symmetry group is therefore $\mathbb{R}^4 \rtimes SO(3)$. Throughout the chapter, we will use the following terminology:

Boost-invariant theories: unbroken
$$ISO(1,3)$$
 (6.12)

Boost-breaking theories: unbroken
$$\mathbb{R}^4 \rtimes SO(3)$$
. (6.13)

In the bootstrap program one has to provide the on-shell data which includes the *on-shell conditions* relating the energy and spatial momentum of each free particle. In boost-invariant theories massless particles satisfy the usual on-shell condition $E^2 - \mathbf{p}^2 = 0$, while in boost-breaking theories many other on-shell conditions are allowed due the reduced symmetry. Below we classify these possibilities:

- *Relativistic*: each free particle satisfies $E^2 c_s^2 \mathbf{p}^2 = 0$ with the speed of sound c_s being *the same* for each particle. Without loss of generality, in this case we can choose to work in units such that $c_s = c = 1$ and we will do this in the rest of the chapter.
- *Linear*: each free particle satisfies $E^2 c_s^2 \mathbf{p}^2 = 0$, where at least two particles have a different c_s .
- General: the on-shell condition for each particle is S(E, p) = 0 and is not captured by the two cases above.

In this chapter we consider the relativistic case where each particle has a Lorentz invariant propagator and leave generalisations to other on-shell conditions for future work. So, we focus on theories where all boosts are broken at the level of the interactions only which will lead us to a natural ansatz for four-particle amplitudes. We therefore combine the energy and spatial momentum into the usual 4-vector p_{μ} satisfying $p^{\mu}p_{\mu} = 0$ for each particle.

6.2.2 Little group scaling and the spinor helicity formalism

Let us now emphasise that the usual classification of massless particles in terms of helicity remains valid for boost-breaking theories. In this subsection we also present the spinor helicity formalism, which for boost-invariant theories has been reviewed in many cases e.g. [30, 28, 31, 211, 210], and for boost-breaking theories was introduced in [171] (see also Appendix C of [181]).

Spacetime translation symmetry alone entails that there exists a basis of one particle states $|\mathbf{p}, E\rangle$, which are the eigenstates of the momentum and energy operators:

$$\hat{p}_i |\mathbf{p}, E\rangle = p_i |\mathbf{p}, E\rangle, \qquad \hat{E} |\mathbf{p}, E\rangle = E |\mathbf{p}, E\rangle.$$
(6.14)

States with the same \mathbf{p} and E may be degenerate and additional quantum numbers are collectively indicated by an index σ i.e. $|\mathbf{p}, E; \sigma\rangle$. An important subgroup of the full Lorentz group is the *little group* which is the group of transformations that leave the 4-momentum p_{μ} invariant. Such transformations map

$$|\mathbf{p}, E; \sigma\rangle \mapsto D_{\sigma}^{\sigma'} |\mathbf{p}, E; \sigma'\rangle.$$
 (6.15)

Single particle states can then be further classified according to their eigenvalues under the little

group. In both boost-invariant and boost-breaking theories, this is the projective $SO(2)^6$, and the states $|\mathbf{p}, E\rangle$ carry a label corresponding to helicity $h = 0, \pm \frac{1}{2}, \pm 1, \ldots$ Clearly the relevant symmetry here is spatial rotations, rather than Lorentz boosts. The helicity of a particle is the same in all frames related by a rotation and changes sign under a spatial reflection. For that reason, we may consider the allowed helicity states for a massless particle of spin S > 0 to be +S and -S.

Throughout this work we will make use of spinor helicity formalism as a powerful tool to present amplitudes in a compact form. This formalism, introduced below, provides a compact way of expressing amplitudes and its simplicity is beautifully captured by the Parke-Taylor formula for gluon scattering [213]. Here we extend these methods along the lines of [171] for application in boost-breaking theories.

We start by using the Pauli matrices (we follow the conventions of [205]) to map the momentum 4-vector p_{μ} into a 2 × 2 matrix⁷

$$p_{\alpha\dot{\alpha}} = \sigma^{\mu}_{\alpha\dot{\alpha}} p_{\mu} = \begin{pmatrix} p_0 + p_3 & p_1 - ip_2 \\ p_1 + ip_2 & p_0 - p_3 \end{pmatrix},$$
(6.16)

where $\sigma^{\mu} = (1, \sigma^{i})$. The dotted and undotted indices transform in the fundamental and antifundamental representation of $SL(2, \mathbb{C})^{8}$ respectively, such that $p_{\alpha\dot{\alpha}}$ transforms in the (1/2, 1/2)representation. The dotted and undotted indices run over two values, e.g. $\alpha = 1, 2$, and in a boostinvariant theory dotted and undotted indices are contracted with the epsilon tensors $\epsilon^{\dot{\alpha}\dot{\beta}}, \epsilon^{\alpha\beta}$. Using $p_{\alpha\dot{\alpha}}$ alone, the only Lorentz invariant quantity we can construct is $p^{\alpha\dot{\alpha}}p_{\alpha\dot{\alpha}} = 2 \det(p) = 2p^{\mu}p_{\mu} = 0$. It follows that $p_{\alpha\dot{\alpha}}$ is at most rank one thereby allowing us to write

$$p_{\alpha\dot{\alpha}} = \lambda_{\alpha}\tilde{\lambda}_{\dot{\alpha}},\tag{6.17}$$

where λ and $\tilde{\lambda}$ are two-component spinors. Note that these objects *are not* Grassmanian, rather they are complex numbers satisfying $\lambda_{\alpha} \tilde{\lambda}_{\dot{\alpha}} = \tilde{\lambda}_{\dot{\alpha}} \lambda_{\alpha}$. We also note that these spinors are not unique and

⁶In the boost-invariant case, the little group for massless particles is ISO(2), but we recover SO(2) if we make the reasonable assumption that the fields transform trivially under the noncompact subgroup representing the translations in ISO(2). (See [212], Chapter 2 for more details.) Once boosts are broken, the little group becomes SO(2) straight away.

⁷Since $\sigma^{\mu}_{\alpha\dot{\alpha}}\bar{\sigma}^{\dot{\beta}\beta}_{\mu} = 2\delta_{\alpha}{}^{\beta}\delta^{\dot{\beta}}{}_{\dot{\alpha}}$ we have $p_{\mu} = \frac{1}{2}\bar{\sigma}^{\dot{\alpha}\alpha}_{\mu}p_{\alpha\dot{\alpha}}$

⁸In 4 dimensions, the group of proper Lorentz transformations is $SO(1,3) \simeq SL(2,\mathbb{C})/\mathbb{Z}_2$. Thus, projective representations of the Lorentz group can be identified with representations of $SL(2,\mathbb{C})$.

are only defined up to a little group, or helicity, transformation. Indeed the transformation

$$(\lambda_{\alpha}, \tilde{\lambda}_{\dot{\alpha}}) \mapsto (t^{-1}\lambda_{\alpha}, t\tilde{\lambda}_{\dot{\alpha}}),$$
 (6.18)

where t is a nonzero complex number, leaves $p_{\alpha\dot{\alpha}}$ invariant. For physical processes, the external momenta are always real and therefore the spinors can be chosen to satisfy the reality condition $\tilde{\lambda}_{\dot{\alpha}} = \pm (\lambda^*)_{\dot{\alpha}}$ and we can restrict the transformation parameter t to a phase. However, to study the analytic structure of the S-matrix we must keep the momenta complex, and therefore the spinors are in general independent.

What scalar quantities can we construct from these spinors? In boost-invariant theories we have the following two inner products

$$\langle ij \rangle = \epsilon^{\alpha\beta} \lambda^{(i)}_{\alpha} \lambda^{(j)}_{\beta}, \qquad [ij] = \epsilon^{\dot{\alpha}\dot{\beta}} \tilde{\lambda}^{(i)}_{\dot{\alpha}} \tilde{\lambda}^{(j)}_{\dot{\beta}}, \qquad (6.19)$$

defined for two particles *i* and *j*. We refer to these products as angle and square brackets, respectively. Since the epsilon tensors are anti-symmetric and the spinors are not Grassmanian, these brackets are anti-symmetric i.e. $\langle ij \rangle = -\langle ji \rangle$ and [ij] = -[ji], which of course implies $\langle ii \rangle = [ii] = 0$. From these brackets we can construct the familiar Mandelstam variables for four-particle scattering amplitudes. Taking all particles as *incoming*, we have

$$s = (p_1 + p_2)^2 = (p_3 + p_4)^2 = \langle 12 \rangle [12] = \langle 34 \rangle [34], \tag{6.20}$$

$$t = (p_1 + p_3)^2 = (p_2 + p_4)^2 = \langle 13 \rangle [13] = \langle 24 \rangle [24], \tag{6.21}$$

$$u = (p_1 + p_4)^2 = (p_2 + p_3)^2 = \langle 14 \rangle [14] = \langle 23 \rangle [23].$$
(6.22)

For our interests, however, we have a reduced set of symmetries and therefore additional scalar quantities are allowed. Indeed, in boost-breaking theories we can mix the dotted and undotted indices by contracting the spinors with $(\bar{\sigma}^0)^{\alpha \dot{\alpha}}$. We therefore have an additional inner product which we denote as

$$(ij) = (\bar{\sigma}^0)^{\alpha \dot{\alpha}} \lambda_{\alpha}^{(i)} \tilde{\lambda}_{\dot{\alpha}}^{(j)}, \qquad (6.23)$$

and refer to as round brackets. As will be explained in section 6.3, for three-particle kinematics only the diagonal components of this new bracket i.e. (ii) are independent objects, while for four-particle

kinematics one of the off-diagonal brackets is independent. For the relativistic on-shell condition, the 0-component of the momentum 4-vector for each particle is the energy of the particle, which we denote by E. The diagonal round brackets pick out precisely this component: $(ii) = 2E_i$.

For spinning particles there is a key piece of on-shell data which we haven't yet discussed: the polarisation tensors. These form non-trivial representations of the little group and therefore encode the helicity of the particle in question. For a spin-S particle we write the rank-S polarisation tensor as a product of S polarisation vectors which in the spinor helicity variables take the form

$$e_{\alpha\dot{\alpha}}^{+} = \frac{\eta_{\alpha}\tilde{\lambda}_{\dot{\alpha}}}{\langle\eta\lambda\rangle}, \qquad e_{\alpha\dot{\alpha}}^{-} = \frac{\lambda_{\alpha}\tilde{\eta}_{\dot{\alpha}}}{[\tilde{\eta}\tilde{\lambda}]},$$
 (6.24)

for +1 and -1 helicity respectively. The form of the polarisation vectors follows from the fact that they should be orthogonal to the corresponding momentum. Indeed,

$$p^{\alpha\dot{\alpha}}e^{+}_{\alpha\dot{\alpha}} = [\tilde{\lambda}\tilde{\lambda}] = 0 = p^{\alpha\dot{\alpha}}e^{-}_{\alpha\dot{\alpha}} = \langle\lambda\lambda\rangle.$$
(6.25)

For each particle, the reference spinors η and $\tilde{\eta}$ are linearly independent from λ and $\tilde{\lambda}$ respectively, but are otherwise arbitrary. Different choices for the reference spinors can alter the polarisation vectors, but only by a gauge transformation, which leaves the amplitude unchanged. We have seen above that for boost-breaking theories we can mix dotted and undotted indices using $(\bar{\sigma}^0)^{\alpha \dot{\alpha}}$. This allows us to make choices for the reference spinors for which the zero-component of the polarisation vectors vanishes [171]. In a gauge invariant theory this choice is as good as any other, but if the underlying Lagrangian is constructed out of SO(3) covariant fields only, then this choice is forced upon us since the fields do not have time components. In this chapter we are assuming that the fields are Lorentz covariant and so we are not restricted to this choice for the reference spinors.

For an *n*-particle scattering amplitude, we have *n* distinct momenta and therefore *n* distinct helicity transformation generators $\hat{\mathcal{H}}_i$, corresponding to rotations of a particle around its momentum vector. If we treat all particles as incoming and represent the initial state as $|p;h\rangle = |p_1;h_1\rangle \otimes \ldots \otimes |p_n;h_n\rangle$, then the *i*th helicity generator is represented on the space of initial states as $\hat{H}_i = id \otimes id \otimes \ldots \otimes \hat{\mathcal{H}}_i \otimes \ldots id$, and we have $\hat{H}_i |p;h\rangle = h_i |p;h\rangle$. The amplitude itself must transform under \hat{H}_i in the same way the initial state does, i.e.

$$\ddot{H}_i \mathcal{A}_n(p;h) = h_i \mathcal{A}_n(p;h), \tag{6.26}$$

which in turn implies that under $\{\lambda^{(i)}, \tilde{\lambda}^{(i)}\} \mapsto \{t_i^{-1}\lambda^{(i)}, t_i\tilde{\lambda}^{(i)}\}$ the amplitude transforms as

$$\mathcal{A}_n(\{\lambda^{(i)}, \tilde{\lambda}^{(i)}; h_i\}) \mapsto \mathcal{A}_n(\{t_i^{-1}\lambda^{(i)}, t_i\tilde{\lambda}^{(i)}; h_i\}) = \prod t_i^{2h_i} \mathcal{A}_n(\{\lambda^{(i)}, \tilde{\lambda}^{(i)}; h_i\}).$$
(6.27)

This little group scaling of the amplitude can very powerfully constrain the allowed structure of the amplitude, see e.g. [30, 28]. For boost-invariant theories it completely fixes the non-perturbative form of the three-particle amplitudes, while in boost-breaking theories it completely fixes the amplitude up to an arbitrary function of the energies of the three particles, as we shall see in section 6.3.

6.2.3 Unitarity, analyticity and the four-particle test

Analytic properties of the S-matrix have been extensively studied in boost-invariant theories. Analyticity, the singularity structure and crossing symmetry of amplitudes are very important aspects of the S-matrix bootstrap. In this chapter we rely on the possibility of extending these essential S-matrix properties to a more general setting and so here we outline why these properties do not require the theory to be invariant under the full Poincaré group.

Let us start with *analyticity* of the S-matrix. By *analyticity*, we mean that once the S-matrix is stripped of the momentum conserving delta function, the remaining factor, when continued into the complex space, is an analytic function of the kinematic variables, except for a finite number of singularities and (possibly) branch cuts. In this chapter we will be considering tree level exchange for four-particle amplitudes and so will not encounter any branch cuts. Our three-particle amplitudes are however non-perturbative and are almost completely fixed by symmetry. An argument for analyticity (away from singularities, which are going to be discussed shortly), which does not rely on the invariance of physics under boosts was presented in [214] and so we will take it for granted that scattering amplitudes are (locally) analytic functions of the kinematic variables discussed above. Our amplitudes will also be crossing symmetric. Crossing symmetry [215] is a symmetry of the S-matrix under the following transformation: for a given particle of momentum p_{μ} in the final state, consider instead its own antiparticle with momentum $-p_{\mu}$ in the initial state. The S-matrix, understood as an analytic function of the complex energies and momenta, must not change under such a transformation. Thus, without loss of generality, we will consider all particles participating in a given process

as incoming (an incoming particle with negative energy is to be interpreted as an outgoing antiparticle).

The most powerful constraint on effective theories and their interactions will come from the singularity structure of the S-matrix. The factorisation theorem, following from locality and unitarity, states that

Theorem 6.2.1. (Factorization Theorem) Singularities of codimension 1 in 4-particle amplitudes may appear at vanishing energies ($E_i = 0$) or else are at most simple poles in the momenta. Each singularity of the latter type is in one-to-one correspondence with an exchange diagram (Fig. 6.2), in the limit when the exchanged particle I goes **on-shell**. The residue of each pole factorises into a product of three-particle amplitudes:

$$\lim_{s=0} (s\mathcal{A}_4) = \mathcal{A}_3(1, 2, -I) \times \mathcal{A}_3(3, 4, I)$$
(6.28)

where s is the propagator of the intermediate particle, and $s \rightarrow 0$ corresponds to the intermediate particle going on-shell.

While the above result is almost trivial in perturbation theory and its intuitive physical meaning is not hard to grasp, it can also be demonstrated with mathematical rigour. Starting from the Weak Causality Postulate (*If initial state consists of wave packets colliding at time* t_1 *and the final state consists of wave packets colliding at time* t_2 , and $t_1 - t_2$ is much larger than the typical spatial width of the wave packets, then the scattering amplitude should be small⁹) and by considering wave packets sharply localized in momentum space, Peres [86] has shown that the existence of an interacting particle of mass $M \neq 0$ leads to a contribution $A_1A_2/(E_I^2 - p_I^2 - M^2 + i\epsilon)$, which is to be identified with processes that involve two collisions of the wave packets (with amplitudes A_1 and A_2 respectively) separated by a macroscopic time interval. Conversely, if the amplitude in the vicinity of a pole takes the form $A_1A_2/(E_I^2 - p_I^2 - M^2 + i\epsilon) + regular terms$, then the first term represents the amplitude for scattering of wave packets through two or more subsequent collisions, which will be non-negligible provided that the 4-vector connecting the collisions is approximately parallel to the 4-momentum (E_I, p_I) . This is then interpreted as a propagating particle of mass M. The argument of [86] does not rely on invariance under boosts¹⁰ and can be easily generalized to on-shell conditions of the form

⁹More rigorously [216]: scattering amplitude should decay faster than any power of $\Delta t = t_1 - t_2$ as $\Delta t \to \infty$.

¹⁰Although the author does fix Lorentz frame to the center of mass frame, this convenient trick serves illustrative and pedagogical purposes only and can be eliminated altogether.

 $E^2 - \omega^2(p) = 0$, provided there is a mass gap. Other derivations of factorisation, which do not rely on invariance under Lorentz boosts and emphasise the important role of unitarity, can be found in [217] and Section 10.2 of [212]. See also [211] for further discussions¹¹.

None of the above proofs can on its own exclude the possibility that the poles corresponding to an intermediate particle going on-shell have order higher than 1. For this we need an additional argument: consider an exchange channel which, according to the Factorization Theorem, leads to a contribution $A_1A_2/(E_I^2 - p_I^2 - M^2 + i\epsilon) + regular terms$ to the amplitude. We want to show that the first term contains only first order pole in $(E_I^2 - p_I^2 - M^2 + i\epsilon)$. The essential observation is that if it contained a higher order pole, then one of the three-particle amplitudes, A_1 or A_2 , would have to be singular on some large subset of the s = 0 hypersurface. But A_1 and A_2 are three-particle amplitudes in a physical configuration (because the original amplitude could be taken to be in the physical configuration and the intermediate particle is on-shell), so they cannot be singular anywhere. This last statement is also confirmed by an explicit calculation starting from (6.48) - this quantity is finite in a generic configuration.

Let us now comment on S-matrix singularities at $E_i = 0$. These do not appear in Lorentz invariant theories, as they would clearly violate Lorentz invariance. More generally, such singularities *cannot* appear if the Lagrangian is local and can be written solely in terms of $X_{\mu_1\mu_2\dots}$, $\eta_{\mu\nu}$, $\epsilon_{\mu\nu\sigma\rho}$, ∂_{μ} and ∂_t (where $X_{\mu_1\mu_2\dots}$ collectively denotes Lorentz covariant fields). This is because the factor $1/E_i$ is generated only when some of the tensor field indices are spatial indices. In that case the associated polarization tensor $e^{\pm S}$ has a vanishing temporal component, so it must have a predetermined reference spinor as we eluded to above:

$$e_{\alpha_i\dot{\alpha}_i}^{+S}(\mathbf{k}) = \prod_{i=1}^{S} \frac{(\epsilon.\tilde{\lambda})_{\alpha_i}\tilde{\lambda}_{\dot{\alpha}_i}}{2k}, \quad e_{\alpha_i\dot{\alpha}_i}^{-S}(\mathbf{k}) = \prod_{i=1}^{S} \frac{\tilde{\lambda}_{\alpha_i}(\epsilon.\lambda)_{\dot{\alpha}_i}}{2k}.$$
(6.29)

We see that $e^{\pm s}(\mathbf{k})$ has a singularity at $E_k \equiv k = 0$, which might therefore appear also in the helicity amplitude by virtue of the relation

$$\mathcal{A}_4 = e^{h_1,\mu_1} e^{h_2,\mu_2} e^{h_3,\mu_3} e^{h_4,\mu_4} A_{4,\mu_1\mu_2\mu_3\mu_4}, \tag{6.30}$$

¹¹While, strictly speaking, there is no rigorous proof of the Factorization Theorem for massless particles, Feynman rules entail that tree-level diagrams in perturbation theory retain the stipulated property. Moreover, there is no known counterexample to the Factorization Theorem for massless particles. With this in mind, we will follow the many papers we have mentioned previously in the context of this theorem and assume that the theorem holds for massless theories.

where $A_{4,\mu_1\mu_2\mu_3\mu_4}$ is the *covariant amplitude*, which only has singularities when an exchanged particle goes on-shell. As we have explained above, we will be assuming that the Lagrangian is written in terms of Lorentz covariant fields so we don't expect such inverse powers of the energies to arise, but in many cases we see that allowing for these inverse powers does not affect our results.

Summarizing, four-particle scattering amplitudes in boost-invariant or boost-violating theories have the following singularity structure:

- The amplitude has only simple poles in the Mandelstam variables *s*, *t* and *u*, as well as poles in the individual energies *E_i*.
- On the s, t and u poles the amplitude factorises into a product of three-particle amplitudes.

These properties form the basis of the *four-particle test* [29]. This test requires the singularity structure of four-particle amplitudes to satisfy these two conditions, and for each pole in s, t or u to be interpreted as the propagation of a physical particle. Ensuring consistency in all three channels (s,t and u) is highly non-trivial and rules out almost all interactions for massless particles in boost-invariant theories, see [29, 28, 208–210, 218]¹². The reason why the test is non-trivial is that the residue on say the s-channel pole can contain inverse powers of t and u, as we shall see. In this chapter we will see that the four-particle test is also very constraining when we allow for boost-breaking interactions.

We will use the factorization theorem to constrain the constructible part of the tree-level fourparticle amplitudes. For this application, it will be sufficient that the tree-level propagator corresponds to a relativistic on-shell condition. If one made the stronger assumption that this is the case also for the full non-perturbative propagator, then one might be able to use our results to derive some constraints on non-perturbative four-particle amplitudes.

It should be noted that for massless particles, the $s \rightarrow 0$ limit of the amplitude makes perfect sense in Minkowski spacetime but this is not the case in curved spacetime. For example, in an FLRW spacetime this limit always takes us outside the validity of the flat-space approximation. Hence, the constraints imposed by Theorem 2.1 apply to flat spacetime but care is required when considering cosmological spacetimes. We discuss this in detail in Section 6.5.

¹²The test was originally formulated using BCFW momentum shifts [160]. Indeed, the authors of [29] demanded that two different BCFW shifts gave rise to the same answer for the four-particle amplitudes. As discussed in [28, 209], the test can actually be formulated as above where only complex factorisation is required.



Figure 6.2: Exchange diagram. Circles represent non-perturbative, exact 3-particle amplitudes.

6.3 Three-particle amplitudes

In this section we construct general on-shell three-particle amplitudes using the spinor helicity techniques outlined in Section 6.2. Then, as an example, we discuss the cases where all three particles are identical.

6.3.1 Non-perturbative structure for all spins

We assume that every particle is massless, has a definite helicity, and satisfies the relativistic on-shell condition $p^{\mu}p_{\mu} = 0$. We take all particles as incoming and therefore by momentum conservation we have

$$p_1^{\mu} + p_2^{\mu} + p_3^{\mu} = 0, \tag{6.31}$$

where 1, 2, 3 label the external particles. The amplitudes only depend on the observable quantities that can be defined on the asymptotic states and these in turn can be fully recovered from the spinors and helicities h_i . The amplitudes are then only a function of $\lambda^{(i)}$, $\tilde{\lambda}^{(i)}$ and h_i . Indeed, written in terms of the spinor helicity variables, (6.31) becomes

$$\lambda_{\alpha}^{(1)}\tilde{\lambda}_{\dot{\alpha}}^{(1)} + \lambda_{\alpha}^{(2)}\tilde{\lambda}_{\dot{\alpha}}^{(2)} + \lambda_{\alpha}^{(3)}\tilde{\lambda}_{\dot{\alpha}}^{(3)} = 0.$$
(6.32)

The simple form of this equation is the main reason why computations are considerably simpler when dealing with relativistic on-shell conditions. For any other on-shell condition, such as linear or general, (6.32) does not hold and the following analysis needs to be modified.

As explained in Section 6.2, the quantities from which we should construct amplitudes are the

three inner products: $\langle ij \rangle$, [ij], (ij). However, momentum conservation and the fact that each particle is on-shell ensures that any contraction of two distinct momenta is zero. Indeed,

$$(p_1 + p_2)^2 = 2p_1 \cdot p_2 = p_3^2 = 0, \tag{6.33}$$

$$(p_2 + p_3)^2 = 2p_2 \cdot p_3 = p_1^2 = 0, \tag{6.34}$$

$$(p_1 + p_3)^2 = 2p_1 \cdot p_3 = p_2^2 = 0.$$
(6.35)

In the spinor helicity variables this translates into

$$\langle 12 \rangle [12] = \langle 13 \rangle [13] = \langle 23 \rangle [23] = 0.$$
 (6.36)

It follows that if $\langle 12 \rangle \neq 0$, we have [12] = 0 but by momentum conservation we have

$$\langle 12 \rangle [23] = -\langle 11 \rangle [13] - \langle 13 \rangle [33] = 0,$$
 (6.37)

and therefore [23] = 0 too. We also have $\langle 12 \rangle [13] = 0$ which requires [13] = 0. So having one angle bracket non-zero requires the three square brackets to vanish and vice versa. This tells us that three-particles amplitudes split up into holomorphic and anti-holomorphic configurations:

Holomorphic kinematics : [12] = [13] = [23] = 0, (6.38)

Anti-holomorphic kinematics :
$$\langle 12 \rangle = \langle 13 \rangle = \langle 23 \rangle = 0.$$
 (6.39)

Furthermore, the off-diagonal components of (ij) are degenerate with other brackets. Indeed for $i \neq j$ we can write

$$(ij)\langle jk\rangle = -(ii)\langle ik\rangle, \qquad (ij)[ik] = -(jj)[jk],$$
(6.40)

which allows us to solve for the off-diagonal components of (ij) for both the holomorphic and anti-holomorphic configurations. The brackets we can use to construct amplitudes are therefore $\langle ij \rangle$, [ij] for $i \neq j$ and (ii). Recalling that for the relativistic on-shell condition $(ii) = 2E_i$, we therefore write the amplitudes as a sum of holomorphic and anti-holomorphic pieces as

$$\mathcal{A}_3(\{\lambda^{(i)}, \tilde{\lambda}^{(i)}; h_i\}) = M_H(\langle ij \rangle, E_i; h_i) + M_{AH}([ij], E_i; h_i).$$
(6.41)

We are now in a position to constrain the amplitude by demanding it scales in the correct way under a helicity transformation $(\lambda^{(i)}, \tilde{\lambda}^{(i)}) \mapsto (t_i^{-1}\lambda^{(i)}, t_i\tilde{\lambda}^{(i)})$. As explained in Section 6.2, under this transformation the amplitude scales as

$$\mathcal{A}_{3}(\{t_{i}^{-1}\lambda^{(i)}, t_{i}\tilde{\lambda}^{(i)}; h_{i}\}) = \prod_{j=1}^{3} t_{j}^{2h_{j}} \mathcal{A}_{3}(\{\lambda^{(i)}, \tilde{\lambda}^{(i)}; h_{i}\}),$$
(6.42)

which constrains the dependence of the angle and square brackets. Note that the diagonal round brackets, or the energies, are invariant under this helicity transformation and so this symmetry does not constrain how they enter the amplitude. First consider M_H , which we can write as

$$M_H(\langle ij \rangle, E_i; h_i) = \langle 12 \rangle^{d_3} \langle 23 \rangle^{d_1} \langle 31 \rangle^{d_2} F^H_{h_1, h_2, h_3}(E_1, E_2, E_3).$$
(6.43)

Demanding the correct scaling of the amplitudes fixes

$$d_1 = h_1 - h_2 - h_3, \tag{6.44}$$

$$d_2 = h_2 - h_3 - h_1, \tag{6.45}$$

$$d_3 = h_3 - h_1 - h_2. ag{6.46}$$

Likewise, for M_{AH} we have

$$M_{AH}([ij], E_i; h_i) = [12]^{-d_3} [23]^{-d_1} [31]^{-d_2} F_{h_1, h_2, h_3}^{AH}(E_1, E_2, E_3).$$
(6.47)

Now consider the three cases h > 0, h < 0 and h = 0 where $h = h_1 + h_2 + h_3$ is the sum of the three helicities. If h > 0, we have $d_1 + d_2 + d_3 < 0$ meaning that the M_H part of the amplitude would become singular in the entire region defined by $\langle ij \rangle = 0$ (as long as $F^H \neq 0$ in that region). Three-particle amplitudes cannot have such singularities, so we require $F^H = 0$ whenever $\langle ij \rangle = 0$. But F^H is just a function of energies, not of the $\langle ij \rangle$ brackets, and it is impossible to generate these brackets from the energies alone. So in fact when h > 0 we require $F^H = M^H = 0$ everywhere. A similar analysis for h < 0 shows that we require $F^{AH} = M^{AH} = 0$ everywhere. For the third possibility, h = 0, both contributions to the amplitude can be non-zero.

We can also argue this by locality of the interactions. Let us define the mass dimension of an object A by dim A where we do not include the functions of energy in the mass dimension. Now since

each angle and square bracket has mass dimension 1, we have $\dim M_H = -h$ and $\dim M_{AH} = h$. The helicity part of the amplitudes cannot have a negative mass dimension as that would require inverse powers of Lorentzian derivatives in the interactions which cannot occur in a local theory. We therefore require $h \le 0$ for the holomorphic configuration and $h \ge 0$ for the anti-holomorphic one.

In conclusion, three-particle amplitudes for boost-breaking theories take the general form

$$\mathcal{A}_{3}(\{\lambda^{(i)},\tilde{\lambda}^{(i)};h_{i}\}) = \begin{cases} \langle 12\rangle^{h_{3}-h_{1}-h_{2}}\langle 23\rangle^{h_{1}-h_{2}-h_{3}}\langle 31\rangle^{h_{2}-h_{3}-h_{1}}F_{h_{1},h_{2},h_{3}}^{H}(E_{1},E_{2},E_{3}), & h \leq 0, \\ \\ [12]^{h_{1}+h_{2}-h_{3}}[23]^{h_{2}+h_{3}-h_{1}}[31]^{h_{3}+h_{1}-h_{2}}F_{h_{1},h_{2},h_{3}}^{AH}(E_{1},E_{2},E_{3}), & h \geq 0. \end{cases}$$

$$(6.48)$$

Note that in our convention particles are arranged cyclically in the order 123, and energy conservation $\sum E_i = 0$ ensures that F^H and F^{AH} can be reduced to functions of two variables only. Thus we will sometimes write

$$F(E_1, E_2) \equiv F(E_1, E_2, E_3 = -E_1 - E_2).$$
(6.49)

We will also drop the H/AH index unless it is necessary. Qualitatively, therefore, the only difference between the boost-invariant (see [30, 28]) and boost-breaking amplitudes is an arbitrary function of the energies that we can add to the latter thanks to the reduced set of symmetries. Our task in Section 6.4 will be to constrain these functions using the four-particle test. To recover the boost-invariant amplitudes one can simply set $F^{H,AH}$ to a constant.

Before going on to discuss some examples, we first show that the functions F^H and F^{AH} are not independent. They are related by a parity transformation (space inversion) P, which does not belong to the connected component of the identity of the Lorentz group. The amplitude can either stay the same (scalar) or inherit a minus sign (pseudoscalar) under P. The transformation of all the 4-momenta $(E, \mathbf{p}) \mapsto (E, -\mathbf{p})$ can be represented in spinor-helicity formalism by transforming the spinors according to¹³

$$\lambda_{\alpha} \mapsto \lambda_{\alpha}' = (-i\tilde{\lambda}_2, i\tilde{\lambda}_1), \qquad \qquad \tilde{\lambda}_{\dot{\alpha}} \mapsto \tilde{\lambda}_{\dot{\alpha}}' = (i\lambda_2, -i\lambda_1), \qquad (6.50)$$

¹³The presence of a factor of i is due to the requirement that the (+) polarization tensor should be transformed exactly into the (-) polarization tensor under spatial reflection.

which leads to $[ij] \mapsto -\langle ij \rangle$ and $\langle ij \rangle \mapsto -[ij]$. The helicities also change sign under P and so the helicity dependent part of the amplitude transforms as

$$[12]^{-d_3}[23]^{-d_1}[31]^{-d_2} \mapsto (-1)^d \langle 12 \rangle^{d_3} \langle 23 \rangle^{d_1} \langle 31 \rangle^{d_2}, \tag{6.51}$$

where $d = d_1 + d_2 + d_3 = -h$, and vice versa. Therefore requiring the amplitude to transform as scalar or psuedoscalar under P fixes

$$F_{h_1,h_2,h_3}^H(E_1,E_2,E_3) = \pm (-1)^h F_{-h_1,-h_2,-h_3}^{AH}(E_1,E_2,E_3),$$
(6.52)

with + for a scalar transformation and – for the pseudoscalar. We will therefore often quote results for F^H or F^{AH} only.

Let us finally emphasise that we have not assumed anything here other than the symmetries of the theory and locality. These amplitudes hold completely non-perturbatively and for any external particles, both bosonic and fermionic¹⁴.

6.3.2 Identical particles: symmetric and alternating polynomials

As an example, in this subsection we discuss the three-particle amplitudes for identical spin-S particles. Note that the spin-statistic theorem implies that S must be an integer in this case i.e. the particles are bosons. This is clear from (6.48) since for fermions each of the brackets has a fractional exponent and therefore when we exchange two fermions the amplitude does not transform into minus itself as it should by Fermi statistics. At the Lagrangian level there is no way to contract the SU(2) indices of three fermions to create a scalar quantity. This is the case for both boost-invariant and boost-breaking theories.

There are two fundamentally distinct helicity configurations with either two or three identical helicities. The corresponding amplitudes have mass dimension S and 3S respectively and so come from different operators. We can read off the amplitudes from (6.48). First consider the lowest dimension amplitudes $(\pm S, \pm S, \mp S)$ which take the form

¹⁴Fermions always come in pairs and so the exponents are always integers.

$$\mathcal{A}_{3}(1^{+S}2^{+S}3^{-S}) = \left(\frac{[12]^{3}}{[23][31]}\right)^{S} F^{AH}_{+S,+S,-S}(E_{1},E_{2}), \tag{6.53}$$

$$\mathcal{A}_3(1^{-S}2^{-S}3^{+S}) = \left(\frac{\langle 12\rangle^3}{\langle 23\rangle\langle 31\rangle}\right)^S F^H_{-S,-S,+S}(E_1,E_2),\tag{6.54}$$

where we have eliminated E_3 by energy conservation. Now, since particles 1 and 2 have the same helicity and they are bosons, the amplitudes must be invariant under their exchange. The spinor helicity part of these amplitudes inherits a factor of $(-1)^S$ under this transformation and so the functions of energy must be symmetric if the particles have even spin and anti-symmetric if they have odd spin:

$$F_{+S,+S,-S}^{AH}(E_1, E_2) = (-1)^S F_{+S,+S,-S}^{AH}(E_2, E_1),$$
(6.55)

$$F_{-S,-S,+S}^{H}(E_1, E_2) = (-1)^S F_{-S,-S,+S}^{H}(E_2, E_1).$$
(6.56)

To make further progress, we will assume that the functions F are polynomials divided by powers of E_1, E_2 and $E_1 + E_2^{15}$:

$$F(E_1, E_2) = \frac{f(E_1, E_2)}{E_1^a E_2^b (E_1 + E_2)^c},$$
(6.57)

It is easy to see that symmetry implies a = b for any spin¹⁶.

Now let us restrict to the case of even S where the functions f are required to be symmetric polynomials. By the fundamental theorem of symmetric polynomials, f can be written purely in terms of *elementary symmetric polynomials*. For n variables, there is a single elementary symmetric polynomial of degree m for all non-negative integers $m \le n$. If we label the n variables as $x_1 \dots x_n$ then the degree-m elementary symmetric polynomial is

$$e_m(x_1, \dots, x_n) = \sum_{1 \le j_1 < j_2 < \dots < j_m \le n} x_{j_1} \dots x_{j_m}.$$
(6.58)

For example, for n = 2 we have

¹⁵The factorisation constraints we derive in Section 6.4 will actually hold for more general functions of the energies too. In local theories with covariant fields we would expect no inverse powers of the energies but our results do indeed apply to more general scenarios.

¹⁶We can naturally assume that a, b and c are minimal. If a > b, then we would have $E_2^{a-b}f(E_1, E_2) = \pm E_1^{a-b}f(E_2, E_1)$ and thus $f(E_1, E_2)$ would be divisible by E_1 , contradicting the assumption that n was minimal.

$$\{1, x_1 + x_2, x_1 x_2\}. \tag{6.59}$$

On the other hand, if S is odd, the functions of energy in the numerators should be *alternating* polynomials. An alternating polynomial¹⁷ is defined by the property

$$\operatorname{Poly}(x_{\sigma(1)},\ldots,x_{\sigma(n)}) = \operatorname{sign}(\sigma)\operatorname{Poly}(x_1,\ldots,x_n), \qquad (6.60)$$

for any permutation σ of the *n* variables. All alternating polynomials can be written as the Vandermonde polynomial v_n multiplied by sums and products of any number of elementary symmetric polynomials and numerical coefficients (it's an ideal on the ring of polynomials). The Vandermonde polynomial is defined as

$$V_n(x_1,...,x_n) \equiv \prod_{1 \le i < j \le n} (x_j - x_i),$$
 (6.61)

and it is an alternating polynomial of order n(n-1)/2. In the case at hand the functions are of two variables (n = 2) and therefore the relevant Vandermonde polynomial is $V_2 = E_1 - E_2$. For the above amplitudes we therefore have

$$f_{+S,+S,-S} = \begin{cases} \operatorname{Poly}(E_1 + E_2, E_1 E_2) & \text{for } S \text{ even,} \\ (E_1 - E_2) \operatorname{Poly}(E_1 + E_2, E_1 E_2) & \text{for } S \text{ odd,} \end{cases}$$
(6.62)

and similarly for $f_{-S,-S,+S}$.

The remaining two three-particle amplitudes have mass dimension 3S and take the form

$$\mathcal{A}_3(1^{+S}2^{+S}3^{+S}) = ([12][23][31])^S F^{AH}_{+S,+S,+S}(E_1, E_2, E_3), \tag{6.63}$$

$$\mathcal{A}_{3}(1^{-S}2^{-S}3^{-S}) = (\langle 12 \rangle \langle 23 \rangle \langle 31 \rangle)^{S} F^{H}_{-S,-S,-S}(E_{1}, E_{2}, E_{3}).$$
(6.64)

Now the amplitudes need to be invariant under the exchange of any two external particles as they all have the same helicity. Thus, in 6.57 we require a = b = c. For even S the functions f must be symmetric polynomials, meaning that they are constructed out of the elementary symmetric polynomials with n = 3, namely

¹⁷Notice that the only object that is anti-symmetric under all possible permutations is zero. That's why anti-symmetric polynomials don't exist. The non-trivial objects are alternating polynomials, which are symmetric or anti-symmetric depending on the sign of the permutation.

$$\{1, x_1 + x_2 + x_3, x_1x_2 + x_2x_3 + x_1x_3, x_1x_2x_3\}.$$
(6.65)

For odd S the functions are constructed from these elementary symmetric polynomials multiplied by the order 3 alternating polynomial V_3 . We therefore have

$$f_{+S,+S,+S} = \begin{cases} \operatorname{Poly}(E_1E_2 + E_1E_3 + E_2E_3, E_1E_2E_3) & \text{for } S \text{ even,} \\ V_3(E_1, E_2, E_3) \operatorname{Poly}(E_1E_2 + E_1E_3 + E_2E_3, E_1E_2E_3) & \text{for } S \text{ odd,} \end{cases}$$
(6.66)

and similarly for $f_{-S,-S,-S}$. Note that for n = 3 we have $E_1 + E_2 + E_3 = 0$ since we are constructing on-shell amplitudes. So there are only two non-trivial elementary symmetric polynomials. Here we did not eliminate E_3 using energy conservation, so as to ensure that the permutation invariance of $F_{+S,+S,+S}$ remains manifest.

Scalar

If the identical particles are three scalars, i.e. S = 0, then the amplitude is simply a function of the energies:

$$\mathcal{A}_3(1^0 2^0 3^0) = F_{0,0,0}(E_1, E_2, E_3).$$
(6.67)

The helicity part of the amplitude disappears because scalars transform in a trivial way. In the boost-invariant case the amplitude is just a constant $F_{0,0,0} = \text{const.}$

Photon

For identical S = 1 particles, each of the four amplitudes presented above requires the functions of energy $F_{\pm 1,\pm 1,\mp 1}$ and $F_{\pm 1,\pm 1,\pm 1}$ to be alternating polynomials, possibly divided by powers of E_1E_2 and $(E_1 + E_2)$. This rules out the possibility of three-particle amplitudes for a photon in a boost-invariant theory, since a constant polynomial cannot be alternating. More generally any odd number of photons cannot self-interact. This well-known fact can be understood at the level of a Lagrangian where three-particle interactions for a single massless vector should be invariant under the U(1) gauge symmetry $A_{\mu} \rightarrow A_{\mu} + \partial_{\mu}\Lambda(x)$. The building block of invariant Lagrangians is the field strength $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ with the indices contracted with $\eta^{\mu\nu}$ or $\epsilon^{\mu\nu\rho\sigma}$ to produce a Lorentz scalar. Three-particle vertices therefore contain at least three derivatives and so the mass dimension of the three-particle amplitudes is dim $A_3 \ge 3$. This is the Lagrangian reason why the $(\pm 1, \pm 1, \pm 1)$ amplitudes vanish since they have mass dimension 1. For the $(\pm 1, \pm 1, \pm 1)$ amplitudes we can try to contract three powers of the field strength. However, all Lorentz scalars cubic in the fields, e.g. $F^{\mu}{}_{\nu}F^{\nu}{}_{\rho}F^{\rho}{}_{\mu}$, $\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\kappa}F_{\sigma}{}^{\kappa}$, vanish by symmetry¹⁸. This Lagrangian interpretation requires us to jump through a few hoops, most notably the introduction of a gauge redundancy to remove the additional degrees of freedom required to write down a manifestly Lorentz invariant and local Lagrangian. The on-shell approach where such redundancies are not required is clearly more efficient and elegant.

In a boost-breaking theory, we can use alternating polynomials in energies to ensure that each of the four three-particle amplitudes have the correct Bose symmetry. It is interesting that we can write down an amplitude of this form even though it has no boost-invariant counterpart. But one must first check if these amplitudes pass the four-particle test before declaring that such a theory is consistent (within our assumptions).

Graviton and higher spins

For identical particles with $S \ge 2$ and S even, we can write down three-particle amplitudes in both boost-invariant and boost-breaking theories, while for particles with S odd we can only write down such amplitudes in a boost-breaking theory, just like for S = 1. Note that the graviton helicity amplitudes are literally the square of the photon amplitudes. When we allow for multiple spin-1 particles, where Bose symmetry in boost-invariant theories is satisfied thanks to the anti-symmetric couplings (the structure constants), the structure of the amplitude is unchanged up to the addition of some colour indices. This simple observation is one of the reasons for the symbolic expression "GR = YM²" [219].

6.4 Four-particle amplitudes and the four-particle test

Having constructed general, non-perturbative three-particle amplitudes, we are now in the position to constrain the almost arbitrary functions of energy using the four-particle test. As explained in Section 6.2, tree-level four-particle amplitudes contain poles and regular pieces. The latter correspond to

¹⁸We can write down non-zero gauge invariant operators at quartic or higher order in the field strength, which describe the interaction of an *even* number of photons. Such terms appear in the Euler-Heisenberg Lagrangian, an effective description of QED below the mass of the electron.

contact diagrams while the former come from particle exchange illustrated in Figure 6.2. When the exchanged particle is taken on-shell, the amplitude approaches a singularity whose residue should factorise into a product of three-particle amplitudes. We use this feature to bootstrap consistent four-particle amplitudes due to exchange diagrams in boost-breaking theories. This bootstrap does not constrain the regular parts of the four-particle amplitude; we are constraining the singularity structure of four-particle amplitudes and therefore the cubic couplings in the process.



Figure 6.3: s, t and u-channel exchange diagrams, respectively.

To illustrate the idea behind this approach, we may first consider a naive attempt at writing down a four-particle amplitude that factorises into three-particle amplitudes. We have three channels, shown in figure 6.3, and so one could initially allow for three separate terms with an order one pole in s, t or u as follows¹⁹

$$\mathcal{A}_4 \stackrel{?}{=} \frac{\mathcal{A}_3(1,2,-I) \times \mathcal{A}_3(3,4,I)}{s} + \frac{\mathcal{A}_3(1,3,-I) \times \mathcal{A}_3(2,4,I)}{t} + \frac{\mathcal{A}_3(1,4,-I) \times \mathcal{A}_3(2,3,I)}{u}$$
(6.68)

where I and -I label the exchanged particle outgoing from the vertex involving particle 1, or incoming into that vertex respectively²⁰. All external particles are incoming. If more than one intermediate particle is allowed, we need to sum over all the species of I. Now it would appear that this amplitude has the residues required by Theorem 6.2.1. However, it is possible that $A_3(1, 2, -I) \times A_3(3, 4, I)$,

¹⁹We remind the reader that we are working with relativistic dispersion relations for each particle meaning that we only encounter poles in the usual boost-invariant Mandelstam variables.

²⁰Throughout our analysis in the spinor helicity variables we send $p_I \to -p_I$ by $\lambda^{(I)} \to \lambda^{(I)}$, $\tilde{\lambda}^{(I)} \to -\tilde{\lambda}^{(I)}$. See Appendix 6.7.1 for a justification of this method.

when analytically continued beyond the loci of s = 0, has a pole at t = 0 or u = 0. In this case, the first term contributes to the t = 0 or u = 0 residue and the formula 6.68 could give an incorrect residue at t = 0. Finding a four-particle amplitude with the correct residues in **all** three channels is therefore a non-trivial matter. This is known as the four-particle test [29, 209], and as we shall see, it allows us to constrain, or altogether eliminate, certain types of cubic interactions in boost-breaking theories.

Before we begin, we must identify a set of SO(3)-invariant variables that are sufficient to fully determine the on-shell data for the scattering of four particles. In addition to the four external helicities, we must use some of the brackets $\langle ij \rangle$, [ij] and (ij), which constitute a complete list of invariants of mass dimension 1. However, not all of these are independent: all but one of the off-diagonal (ij) brackets can be determined in terms of the other brackets and the energies by using momentum conservation.²¹ Therefore, any SO(3) invariant can be written in terms of $\langle ij \rangle$, [ij], E_i and just one of the off-diagonal (ij). These variables are still not all independent, but this won't present a problem for us. On the other hand, it must be emphasized that without at least one (ij) bracket we would be unable to fully determine the kinematic data in the general case. This means that in boost-breaking theories, four-particle amplitudes could depend on one of the (ij)'s and this dependence cannot be eliminated by application of bracket identities.

There is a special class of Lagrangians for which four-particle amplitudes are functions of $\langle ij \rangle$, [*ij*] and E_i only. These Lagrangians take the form

$$\mathcal{L} = \mathcal{L} \left[X_{\mu_1 \mu_2 \dots}, \eta_{\mu\nu}, \epsilon_{\mu\nu\sigma\rho}, \partial_{\mu}, \partial_t \right], \tag{6.69}$$

where $X_{\mu_1\mu_2\dots}$ collectively denotes Lorentz covariant fields. If a physical four-particle amplitude can be written solely in terms of $\langle ij \rangle$, [ij] and E_i , then there exists a Lagrangian of the form (6.69) which generates this amplitude. Such a Lagrangian can be constructed as follows: first, write down a Lorentz-invariant Lagrangian that generates the four-particle amplitude with the energy dependence stripped off, and then insert time derivatives acting on appropriate fields to reinstate the desired energy dependence of the amplitude. Suppose, on the other hand, that a four-particle amplitude in some theory cannot be written without at least one round bracket (ij) (which, as we remarked, cannot be

²¹We verified this via algebraic manipulation in Mathematica.

determined solely in terms of the $\langle ij \rangle$, [ij] and E_i). Then the corresponding Lagrangian must depend on some objects other than the ones listed in (6.69). For example, the Larangian could be constructed out of SO(3) covariant fields rather than Lorentz covariant ones.

As an example of the latter kind of theory, let us consider the Framid EFT [162] which arises from the spontaneous breaking of Poincaré symmetry to an unbroken subgroup of translations and rotations. Indeed, the Framid degrees of freedom are the Goldstone modes of broken Lorentz boosts. With respect to the unbroken SO(3) symmetry, the Framid consistents of three degrees of freedom: a massless tranverse vector and a massless longitudinal scalar with speeds c_T and c_L . Taking $c_L = c_T$, in which the scalar and vector modes have identical propagation speeds as we have been assuming in this work, the Framid Lagrangian up to cubic order in fields takes the form [162]

$$\mathcal{L} = \frac{M_1^2}{2} \left(\dot{\eta}_i^2 - c_L^2 \partial_i \eta_j \partial_i \eta_j \right) + M_1^2 \left(c_L^2 - 1 \right) \eta_i \partial_i \eta_j \dot{\eta}_j + \mathcal{O} \left(\eta^4 \right) \,. \tag{6.70}$$

After defining rescaled fields $\chi_i = c_L M_1 \eta_i$ and replacing t with the rescaled time coordinate $t' = t/c_L$, we obtain

$$\mathcal{L} = \frac{1}{2} \left(\dot{\vec{\chi}}^2 - \partial_i \chi_j \partial_i \chi_j \right) + \left(\frac{c_L^2 - 1}{c_L^2 M_1} \right) \chi_i \partial_i \chi_j \dot{\chi}_j + \mathcal{O} \left(\chi^4 \right) \,. \tag{6.71}$$

Using the above Lagrangian (and rescaled coordinates), we computed the four-particle amplitude $\mathcal{A}_4(1^02^+3^04^-)$ from tree-level exchange to verify and illustrate that it has an explicit dependence on one of the off-diagonal (*ij*), which cannot be eliminated. The result is the simplest, albeit still quite lengthy, if we allow for the dependence on (42), in which case the amplitude reads as follows:

$$\begin{aligned} \mathcal{A}_{4}(1^{0}2^{+}3^{0}4^{-}) &= \frac{1}{4e_{4}} \left(\frac{c_{L}^{2}-1}{c_{L}^{2}M_{1}}\right)^{2} \times \\ &\times \left\{\frac{1}{s} \left[F_{(1,a)}(E_{1}, E_{2}, E_{3}, E_{4}; s, t)(42)^{2} \right. \\ &+ F_{(1,b)}(E_{1}, E_{2}, E_{3}, E_{4}; s, t)[23]\langle 34 \rangle (42) \right. \\ &+ F_{(1,c)}(E_{1}, E_{2}, E_{3}, E_{4}; s, t)[23]^{2}\langle 34 \rangle^{2} \right] \\ &+ \frac{1}{t} \left[F_{(2,a)}(E_{1}, E_{2}, E_{3}, E_{4}; s, t)(42)^{2} \right. \\ &+ F_{(2,b)}(E_{1}, E_{2}, E_{3}, E_{4}; s, t)[23]\langle 34 \rangle (42) \right] \\ &+ \frac{1}{u} \left[F_{(1,a)}(E_{3}, E_{2}, E_{1}, E_{4}; u, t)(42)^{2} \right. \\ &- F_{(1,b)}(E_{3}, E_{2}, E_{1}, E_{4}; u, t)[23]\langle 34 \rangle (42) \end{aligned}$$

$$+F_{(1,c)}(E_3, E_2, E_1, E_4; u, t)[23]^2 \langle 34 \rangle^2] \bigg\}, \qquad (6.72)$$

where functions $F_{(i,x)}$ are defined in Appendix 6.7.2 and $e_4 \equiv E_1 E_2 E_3 E_4$.

Since an ansatz that depends on round brackets would be too general to be constrained effectively, to make progress we will assume that four-particle amplitudes take the form

$$\mathcal{A}_4 = \mathcal{A}_4\left(\langle ij\rangle, [ij], s, t, u, E_i\right),\tag{6.73}$$

meaning that the underlying Lagrangians take the form of (6.69). For more general Lagrangians, we would have to allow for the presence of (42) (or some other off-diagonal (ij)) in the four-particle amplitude. We plan to come back to this in the future.

6.4.1 Single spin-S particle

We begin by constraining the lowest dimension three-particle amplitudes for identical spin-S bosons presented in (6.53), namely the $(\pm S, \pm S, \mp S)$ amplitudes. Consider the four-particle amplitude $\mathcal{A}_4(1^{-S}2^{+S}3^{-S}4^{+S})$ due to exchange of the spin-S particle. By little group scaling we can fix the helicity part of the amplitude leaving only the dependence on the little group invariants (s, t, u, E_i) left to fix by the four-particle test. The amplitude takes the general form

$$\mathcal{A}_4(1^{-S}2^{+S}3^{-S}4^{+S}) = \langle 13 \rangle^{2S} [24]^{2S} \mathcal{G}(s, t, u, E_i), \tag{6.74}$$

and its mass dimension (recall that we don't count the explicit energy dependence in the mass dimension) is

$$\dim \mathcal{A}_4 = 4S + \dim \mathcal{G}. \tag{6.75}$$

Now for exchanges in the *s* and *u* channels both constituent three-particle amplitudes have mass dimension *S* and this can also be achieved in the *t* channel for one of the two possible helicity configurations of the exchanged particle. Since factorisation requires $\lim_{s\to 0} (sA_4) = A_3 \times A_3$, for the case at hand the mass dimension of the four-particle amplitude is

$$\dim \mathcal{A}_4 = 2S - 2. \tag{6.76}$$

~

By equating (6.75) and (6.76) we find that the mass dimension of \mathcal{G} satisfies

$$\dim \mathcal{G} = -2S - 2. \tag{6.77}$$

However, locality dictates that the amplitude can only contain simple poles in s, t and u and so we require dim $\mathcal{G} \ge -6$ yielding the constraint

$$S \le 2. \tag{6.78}$$

This tell us that the above four-particle amplitude is inconsistent for bosonic particles with $S \ge 3$, even in boost-breaking theories. We require the corresponding $(\pm S, \pm S, \mp S)$ amplitudes to vanish, so we set $F_{-S,-S,+S} = F_{+S,+S,-S} = 0$ for $S \ge 3$. This very simple argument leads to a profound result: massless, higher spinning particles cannot have low-energy cubic self-interactions (under the assumption that the underlying Lagrangian is written in terms of covariant fields).

Let us consider this amplitude in more detail for S = 0, 1, 2 where dimensional analysis did not exclude the possibility of consistent factorization. In the *s* and *u* channels there are two distinct diagrams since we have two choices for the helicity configuration of the exchanged particle (see Figure 6.4 for the two *s*-channel possibilities). In the *t* channel there is only one diagram. We therefore have two residues to compute in the *s* and *u* channels and we label these as R_s^{-+} , R_s^{+-} and R_u^{-+} , R_u^{+-} . Using the three-particle amplitudes (6.53) the residue on the s = 0 pole is



Figure 6.4: Two choices for the helicity configuration of the exchanged particle.

$$R_{s} = R_{s}^{-+} + R_{s}^{+-}$$

$$= \left(\frac{\langle I1 \rangle^{3}}{\langle 12 \rangle \langle 2I \rangle}\right)^{S} \left(\frac{[4I]^{3}}{[I3][34]}\right)^{S} F_{-S,-S,+S}(-E_{1} - E_{2}, E_{1})F_{+S,+S,-S}(E_{4}, -E_{3} - E_{4})$$
(6.79)

+
$$\left(\frac{[2I]^3}{[I1][12]}\right)^S \left(\frac{\langle I3\rangle^3}{\langle 34\rangle\langle 4I\rangle}\right)^S F_{+S,+S,-S}(E_2,-E_1-E_2)F_{-S,-S,+S}(-E_3-E_4,E_3),$$

where we have used energy conservation to eliminate E_I . Now in the spinor helicity variables there is not a unique way to approach s = 0. We have $s = \langle 12 \rangle [12] = \langle 34 \rangle [34] = 0$ and this has two main solutions. If [12] = 0, then by momentum conservation we have $0 = [12]\langle 23 \rangle = [14]\langle 34 \rangle$ and so to avoid imposing additional constraints on the kinematics we have to choose $\langle 34 \rangle = 0$. Similarly, if $\langle 12 \rangle = 0$, then [34] = 0 too.

For $[12] = \langle 34 \rangle = 0$, the second term in (6.79) vanishes²² leaving

$$R_{s} = R_{s}^{-+} = \left(\frac{\langle I1\rangle^{3}}{\langle 12\rangle\langle 2I\rangle}\right)^{S} \left(\frac{[4I]^{3}}{[I3][34]}\right)^{S} F_{-S,-S,+S}(-E_{1} - E_{2}, E_{1})F_{+S,+S,-S}(E_{4}, -E_{3} - E_{4})$$
$$= \frac{(\langle I3\rangle^{2}[24]^{2})^{S}}{t^{S}} F_{-S,-S,+S}(-E_{1} - E_{2}, E_{1})F_{+S,+S,-S}(E_{4}, -E_{3} - E_{4}), \tag{6.80}$$

where using conservation of momentum at each vertex we eliminated all factors of I, for example, $\langle 1I \rangle [I4] = \langle 12 \rangle [24]$. For $\langle 12 \rangle = [34] = 0$ the first term vanishes leaving

$$R_{s} = R_{s}^{+-} = \left(\frac{[2I]^{3}}{[I1][12]}\right)^{S} \left(\frac{\langle I3 \rangle^{3}}{\langle 34 \rangle \langle 4I \rangle}\right)^{S} F_{+S,+S,-S}(E_{2}, -E_{1} - E_{2})F_{-S,-S,+S}(-E_{3} - E_{4}, E_{3})$$
$$= \frac{(\langle 13 \rangle^{2} [24]^{2})^{S}}{t^{S}} F_{+S,+S,-S}(E_{2}, -E_{1} - E_{2})F_{-S,-S,+S}(-E_{3} - E_{4}, E_{3}).$$
(6.81)

Again we see how $S \ge 3$ amplitudes are ruled out: for $S \ge 3$, the *s*-channel residue contains higher order poles when t = 0 and so the corresponding amplitude is inconsistent. One may also think that S = 2 is problematic since the denominator is quadratic in t. However, when s = 0 we can write $t^2 = -tu$. Before moving on to the other channels, we note that the residue in the *s*-channel should not differ if we approach the pole in two different ways and so we match the two different expressions for R_s yielding our first constraint on the three-particle amplitudes²³:

$$F_{-S,-S,+S}(-E_1 - E_2, E_1)F_{+S,+S,-S}(E_4, -E_3 - E_4)$$

= $F_{+S,+S,-S}(E_2, -E_1 - E_2)F_{-S,-S,+S}(-E_3 - E_4, E_3).$ (6.82)

²²Once we eliminate I from all brackets, one sees that the numerator vanishes faster than the denominator.

²³Here is a brief justification. Near s = 0, the schematic form of the amplitude is $A \sim s^{-1}(f_1(\lambda)F_1(E) + f_2(\lambda)F_2(E))$, where f_i are functions of the Lorentz invariants and F_i are functions of the energies only. The amplitude has the same dependence on the Lorentz invariants in the two limits, which can then differ only by a function of energies. Hence, we can write $A \sim s^{-1}f(\lambda)F(E)$. Since we can take either of the limits $\langle 12 \rangle \rightarrow 0$ or $[12] \rightarrow 0$ while keeping the energies fixed, we must get the same F(E), which is to be identified with the energy-dependent functions in the main text.

In the boost-invariant limit the two residues are trivially the same.

The u-channel also contains two diagrams and the corresponding residues can easily be obtained from the s-channel ones by interchanging particles 2 and 4. With (6.82) imposed the two residues are equivalent. We have, for example,

$$R_u = R_u^{-+} = \frac{(\langle 13 \rangle^2 [24]^2)^S}{t^S} F_{+S,+S,-S}(E_4, -E_1 - E_4) F_{-S,-S,+S}(-E_3 - E_2, E_3).$$
(6.83)

Finally, the *t*-channel is qualitatively different since it involves two particles of the same helicity on each side of the diagram. There is therefore only a single choice for the exchange particle's helicity if this contribution to the amplitude is to have the same mass dimension as the other channels. The residue is

$$R_{t} = \left(\frac{\langle 13\rangle^{3}}{\langle 3I\rangle\langle I1\rangle}\right)^{S} \left(\frac{[24]^{3}}{[4I][I2]}\right)^{S} F_{-S,-S,+S}(E_{1},E_{3})F_{+S,+S,-S}(E_{2},E_{4})$$

$$= \frac{(\langle 13\rangle^{2}[24]^{2})^{S}}{s^{S}} F_{-S,-S,+S}(E_{1},E_{3})F_{+S,+S,-S}(E_{2},E_{4}).$$
(6.84)

In summary, the residues are

$$R_{s} = \frac{(\langle 13 \rangle^{2} [24]^{2})^{S}}{t^{S}} F_{-S,-S,+S}(-E_{1} - E_{2}, E_{1}) F_{+S,+S,-S}(E_{4}, -E_{3} - E_{4}), \quad (6.85)$$

$$R_t = \frac{(\langle 13 \rangle^2 [24]^2)^S}{s^S} F_{-S,-S,+S}(E_1, E_3) F_{+S,+S,-S}(E_2, E_4),$$
(6.86)

$$R_u = \frac{(\langle 13 \rangle^2 [24]^2)^S}{t^S} F_{+S,+S,-S}(E_4, -E_1 - E_4) F_{-S,-S,+S}(-E_3 - E_2, E_3), \quad (6.87)$$

and are subject to (6.82). Let us now zoom in on the three different allowed values for S.

Scalar

For a single scalar, S = 0, consistent factorisation is trivial. Indeed, each residue is simply a function of the energies and does not contain spurious poles. The consistent four-particle amplitude is

$$\mathcal{A}_{4}(1^{0}, 2^{0}, 3^{0}, 4^{0}) = \frac{F(-E_{1} - E_{2}, E_{1})F(E_{4}, -E_{3} - E_{4})}{s} + \frac{F(E_{1}, E_{3})F(E_{2}, E_{4})}{t} + \frac{F(E_{4}, -E_{1} - E_{4})F(-E_{3} - E_{2}, E_{3})}{u},$$
(6.88)
where $F \equiv F_{0,0,0}$. The only constraint we have on the function of energy is that it should be a symmetric function as explained in Section 6.2.

We can understand this result from a Lagrangian point of view. In the boost-invariant case the three-particle amplitude is a constant with consistent factorisation of the four-particle amplitude for scalar scattering. One may wonder about cubic vertices with derivatives. It is easy to contract the indices in a Lorentz invariant way but these vertices always involve, up to integration by parts, the $\Box = \partial^{\mu}\partial_{\mu}$ operator acting on at least one of the fields and therefore it vanishes on-shell and can be removed by a field redefinition in favour of four-point vertices which only contribute to the regular part of the four-particle amplitude.

In the boost-breaking case we write operators using the usual Lorentzian derivative ∂_{μ} , but also have the freedom to add extra time derivatives. Because any terms with Lorentzian derivatives can be removed by a field redefinition, the only non-trivial three scalar vertices have zero derivatives, corresponding to a constant amplitude, or contain time derivatives only giving rise to functions of energy in the amplitude. A well-known example is the $\dot{\phi}^3$ vertex appearing in the flat space, decoupling limit of the EFT of single-field inflation. Generalisations with more derivatives are easy to write down.

Photon

For a photon, S = 1, consistent factorisation becomes a nontrivial problem: R_s has a pole when t = 0, R_t has a pole when u = 0, and R_u has a pole when s = 0. Therefore the full amplitude must take the form

$$\mathcal{A}_4(1^{-1}2^{+1}3^{-1}4^{+1}) = \langle 13 \rangle^2 [24]^2 \left(\frac{A}{st} + \frac{B}{tu} + \frac{C}{us}\right), \tag{6.89}$$

where A, B and C are constrained by

$$R_s = \langle 13 \rangle^2 [24]^2 \left(\frac{C-A}{u}\right), \tag{6.90}$$

$$R_t = \langle 13 \rangle^2 [24]^2 \left(\frac{A-B}{s}\right), \tag{6.91}$$

$$R_u = \langle 13 \rangle^2 [24]^2 \left(\frac{B-C}{t}\right), \qquad (6.92)$$

where again we have used s + t + u = 0. As explained in Section 6.3, $F_{-1,-1,+1}$ and $F_{+1,+1,-1}$ are proportional with the proportionality factor \pm for parity odd and even theories respectively. Since only their product appears in each residue the following analysis is the same in both cases, so without loss of generality let us take $F = F_{-1,-1,+1} = F_{+1,+1,-1}$. Matching our two expressions for the residues yields

$$C - A = -F(E_2, -E_1 - E_2)F(-E_3 - E_4, E_3),$$
(6.93)

$$A - B = F(E_1, E_3)F(E_2, E_4), (6.94)$$

$$B - C = F(E_4, -E_1 - E_4)F(-E_2 - E_3, E_3),$$
(6.95)

with

$$F(-E_1 - E_2, E_1)F(E_4, -E_3 - E_4) = F(E_2, -E_1 - E_2)F(-E_3 - E_4, E_3),$$
(6.96)

such that the residues in the s and u channels are the same regardless of how we approach the pole. Taking the sum of (6.93), (6.94) and (6.95) yields the main S = 1 factorisation constraint

$$F(E_2, -E_1 - E_2)F(-E_3 - E_4, E_3)$$

$$- F(E_1, E_3)F(E_2, E_4)$$

$$- F(E_4, -E_1 - E_4)F(-E_2 - E_3, E_3) = 0,$$
(6.97)

which must be satisfied for all E_i subject to $E_1 + E_2 + E_3 + E_4 = 0$.

Recall from Section 6.3 that F must be an alternating polynomial (possibly divided by some powers of energies) such that the three-particle amplitudes have the correct Bose symmetry. Since F is an alternating function of two variables, we can write

$$F(x,y) = \frac{(x-y)P[x+y,xy]}{x^m y^m (x+y)^k},$$
(6.98)

with $xy \nmid P[x + y, xy]$ if m > 0 and $(x + y) \nmid P[x + y, xy]$ if k > 0 (\nmid means "does not divide"). Writing the factorisation constraint (6.97) in terms of P, we can prove that it requires $P \equiv 0$. The reason for this is that P, as we show in Appendix 6.7.3, has to satisfy infinitely many distinct constraints of the form $P[x, a_k x^2] = 0 \forall x$ and thus we need $(a_k x^2 - y)$ to divide P[x, y] for all the a_k , which is impossible if P is a nonzero polynomial. We therefore conclude that the four-particle test requires the $(\pm 1, \pm 1, \pm 1)$ three-particle amplitudes for a single photon in a boost-breaking theory (formulated in terms of covariant fields) to vanish: even when boosts are broken there are no consistent three-point vertices for a single photon giving rise to these lowest dimension amplitudes. Note that this result did not require us to impose the additional constraint (6.96) from matching the residues. One may wonder if consistent amplitudes are possible if we include additional particles, but we will show in Section 6.4.2 that additional exchanges do not change this result.

In a theory with only a single photon the four-particle test cannot constrain the other three-particle amplitudes, namely those with $(\pm 1, \pm 1, \pm 1)$ helicities since these amplitudes do not contain inverse powers of brackets and therefore residues constructed out of these amplitudes cannot contain poles. These three-particle amplitudes are therefore only constrained by Bose symmetry which for S = 1tells us that $F_{-1,-1,-1}$ and $F_{+1,+1,+1}$ are alternating functions in the three energies. Amplitudes of the lowest possible dimension are²⁴

$$\mathcal{A}_3(1^{-1}2^{-1}3^{-1}) = g\langle 12\rangle\langle 23\rangle\langle 31\rangle \frac{(E_1 - E_2)(E_2 - E_3)(E_1 - E_3)}{E_1 E_2 E_3},\tag{6.99}$$

$$\mathcal{A}_3(1^{+1}2^{+1}3^{+1}) = \pm g[12][23][31] \frac{(E_1 - E_2)(E_2 - E_3)(E_1 - E_3)}{E_1 E_2 E_3}, \tag{6.100}$$

while the first amplitudes arising from a U(1) gauge invariant theory are

$$\mathcal{A}_3(1^{-1}2^{-1}3^{-1}) = g'\langle 12\rangle\langle 23\rangle\langle 31\rangle(E_1 - E_2)(E_2 - E_3)(E_1 - E_3), \tag{6.101}$$

$$\mathcal{A}_3(1^{+1}2^{+1}3^{+1}) = \pm g'[12][23][31](E_1 - E_2)(E_2 - E_3)(E_1 - E_3), \tag{6.102}$$

where we allow for parity-even and parity odd possibilities and g, g' are coupling constants. All of these amplitudes are consistent since the four-particle test for photon scattering does not impose any conditions on (+1, +1, +1) and (-1, -1, -1) interactions.

Let us briefly comment on the Lagrangian approach to all-plus (and all-minus) amplitudes. Despite the fact that (6.99) - (6.100) are allowed by symmetry and the 4p test, they cannot arise from a gauge invariant cubic term. This is because gauge invariance requires us to construct interactions out of the

 $^{^{24}}$ We acknowledge Maria Alegria Gutierrez's findings on the possible structures of $F_{\pm 1,\pm 1,\pm 1}.$

field strength $F_{\mu\nu}$, which already contains three derivatives, but $F^{\mu}_{\nu}F^{\nu}_{\rho}F^{\rho}_{\mu}$ vanishes identically. In contrast, for (6.101) - (6.102) there exists an underlying local Lagrangian which is gauge invariant. By taking boost-invariant interactions and adding time derivatives we find both a parity-even and parity-odd possibility given by

$$\ddot{F}^{\mu}{}_{\nu}\dot{F}^{\nu}{}_{\rho}F^{\rho}{}_{\mu}, \qquad \epsilon^{\mu\nu\rho\sigma}\ddot{F}_{\mu\nu}\dot{F}_{\rho\kappa}F_{\sigma}{}^{\kappa}. \tag{6.103}$$

In Appendix 6.7.4 we show that the latter interaction does indeed give rise to the purported amplitudes (6.101) - (6.102). The calculation for the first interaction is similar.

In conclusion, boost-breaking theories of a single photon do exist but any gauge invariant cubic interactions require at least 6 derivatives meaning that its low energy consequences are heavily suppressed. In addition, in Section 6.4.3 we will show that in the presence of gravity these interactions do not pass the four-particle test!

Graviton

The graviton, S = 2, is the final case to consider. Here we see that each residue contains a pole in the other two Mandelstam variables and so consistent factorisation is non-trivial. This tells us that a four-particle amplitude with consistent factorisation must take the form

$$\mathcal{A}_4(1^{-2}2^{+2}3^{-2}4^{+2}) = \langle 13 \rangle^4 [24]^4 \frac{A}{stu}, \tag{6.104}$$

with the function A constrained by matching to each residue. Our S = 2 factorisation conditions are

$$-A = F(-E_1 - E_2, E_1)F(E_4, -E_3 - E_4)$$
(6.105)

$$= F(E_1, E_3)F(E_2, E_4), (6.106)$$

$$= F(E_4, -E_1 - E_4)F(-E_3 - E_2, E_3), (6.107)$$

where again we have dropped the subscripts denoting the helicities, and cover both parity even and parity odd cases. We also need to satisfy (6.82).

In Appendix 6.7.3 we show that the only solution to this set of equations, given that F is now a

symmetric polynomial multiplied by inverse energies, is F = const. This reduces the $(\pm 2, \pm 2, \mp 2)$ three-particle amplitudes, and the four-particle amplitude due to these vertices, to the boost-invariant limit. The four-particle amplitude is then what one finds in General Relativity (GR). Indeed, in this boost-invariant limit the three-particle amplitudes have mass dimension 2 which is due to the two-derivative nature of the Einstein-Hilbert action. Note that the minus sign in the overall amplitude is because gravity is an attractive force. We denote the magnitude of the three-gravity coupling as κ .

As with the photon case, we may have anticipated this result from a Lagrangian point of view. In GR the required gauge redundancy is diffeomorphism invariance under which the spacetime coordinates transform. Furthermore, the quantum effective theory of GR is best understood by expanding the Einstein-Hilbert action around the vacuum solution $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. One finds a tower of two-derivative terms with each coupling fixed by diffeomorphisms relating operators at different orders in $h_{\mu\nu}$. Given that in this work the two-derivative kinetic term is assumed to be of the boost-invariant form, adding time derivatives to the cubic vertex would break the (linearised) diffeomorphism symmetry and one would therefore expect issues to arise. However, let us again emphasise that although this Lagrangian interpretation can yield some intuition, the on-shell analysis presented here is preferable given that it is independent of gauge redundancies and field redefinitions. As we shall see in Section 6.4.3, the analysis is also robust against adding additional particles.

Now in contrast to the photon case, here we can constrain the other three-particle amplitudes $(\pm 2, \pm 2, \pm 2)$ thanks to the non-vanishing GR amplitudes. The dimension 6 amplitudes are

$$\mathcal{A}_{3}(1^{-2}2^{-2}3^{-2}) = (\langle 12 \rangle \langle 23 \rangle \langle 31 \rangle)^{2} F_{-2,-2,-2}(E_{1}, E_{2}, E_{3}), \tag{6.108}$$

$$\mathcal{A}_3(1^{+2}2^{+2}3^{+2}) = ([12][23][31])^2 F_{+2,+2,+2}(E_1, E_2, E_3), \tag{6.109}$$

where $F_{-2,-2,-2}$ and $F_{+2,+2,+2}$ are symmetric polynomials. Now consider the four-particle amplitude $\mathcal{A}_4(1^{+2}, 2^{+2}, 3^{+2}, 4^{-2})$. We can arrange the helicities of the exchanged particle such that each residue has mass dimension 8 and going through an analysis mirroring those above we see that the amplitude takes the form

$$\mathcal{A}_4(1^{+2}, 2^{+2}, 3^{+2}, 4^{-2}) = [12]^4 [23]^4 \langle 24 \rangle^4 \frac{B}{stu}, \tag{6.110}$$

and consistent factorisation requires

$$-B = \kappa F_{+2,+2,+2}(E_1, E_2, -E_1 - E_2)$$
(6.111)

$$=\kappa F_{+2,+2,+2}(E_1, E_3, -E_1 - E_3) \tag{6.112}$$

$$=\kappa F_{+2,+2,+2}(E_2, E_3, -E_2 - E_3). \tag{6.113}$$

It is clear that the only solution to this system, for generic energies, is $F_{+2,+2,+2} = \text{const.}$ We therefore also have $F_{-2,-2,-2} = \text{const}$ by parity and so the amplitudes are reduced to their boost-invariant limits.

At the Lagrangian level, these mass dimension 6 three-particle amplitudes are due to terms cubic in the Riemann tensor. Note that there are no three-particle amplitudes with mass dimension 4. One may expect terms quadratic in curvature, R^2 , $R^2_{\mu\nu}$ and $R^2_{\mu\nu\rho\sigma}$, to give rise to mass dimension 4 amplitudes. However, in 4D the Riemann squared term is degenerate with the other two up to the Gauss-Bonnet total derivative and both of these can be removed by a field redefinition since they are proportional to $R_{\mu\nu}$ which vanishes on-shell. One may also wonder about terms with four or more powers of curvature, but these do not contribute to three-particle amplitudes since at cubic order in fluctuations at least one curvature would need to be evaluated on the flat background where it vanishes.

Brief Summary

Let us briefly summarise our results for a single spin-S particle:

- For S = 0 factorisation is trivial with each residue a function of the external energies.
- For S = 1 the four-particle test forces the leading order three-particle amplitudes to vanish. This result assumes that the functions of energies are polynomials divided by some powers of the energies, but does not rely on any specific truncation of such polynomials. The highest dimension three-particle amplitudes are unconstrained by the four-particle test and at the level of a Lagrangian, the leading order gauge invariant vertices are (6.103).
- For S = 2 all three-particle amplitudes are forced to their boost-invariant limit. These are the amplitudes in GR with the addition of a term cubic in curvature. Again we assume that the functions of energies are polynomials divided by some powers of energies and our result does not rely on a truncation of the numerator. Lorentz violation in graviton cubic vertices is therefore impossible for a relativistic on-shell condition, in contrast to the photon.

 For S ≥ 3 the four-particle test cannot be passed and there cannot be any cubic self-interactions for these particles, at least to leading order in derivatives. This is potentially tricky to understand at the level of a Lagrangian, but here simple dimensional analysis and the four-particle test ruled out these vertices.

In the following sections we will see that these results are robust against including additional massless particles.

6.4.2 Couplings to a photon: Compton scattering and beyond

We now move to couplings between spin-S particles and a photon. We take $S \neq 1$ as we will consider multiple spin-1 particles in Section 6.4.4. Apart from this restriction, we allow for both bosonic and fermionic particles. We initially consider Compton scattering $\mathcal{A}_4(1_a^{-S}, 2^{+1}, 3_b^{+S}, 4^{-1})$ to constrain the $(+S, -S, \pm 1)$ amplitudes, allowing for multiple spin-S particles since in the boost-invariant limit a single copy cannot have a U(1) charge. These amplitudes have mass dimension 1 and so correspond to the familiar cubic couplings of a charged particle. We then present a complete analysis, i.e. we constrain all amplitudes that can be constrained, for a theory of a single scalar coupled to a photon. Couplings to a graviton are studied in Section 6.4.3.

Compton scattering

Consider the amplitude $\mathcal{A}_4(1_a^{-S}, 2^{\pm 1}, 3_b^{\pm S}, 4^{-1})$ with dim $\mathcal{A}_4 = 0$. Each residue must have mass dimension 2 which in turn must come from two mass dimension 1 three-particle amplitudes²⁵. First consider the *s*-channel where there are two possibilities for the spin of the exchanged particle. We can exchange a spin-*S* particle or a spin-|S - 2| particle. However, we find that the latter case yields spurious poles for all *S* and so consistency demands that the $(\mp S, \pm (S - 2), \pm 1)$ amplitudes vanish. For the former case we use the three-particle amplitudes

$$A_3(1_a^{-S}, 2_b^{+S}, 3^{-1}) = \langle 12 \rangle^{-1} \langle 23 \rangle^{1-2S} \langle 31 \rangle^{2S+1} F_{ab}^H(E_1, E_2), \tag{6.114}$$

$$A_3(1_a^{-S}, 2_b^{+S}, 3^{+1}) = [12]^{-1} [23]^{2S+1} [31]^{1-2S} F_{ab}^{AH}(E_1, E_2),$$
(6.115)

 $^{^{25}}$ It is not possible to exchange a particle such that one three-particle amplitude is dimensionless and the other has mass dimension 2.

where we have dropped the helicity subscripts on the *F*'s in favour of the internal indices (a, b) labelling the external spin-*S* particles, and have used energy conservation to eliminate E_3 . Computing the *s*-channel residue we find

$$(R_s)_{ab} = \frac{(\langle 14 \rangle [23])^{2S} (\langle 34 \rangle [23])^{2-2S}}{u} \sum_{e} F_{ae}^{AH} (E_1, -E_1 - E_2) F_{eb}^H (-E_3 - E_4, E_3), \quad (6.116)$$

where we have summed over the possible spin-S exchanged particles.

Moving to the *t*-channel, we see that we must exchange a photon to realise the desired mass dimension. A non-zero residue then requires non-zero three-photon amplitudes $(-1, +1, \pm 1)$. In Section 6.4.1 we showed that in the absence of other particles these amplitudes must vanish but since we have now included additional particles, we have to check if this result still holds. Going back to the amplitude $\mathcal{A}_4(1^{-1}, 2^{+1}, 3^{-1}, 4^{+1})$, we see that in the *s* and *u* channels only photon exchange can yield a dimensionless amplitude while in the *t*-channel we can exchange a photon, as we considered in Section 6.4.1, but can also exchange a spin-3 particle. The required three-particle amplitudes are $(\pm 1, \pm 1, \pm 3)$ but we find that such a residue induces spurious poles in *t* and therefore consistency requires these three-particle amplitudes to vanish. So our result in Section 6.4.1 on the absence of a consistent mass dimension 1 three-particle amplitude for photons is unchanged when we allow for additional exchanges. It follows that there is no *t*-channel contribution for Compton scattering.

Finally, for *u*-channel exchange we again find two possibilities for the exchanged particle: we can exchange a spin-S particle or a spin-(S + 2) particle. As in the *s*-channel we find that the latter choice yields spurious poles for all S and so the $(\mp S, \pm (S + 2), \mp 1)$ amplitudes must vanish. For the former case we find that the residue is

$$(R_u)_{ab} = \frac{(\langle 14\rangle [23])^{2S} (\langle 34\rangle [23])^{2-2S}}{s} \sum_e F_{ae}^H (E_1, -E_1 - E_4) F_{eb}^{AH} (-E_3 - E_2, E_3), \quad (6.117)$$

where again we have summed over the possible spin-S exchanged particles. Now we see a fundamental difference between the two cases S < 1 and S > 1. For S > 1, each residue contains a spurious pole in $(\langle 34 \rangle [23])$ meaning that no consistent four-particle amplitude is possible. The four-particle test therefore requires the $(+S, -S, \pm 1)$ three-particle amplitudes to vanish for S > 1, implying that such a particle cannot have a U(1) charge. This result is known in the boost-invariant limit and here we see that it is unchanged when we allow for the breaking of Lorentz boosts. Compton scattering is therefore only possible for low spins with S = 0, 1/2. The test is still non-trivial in these cases, since

consistent factorisation yields the constraints

$$\sum_{e} F_{ae}^{AH}(E_1, -E_1 - E_2) F_{eb}^{H}(-E_3 - E_4, E_3)$$

=
$$\sum_{e} F_{ae}^{H}(E_1, -E_1 - E_4) F_{eb}^{AH}(-E_3 - E_2, E_3),$$
 (6.118)

which needs to be satisfied for all E_i subject to $E_1 + E_2 + E_3 + E_4 = 0$. Again these constraints are the same for parity even and parity odd amplitudes so we will drop the H/AH labels in the following. These factorisation constraints are solved by $F_{ab} = f_{ab}F(E_1 + E_2)$ where f_{ab} is a constant matrix, and F is an arbitrary function of the sum $E_1 + E_2^{26}$. For bosons, f_{ab} needs to be anti-symmetric by Bose symmetry (given the form of (6.114) and (6.115)), and therefore consistent factorisation is not possible for a single scalar which in the boost-invariant limit is the well known fact that a single scalar cannot have a U(1) charge. For two scalars, a consistent boost-breaking amplitude is possible with $F_{ab} = \epsilon_{ab}F(E_1 + E_2)$, and similarly a consistent amplitude exists for a charged S = 1/2 particle. In Appendix 6.7.5 we provide a Lagrangian description of these boost-breaking versions of massless QED with unbroken U(1) gauge symmetry.

Scalar-photon couplings

We now provide a full analysis for a theory of a single scalar coupled to a photon. Many of the possible three-particle amplitudes have already been constrained and our goal in this part is to constrain the others where possible. There are five three-particle amplitudes arising from couplings between the scalar and the photon: $(\pm 1, \pm 1, 0)$, (-1, +1, 0) and $(\pm 1, 0, 0)$. However, we have already considered the $(\pm 1, 0, 0)$ amplitude above and we find that there are no solutions to (6.118) for a single scalar and therefore this amplitude must vanish. In addition, there are two amplitudes involving only the photon: $(\pm 1, \pm 1, \pm 1)$. Finally, there is a single amplitude involving only the scalar: (0, 0, 0).

Let's start by constraining the (-1, +1, 0) amplitude. Consider the four-particle amplitude $\mathcal{A}_4(1^{-1}2^{+1}3^{-1}4^{+1})$ between four photons. By little group scaling this amplitude takes the general form

$$\mathcal{A}_4(1^{-1}2^{+1}3^{-1}4^{+1}) = \langle 13 \rangle^2 [24]^2 \mathcal{G}(s, t, u, E_i).$$
(6.119)

²⁶We haven't shown that there are no other solutions.

Now in the *s*-channel we can exchange a scalar particle, meaning that this residue will have a vanishing mass dimension. This can also be arranged for in the *u*-channel by exchanging a scalar. If these residues are dimensionless, the four-particle amplitude has $\dim A_4 = -2$ which in turn requires $\dim \mathcal{G} = -6$ and so the amplitude must take the form

$$\mathcal{A}_4(1^{-1}2^{+1}3^{-1}4^{+1}) = \langle 13 \rangle^2 [24]^2 \frac{\mathcal{F}(E_i)}{stu}, \tag{6.120}$$

meaning that we require exchanges in *all* channels. In the *t*-channel we would need to exchange a graviton to realise the same mass dimension for the amplitude. However, even in the presence of a graviton the test cannot be passed, since the necessary $(\pm 1, \pm 1, \pm 2)$ amplitudes are forced to vanish by a different test, as we will show in section 6.4.3. Thus, the (-1, +1, 0) three-particle amplitude must vanish.

| Helicities | Amplitude \mathcal{A}_3 | Constraint |
|--------------|--|--------------------------------|
| (-1, -1, +1) | $\langle 12 \rangle^3 / (\langle 23 \rangle \langle 31 \rangle) F$ | F = 0 |
| (-1, -1, -1) | $\langle 12 \rangle \langle 23 \rangle \langle 31 \rangle F$ | alternating F in $(1, 2, 3)$ |
| (-1, -1, 0) | $\langle 12 \rangle^2 F$ | symmetric F in $(1, 2)$ |
| (-1, +1, 0) | $\langle 13 \rangle^2 / \langle 23 \rangle^2 F$ | F = 0 |
| (-1, 0, 0) | $(\langle 12 \rangle \langle 31 \rangle) / \langle 23 \rangle F$ | F = 0 |
| (0,0,0) | F | symmetric F in $(1, 2, 3)$ |

Table 6.1: Constrains on the three-particle amplitudes in a theory of a scalar coupled to a photon

We are therefore left with three distinct three-particle amplitudes and their parity counterparts. The others are forced to vanish. This is summarised in Table 6.1 and one can see that the non-zero amplitudes do not contain inverse powers of the brackets and therefore cannot give rise to spurious poles in four-particle amplitudes. For a theory of a single scalar coupled to a photon, there are therefore no further constraints from the four-particle test. The symmetry constraints on F tell us the minimum number of time derivatives required to write down a consistent boost-breaking interaction. As we discussed above, for the $(\pm 1, \pm 1, \pm 1)$ amplitudes we need at least three time derivatives. For the $(\pm 1, \pm 1, 0)$ and (0, 0, 0) vertices we need at least one and two respectively. The leading order Lagrangian giving rise to these amplitudes is (assuming parity-even interactions only)

$$\mathcal{L} = \frac{1}{2} (\partial \pi)^2 + \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + (a_1 \pi^3 + a_2 \pi^2 \ddot{\pi} + a_3 \dot{\pi}^3 + \dots)$$

$$+ (b_1 \pi + b_2 \dot{\pi} + b_3 \ddot{\pi} + \dots) F_{\mu\nu} F^{\mu\nu} + (c_1 \dot{F}^{\mu}{}_{\nu} \dot{F}^{\nu}{}_{\rho} F^{\rho}{}_{\mu} + \dots), \tag{6.121}$$

where a_i etc are dimensionful Wilson coefficients.

Brief summary

Let us briefly summarise our results for a spin-S particle coupled to a photon:

- Compton scattering is not possible for S > 1, while for S = 0, 1/2 consistent boost-breaking theories of massless scalar and fermionic QED with U(1) gauge symmetry exist. We can write down Lagrangians in each case with generalised boost-breaking gauge symmetries (see Appendix 6.7.5). Along the way we also showed that the absence of (-1, +1, ±1) vertices is robust against adding additional particles and that the (∓S, ±(S 2), ±1) and (∓S, ±(S + 2), ∓1) amplitudes must vanish for S ≠ 1.
- A consistent boost-breaking theory of a single scalar coupled to a photon does exist. Selfinteractions for both particles are possible and so are $\pi\gamma\gamma$ vertices. The leading Lagrangian is presented in (6.121).

6.4.3 Couplings to a graviton: gravitational Compton scattering and beyond

We now move onto couplings between spin-S particles and gravity. This section contains:

- constraints on the $(\pm 2, +S, -S)$ vertices due to gravitational Compton scattering
- a full analysis of all possible three-particle amplitudes in a theory of a single scalar coupled to gravity
- a full analysis of all possible three-particle amplitudes in a theory of a photon coupled to gravity
- an analysis for theory of a massless S = 3/2 particle coupled to gravity a.k.a $\mathcal{N} = 1$ supergravity.

Gravitational Compton scattering

We begin by constraining the leading, mass dimension 2, three-particle amplitudes for spin-S particles coupled to gravity, namely the $(\pm 2, +S, -S)$ amplitudes. We take $S \neq 2$. Consider the four-particle amplitude $\mathcal{A}_4(1^{-S}, 2^{+2}, 3^{-2}, 4^{+S})$ with dim $\mathcal{A}_4 = 2$. As with the photon case above, there are two ways to achieve the required dimension of the residues in s and t channels and a unique way in the *u*-channel. In the *s*-channel we can exchange a spin-*S* particle or a spin-|S - 4| particle. In the latter case we find spurious poles in the residue and so we set the $(\mp S, \pm 2, \pm (S - 4))$ amplitudes to zero for all $S \neq 2$. For spin-*S* exchange, we need the following three-particle amplitudes

$$\mathcal{A}_{3}(1^{-2}, 2^{-S}, 3^{+S}) = \frac{\langle 12 \rangle^{2S+2} \langle 31 \rangle^{2-2S}}{\langle 23 \rangle^{2}} F^{H}_{-2, -S, +S}(E_{1}, E_{2}), \tag{6.122}$$

$$\mathcal{A}_{3}(1^{+2}, 2^{+S}, 3^{-S}) = \frac{[12]^{2S+2}[31]^{2-2S}}{[23]^{2}} F^{AH}_{+2,+S,-S}(E_{1}, E_{2}).$$
(6.123)

Computing the residue we find (for both integer and half-integer S)

$$R_s = -\frac{\langle 13 \rangle^{2S} \langle 34 \rangle^{4-2S} [24]^4}{tu} F^{AH}_{+2,+S,-S}(E_2, -E_1 - E_2) F^{H}_{-2,-S,+S}(E_3, -E_3 - E_4).$$

The ordering of particles is especially important in the fermionic case, where changing the order of two fermions gives rise to a minus sign. Here and in the remaining equations we take particle 1 to always appear before particle 4.

In the *t*-channel, dimensional analysis allows for exchange of a spin-*S* particle and a spin-(S + 4) particle. However in the latter case spurious poles are unavoidable for all *S*. We therefore require the $(\mp S, \mp 2, \pm (S + 4))$ amplitudes to vanish. For spin-*S* exchange we find the residue (for both integer and half-integer *S*)

$$R_t = -\frac{\langle 13 \rangle^{2S} \langle 34 \rangle^{4-2S} [24]^4}{su} F^H_{-2,-S,+S}(E_3, E_1) F^{AH}_{+2,+S,-S}(E_2, E_4).$$
(6.124)

Finally, for *u*-channel exchange there is only a single choice for the spin of the exchanged particle that yields a residue with the desired mass dimension; that particle must be the graviton. The residue therefore depends on the lowest dimension three-graviton amplitude which in Section 6.4.1 we concluded must be reduced to the boost-invariant GR amplitude. However, now that we have included additional particles we must check if that result is robust against allowing for additional exchanges. Going back to the $\mathcal{A}_4(1^{-2}, 2^{+2}, 3^{-2}, 4^{+2})$ amplitude, we see that if the amplitude has dim $\mathcal{A}_4 = 2$ we can only exchange a graviton in the *s* and *u* channels, but in the *t*-channel dimensional analysis allows for S = 2 and S = 6 exchange. In the latter case, however, we find a spurious pole in *t* and so only graviton exchange can yield a consistent amplitude - consistency demands that the $(\pm 2, \pm 2, \mp 6)$ amplitudes are zero. Our result of 6.4.1, i.e. the $(+2, -2, \pm 2)$ amplitudes must be boost-invariant and correspond to those of GR, is robust against including additional massless particles.

We can now go back to gravitational Compton scattering. To compute the u-channel residue, we now need the lowest dimension three-graviton amplitudes. As shown above, these take the form

$$\mathcal{A}_3(1^{-2}, 2^{-2}, 3^{+2}) = \left(\frac{\langle 12 \rangle^3}{\langle 23 \rangle \langle 31 \rangle}\right)^2 \kappa, \tag{6.125}$$

$$\mathcal{A}_3(1^{+2}, 2^{+2}, 3^{-2}) = \left(\frac{[12]^3}{[23][31]}\right)^2 \kappa, \tag{6.126}$$

where κ is related to the Planck mass in GR and we have used the fact that GR is a parity-even theory. Now as we have seen a number of times before, there are two choices for the helicity configuration of the exchanged graviton. The total residue is a sum of the two, $R_u = R_u^{+-} + R_u^{-+}$, but one of these always vanishes once we declare how we approach the *u*-channel pole. We first consider the case of bosons, meaning we can swap the order of any two particles without introducing minus signs, but we keep factors of $(-1)^{2S}$ to make the formulae easy to generalise to the fermionic case. When $[14] = \langle 23 \rangle = 0$ we have $R_u^{+-} = 0$ and

$$R_u^{-+} = -\frac{\langle 13 \rangle^{2S} \langle 34 \rangle^{4-2S} [24]^4}{st} \kappa F_{-2,-S,+S}^H (-E_1 - E_4, E_1), \qquad (6.127)$$

and when $\langle 14 \rangle = [23] = 0$ we have $R_u^{-+} = 0$ and (for bosons)

$$R_u^{+-} = (-1)^{2S+1} \frac{\langle 13 \rangle^{2S} \langle 34 \rangle^{4-2S} [24]^4}{st} \kappa F_{+2,+S,-S}^{AH} (-E_1 - E_4, E_4).$$
(6.128)

If the spin-S particles are fermions, then the expression for R_u^{+-} inherits an overall minus sign (due to the necessity of swapping the order of particles 1 and 4), which conveniently cancels out the $(-1)^{2S}$ factor while R^{-+} is unchanged. The *u*-channel residue for both integer and half-integer S is therefore

$$R_u = -\frac{\langle 13 \rangle^{2S} \langle 34 \rangle^{4-2S} [24]^4}{st} \kappa F_{-2,-S,+S}^H(-E_1 - E_4, E_1), \tag{6.129}$$

subject to

$$F_{-2,-S,+S}^{H}(-E_1 - E_4, E_1) = F_{+2,+S,-S}^{AH}(-E_1 - E_4, E_4),$$
(6.130)

ensuring that the residue is the same regardless of how we approach the pole. This matching condition

ensures that operators generating the amplitudes (6.122) and (6.123) are parity-even.

Now we see from each residue that when 4 - 2S < 0, i.e. $S \ge 5/2$, a consistent four-particle amplitude cannot be constructed due to the additional poles in s. Hence we conclude that the above three-particle amplitudes for a massless particle with $S \ge 5/2$ coupled to gravity are inconsistent and must vanish. In a boost-invariant theory this is the well-known statement that a massless particle with $S \ge 5/2$ cannot couple to gravity, and we see that this statement is unchanged for boost-breaking theories. This is indeed consistent with some recent study in the light-cone formalism in which the only explicitly constructed cubic coupling of higher-spin particles to gravity is non-unitary [220].

For S < 5/2 we can construct a consistent amplitude for gravitational Compton scattering. It takes the form

$$\mathcal{A}_4(1^{-S}2^{+2}3^{-2}4^{+S}) = \langle 13 \rangle^{2S} \langle 34 \rangle^{4-2S} [24]^4 \frac{A}{stu}, \tag{6.131}$$

and consistent factorisation requires

$$-A = F_{+2,+S,-S}^{AH}(E_2, -E_1 - E_2)F_{-2,-S,+S}^H(E_3, -E_3 - E_4)$$

= $F_{-2,-S,+S}^H(E_3, E_1)F_{+2,+S,-S}^{AH}(E_2, E_4)$
= $\kappa F_{+2,+S,-S}^{AH}(-E_1 - E_4, E_4).$ (6.132)

The *F*-functions are related by (6.130) and therefore both can be written as the same *F*. If *F* contained any inverse powers of energies, then the singularities of the three expressions wouldn't match, so *F* must be a polynomial of a degree which we denote as *p*. The above equations then imply that 2p = 2p = p, and therefore p = 0. So only constant solutions are possible: the four-particle test has reduced the amplitudes to their boost-invariant limits! Furthermore, the coupling constants of the $(\pm 2, +S, -S)$ amplitudes are not arbitrary. The equations tell us that they are fixed in terms of the pure gravitational coupling κ : $F_{-2,-S,+S}^H = F_{+2,+S,-S}^{AH} = \kappa$. This is the on-shell derivation of the universality of gravity for elementary massless particles with $S \leq 2$: all particles couple to gravity with the same strength.

Compared to photon Compton scattering considered above, we see some important differences

for gravity. Here boost-breaking interactions are not permitted whereas for a photon coupled to S = 0, 1/2 particles such a breaking is permitted. Here we also see the emergence of the equivalence principle, and allowed couplings to S = 3/2 particles. We attribute these differences to the presence of a three-particle amplitude for three gravitons which does not exist for three photons. The case of a S = 3/2 particle coupled to gravity is particularly interesting. The amplitudes we have considered are those appearing in $\mathcal{N} = 1$ supergravity and here we have seen that boost-breaking versions, with relativistic on-shell conditions, do not exist. We refer the reader to [208] for some very nice results using factorisation when a massless S = 3/2 particle is in the spectrum. These results include: the necessity of gravity, the derivation of super-multiplets, and a proof that having $\mathcal{N} > 8$ requires the presence of a S = 5/2 particle and therefore the test cannot be passed if there is too much supersymmetry. Most of these results come from pole counting and we would therefore expect them to hold for boost-breaking theories with relativistic on-shell conditions too.

Scalar-graviton couplings

We now turn our attention to the boost-breaking theory of a single scalar coupled to gravity. Here we show that for relativistic on-shell conditions *the four-particle test requires all three-particle amplitudes for a scalar coupled to a graviton to be boost-invariant*. We have already seen that the pure graviton three-particle amplitudes are forced to be boost-invariant and so are the $(\pm 2, 0, 0)$ amplitudes. The remaining amplitudes to be discussed are $(\pm 2, \pm 2, 0)$, (+2, -2, 0) and (0, 0, 0).

First consider the (+2, -2, 0) amplitude, which we can easily show is inconsistent in both boost-invariant and boost-breaking theories. This vertex can contribute to *s*-channel exchange in the four-particle graviton amplitude $\mathcal{A}_4(1^{-2}, 2^{+2}, 3^{-2}, 4^{+2})$. This *s*-channel contribution to the amplitude has mass dimension -2 since the residue is dimensionless. However, the scaling of this amplitude under a little group transformation requires it to take the form

$$\mathcal{A}_4(1^{-2}, 2^{+2}, 3^{-2}, 4^{+2}) = \langle 13 \rangle^4 [24]^4 \mathcal{G}(s, t, u, E_i), \tag{6.133}$$

and so if dim $A_4 = -2$ the amplitude cannot be consistent, since simple poles require dim $\mathcal{G} \ge -6$, while dim $\{\langle 13 \rangle^4 [24]^4\} = 8$. We now constrain the $(\pm 2, \pm 2, 0)$ amplitudes using $\mathcal{A}_4(1^{+2}, 2^{+2}, 3^{-2}, 4^0)$ with scalar exchange in the *s*-channel. The contribution to the amplitude from this diagram has mass dimension²⁷ 4. The same mass dimension can be realised in the *t* and *u* channels by exchanging a graviton and using the leading (mass dimension 2) three-graviton amplitudes²⁸. Given that

$$\mathcal{A}_3(1^{+2}, 2^{+2}, 3^0) = [12]^4 F^{AH}_{+2,+2,0}(E_1, E_2), \tag{6.134}$$

the three residues are given by

$$R_s = -\frac{[12]^6 \langle 13 \rangle^2 \langle 23 \rangle^2}{tu} \kappa F^{AH}_{+2,+2,0}(E_1, E_2), \tag{6.135}$$

$$R_t = -\frac{[12]^6 \langle 13 \rangle^2 \langle 23 \rangle^2}{su} \kappa F_{+2,+2,0}^{AH}(E_2, -E_2 - E_4), \tag{6.136}$$

$$R_u = -\frac{[12]^6 \langle 13 \rangle^2 \langle 23 \rangle^2}{st} \kappa F_{+2,+2,0}^{AH}(E_1, -E_1 - E_4).$$
(6.137)

Here we have written $F_{+2,+2,0}^{AH}$ as a function of two energies only and it must be a symmetric function by Bose symmetry. Furthermore, we have used the fact that the (-2, 0, 0) amplitude is boost-invariant with its coupling identical to the graviton self-coupling κ . A consistent amplitude must therefore take the form

$$\mathcal{A}_4(1^{+2}, 2^{+2}, 3^{-2}, 4^0) = [12]^6 \langle 13 \rangle^2 \langle 23 \rangle^2 \frac{B}{stu}, \tag{6.138}$$

with

$$-B = \kappa F_{+2,+2,0}^{AH}(E_1, E_2) \tag{6.139}$$

$$=\kappa F_{+2,+2,0}^{AH}(E_2, -E_2 - E_4) \tag{6.140}$$

$$=\kappa F_{+2,+2,0}^{AH}(E_1, -E_1 - E_4), \tag{6.141}$$

which can only be solved if $F_{\pm 2,\pm 2,0}^{AH} = \text{const}$, thereby reducing the $(\pm 2, \pm 2, 0)$ amplitudes to their boost-invariant limits. Note that the coupling constant for these amplitudes is not fixed in terms of κ .

Finally, we can constrain the pure scalar amplitude (0,0,0) using the four-particle amplitude

²⁷This mass dimension can also be achieved by exchanging a spin-6 particle but such a residue contains spurious poles.

²⁸Another possibility is to exchange a spin-4 particle but in this case the residues again have spurious poles.

 $\mathcal{A}_4(1^0, 2^0, 3^0, 4^{+2})$. If we exchange a scalar in each channel with

$$\mathcal{A}_3(1^0, 2^0, 3^0) = F_{0,0,0}(E_1, E_2), \tag{6.142}$$

the three residues are

$$R_s = -\frac{[34]^2 [24]^2 \langle 23 \rangle^2}{tu} \kappa F_{0,0,0}(E_1, E_2), \tag{6.143}$$

$$R_t = -\frac{[34]^2 [24]^2 \langle 23 \rangle^2}{su} \kappa F_{0,0,0}(E_1, E_3), \tag{6.144}$$

$$R_u = -\frac{[34]^2 [24]^2 \langle 23 \rangle^2}{st} \kappa F_{0,0,0}(E_2, E_3), \tag{6.145}$$

and so the consistent amplitude is

$$\mathcal{A}_4(1^0, 2^0, 3^0, 4^{+2}) = [34]^2 [24]^2 \langle 23 \rangle^2 \frac{C}{stu}, \tag{6.146}$$

with

$$-C = \kappa F_{0,0,0}(E_1, E_2) \tag{6.147}$$

$$=\kappa F_{0,0,0}(E_1, E_3) \tag{6.148}$$

$$=\kappa F_{0,0,0}(E_2, E_3). \tag{6.149}$$

Again, the only solution to these factorisation constraints for generic energies is $F_{0,0,0} = \text{const}$, thereby reducing the three-scalar amplitude to its boost-invariant form, which is simply a constant.

We have therefore seen that *all* three-particle amplitudes, and therefore all three-point vertices, in a theory of a graviton coupled to a scalar (if the Lagrangian depends on covariant fields only) must reduce to their boost-invariant limits. Let us discuss the allowed boost-invariant interactions in more detail. We have discussed the pure gravity vertices at the level of a Lagrangian earlier on. The only allowed pure scalar amplitude is a constant and so the cubic vertex is simply ϕ^3 . The other two allowed interactions mix the scalar and the graviton and have mass dimension 2 and 4. The coupling of the former is the same as the three graviton coupling κ , while the coupling of the latter is independent of κ and is therefore a new Wilson coefficient in the effective action. At the level of a Lagrangian they come from the $(\partial \phi)^2 = g^{\mu\nu} \partial_{\mu} \phi \partial_{\nu} \phi$ and $\phi R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma}$ terms respectively, expanded around the boost-invariant vacuum $g_{\mu\nu} = \eta_{\mu\nu}$, $\phi = 0$. Note that there is no $\phi^2 R$ coupling as this can be removed by a field redefinition going from Jordan to Einstein frame. We can also write down a parity-odd vertex $\phi \epsilon^{\mu\nu\rho\sigma} R_{\mu\nu\kappa\lambda} R^{\kappa\lambda}{}_{\rho\sigma}$. In appendix 6.7.6, we provide further clarifications on why a simple $\dot{\phi}^3$ self-interaction for a scalar coupled to $h_{\mu\nu}$ in Minkowski space is inconsistent.

In [221] it was conjectured that in the flat space, decoupling and slow-roll limit of the EFT of inflation, if the scalar Goldstone has a boost-invariant kinetic term, then the only possible UV completion is a free theory. In this language, the decoupling limit boils down to neglecting all interactions with the metric fluctuations and the slow-roll limit corresponds to neglecting all Lorentz-invariant interactions, such as for example a potential $V(\phi)$. In other words, the conjecture is that any scalar EFT with $c_s = 1$ and boost-breaking interactions cannot be UV completed. The relation of this conjecture to our results is tantalizing but not straightforward. On the one hand, we also found that for $c_s = 1$ boost-breaking interactions are forbidden, but we crucially needed to assume (i) that the scalar is coupled to gravity, (ii) the theory is in Minkowski and (iii) assume a restricted form of the four-particle amplitude. Also, we did not use any constraints coming from a putative UV completion. All our analysis is based on the low-energy EFT. This is to be contrasted with the discussion in [221] where the coupling to gravity does not seem to play a role, while all the constraining power comes from demanding a consistent UV completion. Furthermore, the application of our results to the flat-space limit of FLRW spacetimes clashes with the IR sensitivity of the four-particle test. We will discuss this in Section 6.5.

Photon-graviton couplings

We have seen that when a scalar is coupled to $h_{\mu\nu}$, all three-particle amplitudes and therefore all three-point vertices are required to be boost-invariant by the four-particle test. One may therefore expect the presence of the graviton is forcing boost-invariance upon us when free particles satisfy relativistic on-shell conditions. Here we provide more evidence of this by showing that when a photon is coupled to $h_{\mu\nu}$, all three-point vertices involving this photon have to be boost-invariant. This result can be derived because of the existence of a (boost-invariant) three-point (++-) vertex for gravitons, which is absent for photons.

Let us recap the relevant results we have derived so far. We have shown that the pure graviton

three-particle amplitudes are boost-invariant. The lowest dimension photon amplitudes are forced to vanish by the test, while boost-breaking possibilities have not yet been ruled out for the (+ + +) and (- - -) three photon interaction. Now, for mixed amplitudes, we have four possibilities (plus their parity counterparts) left to consider:

$$(+2, +2, +1), (+2, +2, -1), (+1, +1, +2), (+1, +1, -2).$$
 (6.150)

First consider the dimensionless choice (+1, +1, -2). These amplitudes have both holomorphic and anti-holomorphic parts, and contribute to e.g. *u*-channel diagram for the $\mathcal{A}_4(1^{+1}, 2^{-1}, 3^{+2}, 4^{-2})$ amplitude via a photon exchange. The dimensionality of this amplitude is

$$\dim\{\mathcal{A}_4\} = 0 + 0 - 2 = -2. \tag{6.151}$$

On the other hand, to achieve correct helicity scalings, we need,

$$\mathcal{A}_4 \sim [13]^2 [23]^2 \langle 24 \rangle^4 \mathcal{G}(s, t, u, E_i),$$
 (6.152)

but then dim{ $\mathcal{G}(s, t, u, E_i)$ } = -10 < -6, which yields a contradiction. We therefore fail the test, which means these amplitudes must vanish. Note that this is the case for both boost-invariant and boost-breaking theories. In [208] it was argued that all dimensionless amplitudes, other than the pure scalar one, must vanish by virtue of the test. This result is based on pole counting so we expect those general results to be valid in our case too.

Now consider pure graviton scattering via the amplitude $A_4(1^{-2}2^{-2}3^{+2}4^{+2})$ which by the little group scaling takes the form

$$\mathcal{A}_4(1^{-2}2^{-2}3^{+2}4^{+2}) = \langle 12 \rangle^4 [34]^4 \mathcal{G}(s, t, u, E_i).$$
(6.153)

Now if we allow for a photon to be exchanged in the *s*-channel, the residue can have mass dimension 6 if we use the (+2, +2, -1) amplitudes and their parity counterparts. This contribution to the amplitude therefore has mass dimension 4 and by comparing to (6.153) we see that we need a *t* or *u* channel exchange to construct a consistently factorising amplitude. However, to achieve the required same mass dimension in either the *t* or *u* would require the exchange of a spin 3 particle with

non-zero $(+2, -2, \pm 3)$ amplitudes. But such amplitudes are not permitted²⁹. It is therefore impossible to achieve mass dimension 6 residues in the *t* and *u* channels of $\mathcal{A}_4(1^{-2}2^{-2}3^{+2}4^{+2})$ and so the $(\pm 2, \pm 2, \pm 1)$ amplitudes must vanish. This is the case for both boost-invariant and boost-breaking theories considered here.

Now consider the $\mathcal{A}_4(1^{+1}, 2^{+1}, 3^{+1}, 4^{-2})$ amplitude which we can use to constrain the (+1, +1, +1) interactions. The process is very similar to what we have seen a number of times. If we exchange a photon in the *s*-channel, we can construct a residue using the (+1, +1, +1) and (+1, -1, -2) amplitudes. The former has not yet been constrained beyond Bose symmetry, while the latter is required to be boost-invariant. By exchanging a photon in the other channels too we find a non-trivial factorisation constraint which fixes $F_{+1,+1,+1} = 0^{30}$. So in the presence of gravity, under the assumptions we made, all three-particle amplitudes involving three photons must vanish: there are no cubic self-interactions for a gravitationally coupled photon in a boost-breaking theory with $h_{\mu\nu}$ and A_{μ} fields, just as is the case for a boost-invariant one.

We have two more sets of amplitudes to constrain: (+1, +1, +2) and (+2, +2, +1) (and their parity counterparts). We find that both are forced to their boost-invariant limit using the four-particle test applied to $\mathcal{A}_4(1^{+1}, 2^{+1}, 3^{+2}, 4^{-2})$ and $\mathcal{A}_4(1^{+1}, 2^{+2}, 3^{+2}, 4^{-2})$ respectively. In both cases we include all possible exchanges allowed by dimensional analysis and find that any amplitudes involving higher spin (S > 2) particles are inconsistent. The coupling of (+1, +1, +2) corresponds to a new Wilson coefficient unrelated to the gravitational coupling κ . Meanwhile, the (+2, +2, +1) amplitudes are forced to vanish by Bose symmetry.

In conclusion, all three-particle amplitudes, in theories formulated in terms of covariant fields, are forced to their boost-invariant limits when we have a photon and a graviton in the spectrum. Pure photon vertices are constrained to vanish. The only allowed amplitudes that mix the photon and the graviton are $(+1, -1, \pm 2)$, $(\pm 1, \pm 1, \pm 2)$. At the level of a Lagrangian, the parity even operators are the Maxwell kinetic term $F^{\mu\nu}F_{\mu\nu} = g^{\mu\nu}g^{\rho\sigma}F_{\mu\rho}F_{\nu\sigma}$, and the non-minimal coupling

²⁹Indeed, if we allow for graviton exchange in the *s*-channel of the $\mathcal{A}_4(1^{-3}, 2^{+2}, 3^{-2}, 4^{+3})$, we see that the residue contains a $1/t^3$ piece and therefore the $(+2, -2, \pm 3)$ amplitudes are forced to vanish.

³⁰We could also exchange a S = 4 particle to find residues with the same mass dimension, but these additional exchanges lead to spurious poles.

term $F^{\mu\nu}F^{\rho\sigma}R_{\mu\nu\rho\sigma}$ expanded around the vacuum $g_{\mu\nu} = \eta_{\mu\nu}$, $A_{\mu} = 0$. Parity-odd amplitudes come from $\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}$ and $\epsilon^{\mu\nu\lambda\kappa}F_{\lambda\kappa}F^{\rho\sigma}R_{\mu\nu\rho\sigma}$.

Brief summary

- We have seen that massless particles with S ≥ 5/2 cannot couple to gravity under our assumptions, while particles with S < 5/2 can consistently couple to gravity, in which case the test yields universality of the gravitational couplings. No boost-breaking interactions are permitted. Along the way we also showed that allowing for additional particles does not change the fact that the lowest dimension vertices containing three gravitons must be boost-invariant and given by GR. We also saw that the (∓S, ±(S 4), ±2), (∓S, ±(S + 4), ∓2) amplitudes must vanish since for all S ≠ 2 they yield spurious poles in gravitational Compton scattering.
- We have perfomed a full analysis for the cases of a graviton coupled to a scalar or a photon. In each case we find that *all* three-point vertices, including the self-interactions of the scalar or photon are forced to their boost-invariant limits.

6.4.4 Multiple S = 1 particles

We now move on to considering multiple particles of the same spin. Consistent factorisation is trivial for multiple scalar particles since the three-particle amplitudes remain only functions of the energies and therefore products of these amplitudes cannot yield singularities. In this section we will focus on multiple S = 1 particles which we take to come in multiplets and therefore carry an additional colour index, a = 1, 2, ..., N. Our goal is to constrain the interactions between these particles in a boost-breaking theory formulated in terms of covariant fields. Recall that for a single particle (N = 1), the $(\pm 1, \pm 1, \pm 1)$ amplitudes are excluded by the four-particle test, whereas boost-breaking $(\pm 1, \pm 1, \pm 1)$ amplitudes are allowed (as long as gravity is decoupled).

The lowest mass dimension three-particle amplitudes are

$$\mathcal{A}_{3}(1_{a}^{+1}2_{b}^{+1}3_{c}^{-1}) = \frac{[12]^{3}}{[23][31]}F_{abc}^{AH}(E_{1}, E_{2}),$$
(6.154)

$$\mathcal{A}_{3}(1_{a}^{-1}2_{b}^{-1}3_{c}^{+1}) = \frac{\langle 12 \rangle^{3}}{\langle 23 \rangle \langle 31 \rangle} F^{H}_{abc}(E_{1}, E_{2}), \qquad (6.155)$$

where we have eliminated E_3 by energy conservation and have dropped the helicity subscripts on $F^{H/AH}$ in favour of the colour indices. The relationship between F^H and F^{AH} is³¹

$$F_{abc}^{H}(E_1, E_2) = \pm F_{abc}^{AH}(E_1, E_2), \tag{6.156}$$

with the -/+ sign corresponding to parity even/parity odd amplitudes respectively, by (6.52). In addition, Bose symmetry constrains the functions to satisfy

$$F_{abc}^{H}(E_1, E_2) = -F_{bac}^{H}(E_2, E_1),$$
(6.157)

$$F_{abc}^{AH}(E_1, E_2) = -F_{bac}^{AH}(E_2, E_1).$$
(6.158)

Now consider the amplitude $\mathcal{A}_4(1_a^{-1}2_b^{+1}3_c^{-1}4_d^{+1})$ with S = 1 exchange in each channel. If the amplitude has mass dimension 2, then there are two choices for the helicity in the *s* and *u* channels, and a unique choice for the *t*-channel. Remembering to take proper care of the ordering of indices and energies, we find the two residues to be

$$R_s^{+-} = \sum_e \frac{\langle 13 \rangle^2 [24]^2}{t} F_{bea}^{AH}(E_2, -E_1 - E_2) F_{ecd}^H(-E_3 - E_4, E_3), \tag{6.159}$$

$$R_s^{-+} = \sum_e \frac{\langle 13 \rangle^2 [24]^2}{t} F_{eab}^H(-E_1 - E_2, E_1) F_{dec}^{AH}(E_4, -E_3 - E_4), \tag{6.160}$$

summing over the exchanged particle colour e. Matching these two residues yields our first constraint on the three-particle amplitudes:

$$\sum_{e} F_{bea}^{AH}(E_2, -E_1 - E_2) F_{ecd}^H(-E_3 - E_4, E_3)$$

=
$$\sum_{e} F_{eab}^H(-E_1 - E_2, E_1) F_{dec}^{AH}(E_4, -E_3 - E_4).$$
 (6.161)

Next consider the u-channel. The two residues are

$$R_u^{+-} = -\sum_e \frac{\langle 13 \rangle^2 [24]^2}{s} F_{dea}^{AH}(E_4, -E_1 - E_4), F_{ecb}^H(-E_3 - E_2, E_3)$$
(6.162)

$$R_u^{-+} = -\sum_e \frac{\langle 13 \rangle^2 [24]^2}{s} F_{ead}^H(-E_1 - E_4, E_1) F_{bec}^{AH}(E_2, -E_2 - E_3), \qquad (6.163)$$

³¹We assume that the parity transformation commutes with the internal symmetry group, so that particle a is mapped to particle a under P.

and these are equivalent thanks to (6.161). Finally, the t-channel residue is

$$R_t = -\sum_e \frac{\langle 13 \rangle^2 [24]^2}{u} F_{ace}^H(E_1, E_3) F_{bde}^{AH}(E_2, E_4).$$
(6.164)

The full amplitude must therefore take the form

$$A_4(1_a^{-1}, 2_b^{+1}, 3_c^{-1}, 4_d^{+1}) = \langle 13 \rangle^2 [24]^2 \left(\frac{A_{abcd}}{st} + \frac{B_{abcd}}{su} + \frac{C_{abcd}}{tu} \right),$$
(6.165)

with consistent factorisation fixing

$$A_{abcd} - B_{abcd} = \sum_{e} F_{bea}^{AH}(E_2, -E_1 - E_2) F_{ecd}^{H}(-E_3 - E_4, E_3),$$

$$C_{abcd} - A_{abcd} = -\sum_{e} F_{ace}^{H}(E_1, E_3) F_{bde}^{AH}(E_2, E_4),$$

$$B_{abcd} - C_{abcd} = -\sum_{e} F_{dea}^{AH}(E_4, -E_1 - E_4) F_{ecb}^{H}(-E_3 - E_2, E_3).$$
 (6.166)

Taking the sum of these equations yields

$$\sum_{e} F_{bea}^{AH}(E_2, -E_1 - E_2) F_{ecd}^{H}(-E_3 - E_4, E_3) - \sum_{e} F_{ace}^{H}(E_1, E_3) F_{bde}^{AH}(E_2, E_4)$$
(6.167)

$$-\sum_{e} F_{dea}^{AH}(E_4, -E_1 - E_4) F_{ecb}^H(-E_2 - E_3, E_3) = 0,$$
(6.168)

which is our main factorisation constraint and must be satisfied with (6.161) subject to $E_1 + E_2 + E_3 + E_4 = 0$.

Now in the boost-invariant limit we have $F_{abc}^{H} = f_{abc} = \text{const}$, $F_{abc}^{AH} = \mp f_{abc} = \text{const}$. Under the assumption of complete antisymmetry of f_{abc} , matching the residues is trivial, but the primary factorisation constraint yields

$$\sum_{e} f_{abe} f_{ecd} + \sum_{e} f_{ace} f_{edb} + \sum_{e} f_{ade} f_{ebc} = 0.$$
(6.169)

The amplitudes in this case are those of Yang-Mills and we see that consistent factorisation of the four-particle amplitude forces the coupling constants to satisfy the familiar Jacobi identity. Note that

we have made no reference to an underlying Lie-algebra; this result follows from the basic physical principles of unitarity and locality.

Coming back to the boost-breaking case, the system of equations is very difficult to solve in general. To make progress, we make the assumption that $F_{abc}^{H} = f_{abc}F(E_1, E_2)$, $F_{abc}^{AH} = \mp f_{abc}F(E_1, E_2)$ with f_{abc} the usual couplings of Yang-Mills theory. Our three-particle amplitudes are therefore of the Yang-Mills form multiplied by a function of the energies. Bose symmetry requires these functions to be *symmetric* in the exchange of their two arguments, since f_{abc} are fully antisymmetric. Our factorisation constraint now becomes

$$\sum_{e} f_{bea} f_{ecd} F(E_2, -E_1 - E_2) F(-E_3 - E_4, E_3)$$

-
$$\sum_{e} f_{ace} f_{bde} F(E_1, E_3) F(E_2, E_4)$$

-
$$\sum_{e} f_{dea} f_{ecb} F(E_4, -E_1 - E_4) F(-E_2 - E_3, E_3) = 0.$$
 (6.170)

Now if we don't want to impose additional constraints on f_{abc} , consistent factorisation requires

$$F(E_1, E_3)F(E_4, E_2) \tag{6.171}$$

$$=F(E_2, -E_1 - E_2)F(-E_3 - E_4, E_3)$$
(6.172)

$$=F(E_4, -E_1 - E_4)F(E_3, -E_2 - E_3).$$
(6.173)

Upon using (6.161), we see that this constraint is exactly the same as the constraint on the graviton three-particle amplitude (6.105). As shown in Appendix 6.7.3, the only solution is F = const and therefore consistent factorisation requires the three-particle amplitudes to take their boost-invariant, Yang-Mills form. One may have expected the constraints for multiple S = 1 particles to be equivalent to a single S = 2 particle due to the kinematic-colour duality relating these amplitudes [219].

6.5 Mind the gap: amplitudes and the flat-space limit of cosmology

In this section, we discuss the connection of our results to cosmology. Instead of considering the most general scenario, for concreteness we focus on theories of a single scalar field minimally coupled to gravity, as they are both simple and relevant for models of inflation and dark energy. For so-called P(X)-theories, to be defined below, we will confirm our findings that in Minkowski all interactions

must be Lorentz invariant if we impose that the scalar speed of propagation c_s is the same as that of the graviton, c = 1, and require that the graviton be described in terms of a covariant Lagrangian (at least on the level of the free theory). Then, we consider the case in which the background is an FLRW spacetime with non-vanishing Hubble parameter, $H \neq 0$, and we study the sub-Hubble limit, i.e. we imagine performing a scattering experiment in a small laboratory of size $L \ll H^{-1}$, and describe the results in terms of flat-space amplitudes. Our main observation is that for arbitrarily small but non-vanishing H, it is always possible to find amplitudes that break boosts by any amount, within the validity of the Effective Field Theory (EFT), and no violations of unitarity or locality seem to arise. We argue that, despite the appearance, this observation does not imply any pesky physical discontinuity. Rather, we interpret this finding as the fact that *the constraining power of unitarity and locality through consistent factorization for massless theories is extremely fragile to IR modifications*. An analogous principle has already been established in Lorentz invariant contexts, where many interactions prohibited in flat space have consistent counterparts in AdS, regardless of the AdS radius - see [222] and references therein. Nonetheless, we decided to illuminate this issue further by discussing FLRW backgrounds which are more closely related to cosmology.

Sensitivity to IR modifications in cosmological scenarios is to be expected on the following grounds. Factorization happens when s, t or u go to zero and that's where all the constraining power of the four-particle test comes from. But this regime cannot be reached within the validity of the sub-Hubble limit. Indeed, for a flat-space approximation of FLRW spacetime to make sense, we need to require that the quantum uncertainty Δx on the spacetime position of the scattering particles is well within a Hubble volume $\Delta x \ll H^{-1}$. But then by the uncertainty principle

$$\Delta p \ge \frac{1}{2\Delta x} \gg H \quad \Rightarrow \quad \Delta s, \Delta t, \Delta u \gg H^2,$$
(6.174)

and therefore we always have an uncertainty in the Mandelstam variables of order H^2 . In FLRW spacetime, we cannot meaningfully distinguish, say, a pole at s = 0 from one at $s = H^2$. In more physical terms, as long as $H \neq 0$, we cannot experimentally reach the poles corresponding to massless on-shell particles while neglecting the expansion of the universe. Our finding that in the presence of an interacting spin-2 particle boost-breaking interactions cannot satisfy consistent factorization on s, t, u = 0, respectively, does not seem to matter in FLRW spacetime where this kinematic regime cannot be reached in the flat-space limit.

The suspicious reader might complain that our results suggest the presence of an unphysical discontinuity as $H \rightarrow 0$, but this is not the case. In the deep IR of the theory, a background with $H \neq 0$ is always very different from one with H = 0 because of the presence of a Hubble "horizon". So it is to be expected that any IR property of the theory for $H \rightarrow 0$ might be different from the corresponding one at H = 0. In other words, one cannot engineer a continuous series of physical thought experiments that give a discontinuous set of results and so there is no problem with our claims in this section.

Before proceeding, let's stress that there might be other obstructions to Lorentz breaking interactions when $c_s = 1$, which we don't capture in our analysis. For example, [221] conjectured that for the theory to have a local and unitary Lorentz invariant UV-completion, all Lorentz-breaking interactions for a single scalar with non-linear boosts must vanish as $c_s \rightarrow 1$. Also, recently [223] found some related obstructions considering perturbative unitarity in the sub-Hubble limit, where they showed that the window of validity of an EFT description for amplitudes shrinks to zero when $c_s \rightarrow 1$ in the presence of $\dot{\phi}^3$ interactions.

6.5.1 The absence of boost-breaking interactions in Minkowski

For concreteness, consider so-called P(X) theories minimally coupled to gravity with action

$$S = -\int d^4x \sqrt{-g} \left[\frac{M_{\rm P}^2}{2} R + P(X) \right], \qquad \qquad X \equiv \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi, \qquad (6.175)$$

which is a good toy model to study the spontaneous breaking of boosts while preserving time translations. The homogeneous equations of motion for the background $\phi(t)$ and the scale factor a(t) are

$$3M_{\rm P}^2 H^2 - 2XP_X + P = 0, \quad -M_{\rm P}^2 \dot{H} = XP_X, \quad \ddot{\phi} \left(P_X + 2XP_{XX} \right) + 3H\dot{\phi}P_X = 0.$$
(6.176)

The Lagrangian for perturbations $\varphi(t, \vec{x})$ is

$$= \frac{1}{2}(P_X + 2XP_{XX})\dot{\varphi}^2 - \frac{1}{2}P_X\partial_i\varphi\partial^i\varphi + \frac{1}{6}P_{XXX}\dot{\phi}^3\dot{\varphi}^3 + \dots, \qquad (6.177)$$

where the dots stand for higher derivatives of P(X) with respect of X, which will not be relevant for this discussion (they could be chosen to vanish if desired). The speed of sound is found to be

$$c_s^2 = \frac{P_X}{P_X + 2XP_{XX}} \,. \tag{6.178}$$

In this class of theories, it is only possible to have a well-defined solution in Minkowski spacetime with $c_s = 1$ if X = 0, in which case all interactions are Lorentz invariant. To see why, note that the following three assumptions cannot all be satisfied at the same time:

• Spontaneously broken boosts: This implies $X \neq 0$. From the equations of motion, setting H = 0 and $P_X = 0$ as appropriate for Minkowski, we get

$$\ddot{\phi}(P_X + 2XP_{XX}) = 0 \quad \Rightarrow \quad \ddot{\phi} = 0 \text{ or } P_X + 2XP_{XX} = 0.$$
 (6.179)

The second option is the cuscuton [224], which is non-dynamical and so not relevant for the present discussion. From $\ddot{\phi} = 0$ we deduce that X is constant, and so if it is non-vanishing it remains so for all times.

• Luminal propagation: This implies $c_s = 1$ and so

$$c_s^2 = \frac{P_X}{P_X + 2XP_{XX}} \stackrel{!}{=} 1 \quad \Rightarrow \quad P_X \neq 0 \& (P_{XX} = 0 \text{ or } X = 0).$$
 (6.180)

• Minkowski spacetime with dynamical gravity: This implies $g_{\mu\nu} = \eta_{\mu\nu}$ and so

$$\begin{cases} 3M_{\rm P}^2 H^2 = 2XP_X - P \stackrel{!}{=} 0\\ -M_{\rm P}^2 \dot{H} = XP_X \stackrel{!}{=} 0 \end{cases} \Rightarrow P = 0 \& (P_X = 0 \text{ or } X = 0). \tag{6.181}$$

Combining the above requirements we arrive at a contradiction: if we insist that $X \neq 0$, so that a Lorentz violation is in principle possible, then the luminality and Minkowski requirements are incompatible because the former leads to $P_X \neq 0$, while the latter entails $P_X = 0$. While we don't discuss it here in detail, the above result also applies to theories with higher derivatives. Intuitively, this stems from the fact that the higher derivative terms vanish when evaluated on the linearly time-dependent background we considered above.

This discussion confirms and complements our result that coupling to gravity (i.e. the covariant $h_{\mu\nu}$) in Minkowski enforces Lorentz invariance. On the one hand, our amplitude discussion is more general as it does not assume a P(X) Lagrangian. On the other hand, the above discussion generalized our findings in that it shows, for P(X) theories, that all *n*-particle amplitudes must be Lorentz invariant if the scalar propagates at the same speed as the graviton. In appendix 6.7.6, we provide further clarifications on why a simple $\dot{\phi}^3$ theory coupled to gravity is inconsistent in Minkowski space.

6.5.2 Boost-breaking interactions in the sub-Hubble limit

The attentive reader will have noticed that when $P_X = 0 = P_{XX}$, the speed of sound is ill defined, $c_s \stackrel{?}{=} 0/0$. In particular, the order of taking the limits matters: if we first impose Minkowski by setting $P_X = 0$, then $c_s = 0$ for any finite P_{XX} ; while if we first impose $c_s = 1$ by setting $P_{XX} = 0$, then we can take the Minkowski limit of FLRW, $P_X \rightarrow 0$, without changing the value of c_s . In this section, we discuss in detail this second possibility and find that in this case, Lorentz-breaking interactions are allowed within the regime of validity of the EFT. Let us now study how the Minkowski and $c_s^2 = 1$ solutions are *approached* from an FLRW solution.

Let us first assume the value \bar{X} of X(t) at some time is such that

$$P_{XX}(\bar{X}) = 0$$
 but $P_X(\bar{X}) \neq 0$. (6.182)

Expanding around it, we find

$$P_{XX}(X) = P_{XX}(\bar{X}) + (X - \bar{X})P_{XXX}(\bar{X}) + \mathcal{O}((X - \bar{X})^2), \qquad (6.183)$$

$$= (X - \bar{X})P_{XXX}(\bar{X}) + \mathcal{O}((X - \bar{X})^2).$$
(6.184)

The background equations of motion to zeroth order in $X - \bar{X}$ are

$$P_X(\bar{X})\left(\ddot{\phi} + 3H\dot{\phi}\right) + \mathcal{O}\left(X - \bar{X}\right) = 0, \qquad (6.185)$$

$$\dot{\phi}\left(\ddot{\phi}+3H\dot{\phi}\right)\simeq\mathcal{O}\left(X-\bar{X}\right)\,,$$
(6.186)

$$\dot{X} + 6HX \simeq \mathcal{O}\left(X - \bar{X}\right),$$
 (6.187)

and so are solved by $X \propto a^{-6}$. More usefully, for a small time interval $\Delta t \ll H^{-1}$, we can write

$$X = \bar{X} + \bar{X}\Delta t + \mathcal{O}((X - \bar{X})^2)$$
(6.188)

$$\Rightarrow \quad \frac{X - \bar{X}}{\bar{X}} \simeq -6H\Delta t + \mathcal{O}\left(\left(X - \bar{X}\right)^2\right). \tag{6.189}$$

So we find that, unlike in Minkowski where a constant X is always a solution, in FLRW we have to take into account that X evolves with time at some rate set by H.

Consider now the theory of perturbations in (6.177). Since X depends on time and we don't want to assume P(X) is just linear in X, which corresponds to the free theory, we cannot set $c_s^2 = 1$ at all times, but only at the time corresponding to $X = \overline{X}$ where P_{XX} happens to vanish. We can Taylor expand around $c_s - 1 \rightarrow 0$ and re-write c_s as

$$c_s^2 = \frac{P_X}{P_X + 2XP_{XX}} \tag{6.190}$$

$$=1-\frac{2XP_{XX}}{P_X}+\mathcal{O}\left(\left(\frac{2XP_{XX}}{P_X}\right)^2\right)$$
(6.191)

$$= 1 - \frac{2\bar{X}(X - \bar{X})P_{XXX}(\bar{X})}{P_X(\bar{X})} + \mathcal{O}\left(\left(X - \bar{X}\right)^2\right).$$
(6.192)

Using (6.189) for the time evolution of X, this becomes

$$1 - c_s^2 = -\frac{12H\Delta t\bar{X}^2 P_{XXX}(\bar{X})}{P_X(\bar{X})} + \mathcal{O}\left(\left(X - \bar{X}\right)^2\right).$$
(6.193)

Now we want to ask whether we can keep $1 - c_s^2$ arbitrary small while performing a subHubble scattering experiment in which some φ particles interact via the (spontaneously) boost-breaking coupling $\dot{\varphi}^3$ in the Lagrangian (6.177). We canonically normalize φ to φ_c and extract the cutoff scale Λ of the $\dot{\varphi}_c^3$ operator

$${}_{2} = \frac{1}{2} \left[\frac{1}{c_{s}^{2}} \dot{\varphi}_{c}^{2} - \frac{1}{2} \partial_{i} \varphi_{c} \partial^{i} \varphi_{c} \right] + \frac{\sqrt{2}}{3} \frac{X^{3} P_{XXX}}{(XP_{X})^{3/2}} \dot{\varphi}_{c}^{3}$$
(6.194)

$$\equiv \frac{1}{2} \left[\frac{1}{c_s^2} \dot{\varphi}_c^2 - \frac{1}{2} \partial_i \varphi_c \partial^i \varphi_c \right] + \frac{\dot{\varphi}_c^3}{\Lambda^2} \,. \tag{6.195}$$

Since we rescaled by P_X , which is time dependent, we also pick up additional terms proportional to

 $\partial_t P_X$, such as a mass term. We have neglected writing these terms because, around $X = \overline{X}$,

$$\partial_t P_X(X) = P_{XX}(X)\dot{X} \tag{6.196}$$

$$\simeq -6H\Delta t\bar{X}(X-\bar{X})P_{XXX}(\bar{X}) + \dots, \qquad (6.197)$$

$$\simeq 36 (H\Delta t)^2 \bar{X}^2 P_{XXX}(\bar{X}) + \dots,$$
 (6.198)

which is suppressed by at least two powers of $H\Delta t$. As long as we can neglect the expansion of the universe for some time $\Delta t \ll H^{-1}$, we can also neglect these additional terms.

Since P_{XXX} sets both the scale for the time evolution of $1 - c_s^2$ and the strength of the interaction we re-write

$$1 - c_s^2 = -\frac{36}{\sqrt{2}} \frac{H\Delta t \sqrt{XP_X}}{\Lambda^2} + \mathcal{O}\left(\left(X - \bar{X}\right)^2\right)$$
(6.199)

$$= -\frac{36}{\sqrt{2}} \left(\frac{E^2}{\Lambda^2}\right) \left(\frac{\sqrt{-\dot{H}M_{\rm P}^2}}{E^2}\right) (H\Delta t) + \mathcal{O}\left(\left(X - \bar{X}\right)^2\right), \qquad (6.200)$$

where we introduced the dummy factor E to represent the energy scale of the scattering process. For the scattering to happen effectively in flat space we need $E^2 \gg H^2$, $|\dot{H}|$. To resolve energies of order E while being able to neglect the expansion of the universe during the experiment, we need the experiment to last a time $H^{-1} \gg \Delta t \gg E^{-1}$. Finally, perturbativity requires $E \ll \Lambda$. Then

$$1 - c_s^2 \gg -\frac{36}{\sqrt{2}} \left(\frac{E}{\Lambda}\right)^2 \left(\frac{\sqrt{-\dot{H}}}{E}\right) \left(\frac{H}{E}\right) \left(\frac{M_{\rm P}}{E}\right) + \mathcal{O}\left(\left(X - \bar{X}\right)^2\right). \tag{6.201}$$

The first three factors must be much smaller than one while $M_{\rm P}/E$ must be much larger than one. Summarizing, we want the hierarchy of scales

$$H, \sqrt{-\dot{H}} \ll E \ll \Lambda \ll M_{\rm P}, \qquad (6.202)$$

while keeping $1 - c_s^2$ arbitrary small. This is always possible to achieve for any desired E/Λ , which parameterizes the strength of the cubic interaction), and Λ/M_P simply by taking H, $\sqrt{-\dot{H}}$ sufficiently small.

The upshot of this discussion is that we can find solutions for which a scattering experiment in a

small lab in an FLRW spacetime gives Lorentz-breaking amplitudes for massless particles that all move at the same speed to arbitrary but finite precision. Given the assumptions we have made about four-particle amplitudes, our results have shown that if this happened in Minkowski spacetime, there would be a violation of unitarity and/or locality for the amplitudes. But in FLRW those configurations cannot be reached while still neglecting corrections due to the expansion of the universe.

6.6 Discussion and conclusion

In this chapter we studied scattering amplitudes for massless, luminal, relativistic particles of any spin without demanding Lorentz invariance of the interactions. This is relevant for many systems that break Lorentz boosts spontaneously, as in cosmology or condensed matter physics. We focussed exclusively on on-shell particles and discussed (analytically continued) amplitudes without reference to unphysical structures such as gauge invariance or off-shell particles. The on-shell approach considerably simplifies the treatment of spinning particles, and our conclusions are independent of perturbative field redefinitions.

We systematically derived all possible massless three-particle amplitudes consistent with spacetime translations and rotations and constrained them using unitarity and causality via the requirement that four-particle amplitudes consistently factorize on simple poles into the product of two threeparticle amplitudes, a.k.a. the four-particle test [29]. We found that a large number of three-particle amplitudes fail the test and therefore cannot arise in any local, unitary perturbative theory around Minkowski spacetime. One result that stands out is that the existence of an interacting graviton, namely a massless spin-2 particle, enforces all cubic interactions involving particles coupled to it to be Lorentz invariant, including those interactions that do not involve the graviton. This is quite remarkable because, in the absence of a graviton, there could be infinitely many Lorentz-breaking interactions. As a concrete and simple example, consider the theory of a single scalar, for which we can write down infinitely many local interactions of the form $(\partial_t^{n_1}\phi)(\partial_t^{n_2}\phi)(\partial_t^{n_3}\phi)$ for any positive integers $n_{1,2,3}$. These interactions are not equivalent on-shell, generically giving different amplitudes, yet they are all allowed by the four-particle test. Our results show that in Minkowski, none of these Lorentz-breaking interactions can be consistently coupled to gravity! Although the form of the three-particle amplitudes that we have derived are completely general, in order to make progress we assumed that the helicity scaling of four-particle amplitudes are fixed by angle and square brackets rather than round ones. As we explained in Section 6.4, this amounts to assuming that the underlying Lagrangian is a function of Lorentz covariant fields with the breaking of boosts driven solely by time derivatives. It would be very interesting to work with a more general ansatz for the four-particle amplitudes such that we can constrain theories constructed out of SO(3) covariant fields.

Finally, we have discussed the relation of our analysis to cosmological models, in which spacetime can be approximated as flat only locally, but is never flat asymptotically. We found that, contrary to what happens in Minkowski, one can find models of a massless luminal scalar coupled to dynamical gravity in which sub-Hubble scattering is boost-breaking while no violations of unitarity and locality arise in the IR within the validity of the required approximations. We interpreted this as the observation that *the four-particle test is IR-sensitive* and the expansion of the universe provides an IR modification of the on-shell conditions. This finding mirrors the analogous findings for Lorentz invariant theories, where the four-particle test is not applicable if one deviates ever so slightly from asymptotically flat space [222].

One of our main motivations for studying boostless amplitudes was to use the results to constrain and perhaps fully bootstrap cosmological correlators when de Sitter boosts are not a symmetry of the theory. Our findings shows yet another reason why several clarifications need to be added to the simplistic slogan that the residue of the k_T pole of cosmological correlators is the Minkowski amplitude. In particular, we have shown that consistent factorization (Theorem 2.1) imposes severe constraints on Minkowski amplitudes, but these constraints don't necessary apply to the residue of the total-energy pole of correlators in (6.2). This issue will be discussed in detail elsewhere.

There are several ways in which our results could be extended.

• We used the consistent factorization of four-particle amplitudes to constrain three-particle amplitudes. It would be desirable to extend our analysis to higher *n*-particle amplitudes. For example, we expect that the coupling to a massless graviton will enforce *all* interactions to be Lorentz invariant. While the pedestrian methods we used in this chapter are probably ill-suited

to prove this more general result, one would probably want to harvest the power of on-shell recursion relations.

- It would be interesting to study how unitarity and locality constrains scattering experiments in the sub-Hubble limit of FLRW spacetime. This requires modification of the standard on-shell methods and an analysis will appear elsewhere.
- It would be interesting to extend our analysis to more general on-shell conditions where different particles can have different speeds, and to allow for a more general form of the four-particle amplitudes such that we capture the type of theories derived in [162].

6.7 Appendices

6.7.1 Spinor variables and discrete transformations

In this appendix we prove two important results for spinor representations of lightlike momenta, namely their transformation law under spatial reflection and the prescription for transforming the spinors so as to flip the sign of the exchanged particle's energy and momentum, which is necessary to compute the residues correctly.

Spatial reflection

Under the spatial reflection with respect to the origin, lightlike momentum p^{μ} tranforms as

$$(E, \mathbf{p}) \mapsto (E, -\mathbf{p}). \tag{6.203}$$

To the original momentum p^{μ} we associate a pair of spinors $(\lambda^{\alpha}, \tilde{\lambda}^{\dot{\alpha}})$. One choice is

$$\lambda = \left(\sqrt{p^0 + p^3}, \frac{p^1 + ip^2}{\sqrt{p^0 + p^3}}\right)^T, \ \tilde{\lambda} = \left(\sqrt{p^0 + p^3}, \frac{p^1 - ip^2}{\sqrt{p^0 + p^3}}\right).$$
(6.204)

Spinor helicity variables corresponding to the new momentum must be of the form

$$\lambda'_{\alpha} = a\epsilon^{\dot{\beta}}_{\alpha}\tilde{\lambda}_{\dot{\beta}}, \ \tilde{\lambda}'_{\dot{\alpha}} = a^{-1}\epsilon^{\beta}_{\dot{\alpha}}, \lambda_{\beta}$$
(6.205)

i.e.

$$\lambda' = a(\tilde{\lambda}_2, -\tilde{\lambda}_1)^T, \ \tilde{\lambda}' = a^{-1}(\lambda_2, -\lambda_1).$$
(6.206)

It is easy to check that these new variables do indeed give $p'^{\mu} = (E, -\mathbf{p})$. Now we must fix the coefficient a. To do this, we have to take a look at polarization tensors.

Consider an exchange diagram with an exchanged particle of spin-1. Suppose at the left-hand side vertex, there is an outgoing particle of helicity +1 (equivalent to an incoming antiparticle of helicity -1). Then the same particle (with helicity +1) is incoming at the right-hand side vertex. The +1 polarization vector ξ^+ of the exchanged particle is mapped to $P\xi^+$ under spatial reflection P. But we also require, for consistency, that it be mapped to the -1 polarization vector of the particle with reversed momentum. The spatial reflection of ξ^+ is, in terms of spinor variables,

$$P\xi^{+}_{\alpha\dot{\alpha}}(\mathbf{p}) = \frac{\epsilon^{\dot{\beta}}_{\alpha}\epsilon^{\ \beta}_{\dot{\alpha}}\mu_{\beta}\tilde{\lambda}_{\dot{\beta}}}{\langle\mu,\lambda\rangle},\tag{6.207}$$

where we used (6.205), and μ is a reference spinor. Now, the -1 polarization vector relative to $-\mathbf{p}$ momentum is

$$\xi_{\alpha\dot{\alpha}}^{-}(-\mathbf{p}) = \frac{\lambda_{\alpha}^{\prime}\tilde{\zeta}_{\dot{\alpha}}^{\prime}}{[\tilde{\lambda}^{\prime},\tilde{\zeta}^{\prime}]} = -a^{2}\frac{\epsilon_{\alpha}^{\beta}\epsilon_{\dot{\alpha}}^{\beta}\zeta_{\beta}\tilde{\lambda}_{\dot{\beta}}}{\langle\zeta,\lambda\rangle}.$$
(6.208)

Setting $\zeta = \mu$ and comparing the two expressions, we conclude that $a^2 = -1$, i.e. $a = \pm i$. Thus, the prescription for mapping $(E, \mathbf{p}) \mapsto (E, -\mathbf{p})$ is (for example),

$$\lambda' = (-i\tilde{\lambda}_2, i\tilde{\lambda}_1)^T, \quad \tilde{\lambda}' = (i\lambda_2, -i\lambda_1).$$
(6.209)

Under spatial reflection, the two inner products then transform as, e.g,

$$[12] \quad \mapsto \quad [1'2'] = \langle 21 \rangle = -\langle 12 \rangle, \tag{6.210}$$

$$\langle 12 \rangle \mapsto \langle 1'2' \rangle = [21] = -[12].$$
 (6.211)

This transformation law leads to consistent results for various 3p amplitudes - see, for example, Appendix 6.7.4.

The $p_I \mapsto -p_I$ prescription

Consider again a diagram in which a particle with helicity +1 is being exchanged. Let's transform this diagram under TP. Then the polarization 4-vector of the intermediate particle flips its sign: $\xi^{\mu} \mapsto -\xi^{\mu}$. On the other hand, this new 4-vector must be precisely the +1 polarization vector relative to $-p_I$ (helicity of the exchanged particle doesn't change under TP). Schematically, the ± 1 polarization vector is proportional to $(\tilde{\lambda}/\lambda)^{\pm 1}$. Thus, if $p_I \leftrightarrow (\lambda, \tilde{\lambda})$, then we must have $-p_I \leftrightarrow (\lambda, -\tilde{\lambda})$ (or $(-\lambda, \tilde{\lambda})$) to give consistent polarization vectors. We extrapolate this conclusion to spins other than 1. This convention produces the correct relative signs in the amplitudes - see, for example, the discussion in Section 6.4.3.

6.7.2 Formulas for the framid amplitude

Here we list the functions we used in (6.72) to write down the framid exchange four-particle amplitude $\mathcal{A}_4(1^0, 2^+, 3^0, 4^-)$:

$$F_{(1,a)}(E_1, E_2, E_3, E_4; s, t) = -4e_4 E_{12}^2 - 2s E_{12} E_{23} f + s^2 (E_1 - E_2)(E_3 - E_4) - \frac{s^2}{t} g,$$
(6.212)

$$F_{(1,b)}(E_1, E_2, E_3, E_4; s, t) = 12e_4E_{12} + \frac{2s}{t}(E_2 - E_4)g + 3sE_{24}f,$$
(6.213)

$$F_{(1,c)}(E_1, E_2, E_3, E_4; s, t) = -9e_4 + \frac{4E_2E_4}{t}g,$$
(6.214)

$$F_{(2,a)}(E_1, E_2, E_3, E_4; s, t) = 4e_4(E_1^2 + E_1E_3 + E_3^2) + t^2f - st(E_1 - E_3)(E_2 - E_4) + sE_1E_3(E_1 - E_3)(E_2 - E_4)$$
(6.215)
$$- tE_1E_3\left(-E_{13}^2 + E_2E_3\left(1 + \frac{2E_4}{E_1}\right) + E_1E_4\left(1 + \frac{2E_2}{E_3}\right)\right),$$
$$F_{(2,b)}(E_1, E_2, E_3, E_4; s, t) = 2(E_1 - E_3)(E_2^2 + E_2E_4 + E_4^2)(t - E_1E_3),$$
(6.216)

$$F_{(2,b)}(E_1, E_2, E_3, E_4; s, t) = 2(E_1 - E_3)(E_2^2 + E_2 E_4 + E_4^2)(t - E_1 E_3),$$
(6.21)

where we used

$$f = E_1 E_4 + E_2 E_3, (6.217)$$

$$g = 4e_4 + \frac{1}{2}E_1E_3(2E_1 + E_2)(2E_3 + E_4) + sf, \qquad (6.218)$$

$$E_{ij} = E_i + E_j, agenum{6.219}{6.219}$$

$$e_4 = E_1 E_2 E_3 E_4. (6.220)$$

For completeness, we also list all on-shell, three-particle amplitudes for the framid, in the case of equal speeds $c_L = c_T$. We find

$$A_3(1^+2^+3^+) = 0, (6.221)$$

$$A_3(1^+2^+3^-) = \sqrt{2}g(E_1 - E_2) \frac{[12]^3}{[23][31]}, \qquad (6.222)$$

$$A_3(1^+2^+3^0) = g[12]^2, (6.223)$$

$$A_3(1^+2^-3^0) = \frac{1}{2}g(21)^2, (6.224)$$

$$A_3(1^+2^03^0) = -\frac{1}{\sqrt{2}}g\left(E_1 + 2E_2\right)\frac{[12][31]}{[23]}, \qquad (6.225)$$

$$A_3(1^0 2^0 3^0) = 2g \left(E_1 E_2 + E_2 E_3 + E_3 E_1 \right).$$
(6.226)

where

$$g = \frac{c_L^2 - 1}{c_L^2 M_1}.$$
(6.227)

6.7.3 Solutions to constraints on $F(E_i)$

In this appendix we provide proofs that the only rational functions of the form

$$F(x,y) = \frac{f(x,y)}{x^n y^m (x+y)^k}$$
(6.228)

that solve (6.97) and (6.105) are F = 0 and F =const respectively.

Photon constraint

We begin with the constraint (6.97). We allow F to take the form (6.228) and we have already shown that the antisymmetry in the first two arguments of F requires n = m. Thus $F(x, y) = (xy)^{-m}(x+y)^{-k}f(x,y)$, where the function f must be alternating in its two variables. We therefore write f(x,y) = (x-y)P[x+y,xy] where P is another polynomial. Our factorisation constraint (6.97) is then

$$0 = (-1)^{k} \frac{(E_{1} - E_{3})(E_{2} - E_{4})}{E_{1}^{m} E_{2}^{m} E_{4}^{m}(E_{1} + E_{3})^{2k}} P[E_{1} + E_{3}, E_{1}E_{3}]P[E_{2} + E_{4}, E_{2}E_{4}]$$

$$+ (-1)^{m} \frac{(E_{1} + 2E_{2})(2E_{3} + E_{4})}{E_{1}^{k} E_{2}^{m} E_{4}^{k}(E_{1} + E_{2})^{2m}} P[-E_{1}, -E_{2}(E_{1} + E_{2})]P[-E_{4}, -E_{3}(E_{3} + E_{4})]$$

$$- (-1)^{m} \frac{(E_{1} + 2E_{4})(E_{2} + 2E_{3})}{E_{1}^{k} E_{2}^{k} E_{4}^{m}(E_{1} + E_{4})^{2m}} P[-E_{1}, -E_{4}(E_{1} + E_{4})]P[-E_{2}, -E_{3}(E_{2} + E_{3})].$$
(6.229)
First, we are going to assume that P is non-zero and, by examining the singularities, deduce that m = k = 0. By assumption, P[x, y] is not divisible by x or y, so the first term in (6.229) is singular at $E_1 + E_3 = 0$ for k > 0 while neither the second nor the third term are singular there. Thus, we must have k = 0. By a similar argument, we also have m = 0. Our main equation thus simplifies to

$$(E_1 - E_3)(E_2 - E_4)P[E_1 + E_3, E_1E_3]P[E_2 + E_4, E_2E_4]$$

+(E_1 + 2E_2)(2E_3 + E_4)P[-E_1, -E_2(E_1 + E_2)]P[-E_4, -E_3(E_3 + E_4)]
-(E_1 + 2E_4)(E_2 + 2E_3)P[-E_1, -E_4(E_1 + E_4)]P[-E_2, -E_3(E_2 + E_3)] = 0, \quad (6.230)

and this equation must be satisfied for all energies subject to $E_1 + E_2 + E_3 + E_4 = 0$. Now we will aim to show

$$\left(P\left[x, \frac{(3 \cdot 2^{n+1} - 2)}{(3 \cdot 2^{n+1} - 1)^2}x^2\right] = 0 \quad \forall x \quad \text{OR} \quad P\left[x, 3 \cdot 2^n \cdot \frac{3 \cdot 2^n - 1}{(3 \cdot 2^{n+1} - 1)^2}x^2\right] = 0 \quad \forall x\right) \forall n \in \mathbb{Z}_{\geq 0},$$
(6.231)

which entails $P \equiv 0$. The reason for this is that P would have to satisfy infinitely many distinct constraints of the form $P[x, a_k x^2] = 0 \quad \forall x$ (it is easy to check that a_k are indeed distinct) and thus we would need $(a_k x^2 - y) \mid P[x, y]$ for all the a_k , which is impossible if P is a nonzero polynomial.

To prove (6.231), let

$$E_1^{(n)} = (3 \cdot 2^{n+1} - 2)x,$$

$$E_2^{(n)} = -(3 \cdot 2^n)x,$$

$$E_3^{(n)} = x,$$

$$E_4^{(n)} = -(3 \cdot 2^n - 1)x,$$
(6.232)

for n = 0, 1, 2, ... Note that $E_1 = -2E_4$ for any n, in which case the third term in (6.230) vanishes and the main equation becomes

$$(E_3 + 2E_4)P[E_3 - 2E_4, -2E_3E_4]P[E_2 + E_4, E_2E_4]$$

= 2(2E_3 + E_4)P[2E_4, -E_2(E_2 - 2E_4)]P[-E_4, -E_3(E_3 + E_4)]. (6.233)

Taking n = 0, we get

$$-3xP[5x, 4x^2]P[-5x, 6x^2] = 0, (6.234)$$

so

$$P[5x, 4x^2] = 0 \quad \forall x \quad \text{OR} \quad P[-5x, 6x^2] = 0 \quad \forall x,$$
 (6.235)

or equivalently,

$$P[x, \frac{4}{25}x^2] = 0 \quad \forall x \quad \text{OR} \quad P[-x, \frac{6}{25}x^2] = 0 \quad \forall x,$$
 (6.236)

which is precisely the condition from (6.231) for n = 0. Now we will prove (6.231) for any n > 0 by induction. Suppose (6.231) is true for some n - 1. Then set E_i to the values specified in (6.232). We get

$$(3 - 3 \cdot 2^{n+1})xP[(3 \cdot 2^{n+1} - 1)x, (3 \cdot 2^{n+1} - 2)x^2]P[-(3 \cdot 2^{n+1} - 1)x, 3 \cdot 2^n(3 \cdot 2^n - 1)x^2] = 2(3 - 3 \cdot 2^n)xP[(3 \cdot 2^n - 1)x, (3 \cdot 2^n - 2)x^2]P[-(3 \cdot 2^{n+1} - 2)x, 3 \cdot 2^n \cdot (3 \cdot 2^n - 2)x^2].$$

$$(6.237)$$

The right hand side is zero by virtue of the previous induction step. Thus, the left hand side is also zero, which entails

$$P[x, \frac{(3 \cdot 2^{n+1} - 2)}{(3 \cdot 2^{n+1} - 1)^2} x^2] = 0 \quad \forall x \quad \text{OR} \quad P[x, 3 \cdot 2^n \cdot \frac{3 \cdot 2^n - 1}{(3 \cdot 2^{n+1} - 1)^2} x^2] = 0 \quad \forall x, \quad (6.238)$$

thereby completing the proof. This proves that there are no consistent $(+1, -1 \pm 1)$ amplitudes under the assumption made in (6.69) and discussed in that section.

Graviton constraint

We now show that the only solution to the system of equations³² (6.105) is F = const thereby reducing the $(+2, -2, \pm 2)$ amplitudes to their boost-invariant limits.

Here F must be of the form

$$F(x,y) = \frac{f(x,y)}{x^m y^m (x+y)^k}$$
(6.239)

 $^{^{32}}$ In fact, we need only 2 equations - those relating the second, third and fifth expression in (6.105) - and we can drop the condition that the residue must be the same regardless of how the pole is approached.

where f is a symmetric polynomial, so f(x, y) = P[x + y, xy] for some polynomial P. Thus (6.105) takes the form

$$\frac{(-1)^{k+m}}{E_2^m E_3^m E_4^m (E_1 + E_3)^{2k}} P[E_1 + E_3, E_1 E_3] P[E_2 + E_4, E_2 E_4]
= \frac{1}{E_2^k E_3^k E_4^m (E_1 + E_2)^{2m}} P[-E_2, -E_1(E_1 + E_2)] P[-E_3, -E_4(E_3 + E_4)]
= \frac{1}{E_2^m E_3^k E_4^k (E_1 + E_4)^{2m}} P[-E_4, -E_1(E_1 + E_4)] P[-E_3, -E_2(E_2 + E_3)].$$
(6.240)

As in the case of the photon, we see that singularities generally don't match. If k > 0, then the first line contains a singularity at $E_1 + E_3 = 0$ which does not appear in the other two expressions. If m > 0, then the second line has a singularity at $E_1 + E_2 = 0$ which does not correspond to the behaviour of the other two functions. Thus, we must have m = k = 0 and the equations become

$$P[E_1 + E_3, E_1E_3]P[E_2 + E_4, E_2E_4]$$

= $P[-E_2, -E_1(E_1 + E_2)]P[-E_3, -E_4(E_3 + E_4)]$
= $P[-E_4, -E_1(E_1 + E_4)]P[-E_3, -E_2(E_2 + E_3)].$ (6.241)

This must hold for any E_i that satisfy $\sum_i E_i = 0$. Now if we let $E_1 = E_2 = 0$, $E_3 = -E_4 = E$, our constraint becomes

$$P[E,0]P[-E,0] = P[0,0]P[-E,0] = P[E,0]P[-E,0],$$
(6.242)

and so P[-E, 0](P[E, 0] - P[0, 0]) = 0. This implies that P[-E, 0] = 0 for all E or P[E, 0] = P[0, 0] for all E. But the first alternative entails the latter, so we can just assume

$$P[E,0] = P[0,0] := P_0 \quad \forall E.$$
(6.243)

Now let $E_1 + E_2 = E_3 + E_4 = 0$. Our factorisation constraint is then

$$P[E_1 + E_3, E_1 E_3] P[-(E_1 + E_3), E_1 E_3]$$
(6.244)

$$=P[E_1, 0]P[-E_3, 0] \tag{6.245}$$

$$=P[E_3, -E_1(E_1 - E_3)]P[-E_3, -E_1(E_1 - E_3)].$$
(6.246)

Because E_1 and E_3 are effectively independent variables, we can write $x = E_1 + E_3$, $y = E_1E_3$ and find that the following equation must hold for all x, y:

$$P[x,y]P[-x,y] = P_0^2.$$
(6.247)

It is then easy to show (e.g. by observing that any zero of P[x, y] would correspond to a singularity of P[-x, y], which a polynomial cannot have) that the only polynomial solution to this equation is $P[x, y] = P_0$.

6.7.4 Tree level three-point amplitudes for broken Maxwell theory

Maxwell theory of electromagnetism is a Lorentz invariant theory of a massless spin-1 particle, with just two degrees of freedom corresponding to the two helicities ± 1 of the photon. The quadratic Lagrangian is

$$\mathcal{L}_2 = \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{6.248}$$

where $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$. By counting first class and second class constraints, one can show that the free theory indeed has two degrees of freedom. This is because A_0 is non-dynamical and we also have a one-dimensional gauge freedom. In the boost-invariant theory, there are no cubic interactions, as we have shown in Section 6.3. Interactions can only start at quartic order in the fields.

As for the boost-breaking amplitudes in a theory of a single photon, we have shown that they are allowed: they are the $(\pm 1, \pm 1, \pm 1)$ amplitudes with at least three powers of energy as dictated by Bose symmetry. The simplest such amplitudes are

$$\mathcal{A}_3(1^{-1}2^{-1}3^{-1}) = g\langle 12\rangle\langle 23\rangle\langle 31\rangle(E_1 - E_2)(E_2 - E_3)(E_1 - E_3), \tag{6.249}$$

$$\mathcal{A}_3(1^{+1}2^{+1}3^{+1}) = \pm g[12][23][31](E_1 - E_2)(E_2 - E_3)(E_1 - E_3), \tag{6.250}$$

and in Section 6.4.1 we suggested that such amplitudes arise from

$$\ddot{F}^{\mu}{}_{\nu}\dot{F}^{\nu}{}_{\rho}F^{\rho}{}_{\mu}, \qquad \epsilon^{\mu\nu\rho\sigma}\ddot{F}_{\mu\nu}\dot{F}_{\rho\kappa}F_{\sigma}{}^{\kappa}, \tag{6.251}$$

operators in the Lagrangian. In this Appendix we consider the second of these operators showing that

it does indeed give rise to the parity-odd form of the above amplitudes. Extending the following to the first of these operators is straightforward and yields the parity-even form of the above amplitudes.

We will use the following, elegant identity:

$$\epsilon^{\mu\nu\rho\sigma} p^{1}_{\mu} p^{2}_{\nu} p^{3}_{\rho} p^{4}_{\sigma} = -4i \left(\langle 12 \rangle [23] \langle 34 \rangle [41] - [12] \langle 23 \rangle [34] \langle 41 \rangle \right), \tag{6.252}$$

which is valid for any four, null 4-momenta (not necessarily conserved). The identity can be proven efficiently using symbolic manipulation in Mathematica. The tree-level, (+1, +1, +1), S-matrix element $S^+_{3\to 0}$ due to $\epsilon^{\mu\nu\rho\sigma}\ddot{F}_{\mu\nu}\dot{F}_{\rho\kappa}F_{\sigma}{}^{\kappa}$ is

$$\begin{split} S_{3\to0}^{+} &= \langle 0|(-i) \int d^{3}x dt H_{int}(x,t) \left[\prod_{i=1}^{3} \sqrt{2E_{i}} a_{\mathbf{p}_{i}^{+\dagger}}^{+\dagger} \right] |0\rangle \\ &= ig' \int d^{3}q_{1} d^{3}q_{2} d^{3}q_{3} \delta^{(4)} \left(\sum q_{i}^{\mu} \right) \\ &\times \sum_{\Lambda_{1,2,3}} \epsilon_{\mu\nu\rho\sigma} E_{q1}^{2} \left(q_{1}^{\mu} \xi_{1}^{\Lambda_{1},\nu} - q_{1}^{\nu} \xi_{1}^{\Lambda_{1},\mu} \right) E_{q2} \left(q_{2}^{\rho} \xi_{2}^{\Lambda_{2},\alpha} - q_{2}^{\alpha} \xi_{2}^{\Lambda_{2},\rho} \right) \left(q_{3}^{\sigma} \xi_{3,\alpha}^{\Lambda_{3}} - q_{3,\alpha} \xi_{3}^{\Lambda_{3},\sigma} \right) \\ &\times \sum_{\sigma \in S_{3}} \left(\delta(p_{\sigma(1)} - q_{1}) \delta(p_{\sigma(2)} - q_{2}) \delta(p_{\sigma(3)} - q_{3}) \delta_{+,\Lambda_{1}} \delta_{+,\Lambda_{2}} \delta_{+,\Lambda_{3}} \right) \\ &= ig' \delta^{(4)} \left(\sum p_{i}^{\mu} \right) \epsilon_{\mu\nu\rho\sigma} E_{1}^{2} E_{2} \\ &\times \left(p_{1}^{\mu} \xi_{1}^{+,\nu} - p_{1}^{\nu} \xi_{1}^{+,\mu} \right) \left(p_{2}^{\rho} \xi_{2}^{+,\alpha} - p_{2}^{\alpha} \xi_{2}^{+,\rho} \right) \left(p_{3}^{\sigma} \xi_{3,\alpha}^{+} - p_{3,\alpha} \xi_{3}^{+,\sigma} \right) + 5 \text{ perms.} \\ &= 2ig' \delta^{(4)} \left(\sum p_{i}^{\mu} \right) \epsilon_{\mu\nu\rho\sigma} E_{1}^{2} E_{2} p_{1}^{\mu} \xi_{1}^{+,\nu} \left(p_{2}^{\rho} \xi_{2}^{+,\alpha} - p_{2}^{\alpha} \xi_{2}^{+,\rho} \right) \left(p_{3}^{\sigma} \xi_{3,\alpha}^{+} - p_{3,\alpha} \xi_{3}^{+,\sigma} \right) + 5 \text{ perms.} \end{split}$$

Once we expand the product of two brackets into a sum, each permutation seems to include four terms, but one of these trivially vanishes as it involves a factor $p_2 \cdot p_3 = 0$. We therefore have

$$S_{3\to0}^{+} = 2ig'\delta^{(4)}\left(\sum p_i^{\mu}\right)\epsilon_{\mu\nu\rho\sigma}E_1^2E_2p_1^{\mu}\xi_1^{+,\nu} \\ \times \left(p_2^{\rho}p_3^{\sigma}(\xi_2^+\cdot\xi_3^+) - \xi_2^{+,\rho}p_3^{\sigma}(p_2\cdot\xi_3^+) - p_2^{\rho}\xi_3^{+,\sigma}(p_3\cdot\xi_2^+)\right) + 5 \text{ perms.}$$
(6.253)

Using (6.252), we get

$$S_{3\to0}^{+} = 2ig'\delta^{(4)} \left(\sum p_i^{\mu}\right) (-4i)E_1^2 E_2 \left\{ \left(\langle 1\xi_1 \rangle [\xi_1 2] \langle 23 \rangle [31] - [1\xi_1] \langle \xi_1 2 \rangle [23] \langle 31 \rangle \right) \left(\xi_2^+ \cdot \xi_3^+ \right) - \left(\langle 1\xi_1 \rangle [\xi_1 \xi_2] \langle \xi_2 3 \rangle [31] - [1\xi_1] \langle \xi_1 \xi_2 \rangle [\xi_2 3] \langle 31 \rangle \right) \left(p_2 \cdot \xi_3^+ \right) - \left(\langle 1\xi_1 \rangle [\xi_1 2] \langle 2\xi_3 \rangle [\xi_3 1] - [1\xi_1] \langle \xi_1 2 \rangle [2\xi_3] \langle \xi_3 1 \rangle \right) \left(p_3 \cdot \xi_2^+ \right) \right\} + 5 \text{ perms.} \quad (6.254)$$

(Spinors constructed from the momenta are written as numbers 1, 2, 3; spinors constructed from the polarization vectors are written as ξ_i .) Recall that for three-particle, on-shell interactions, we have $\langle ij \rangle = 0$ for all i, j or [ij] = 0 for all i, j; so the first line vanishes. We also have $[1\xi_1] = 0$, so all terms involving this factor vanish as well. Thus,

$$S_{3\to0} = -8g'\delta^{(4)} \left(\sum p_i^{\mu}\right) E_1^2 E_2 \left\{ \left(\langle 1\xi_1 \rangle [\xi_1\xi_2] \langle \xi_2 3 \rangle [31] \right) \left(p_2 \cdot \xi_3^+ \right) + \left(\langle 1\xi_1 \rangle [\xi_1 2] \langle 2\xi_3 \rangle [\xi_3 1] \right) \left(p_3 \cdot \xi_2^+ \right) \right\} + 5 \text{ perms.}$$
(6.255)

To make further progress, we have to choose a concrete spinor representation of the polarization vectors ξ_i . Recall that

$$\xi_{a\dot{a}}^{+}(\mathbf{p}) = \frac{\eta_a \lambda_{\dot{a}}}{\langle \eta, \lambda \rangle},$$

with an almost arbitrary reference spinor η . At this point, we are free to make a choice that breaks the Lorentz symmetry and we do so such that

$$\xi_i^+ = \frac{(\epsilon . \tilde{\lambda}_i^T) \tilde{\lambda}_i}{(ii)}.$$
(6.256)

So $\eta_{i,1} = \tilde{\lambda}_{i,2}$ and $\eta_{i,2} = -\tilde{\lambda}_{i,1}$. Then, we have the following identities:

$$\langle i\xi_j^+ \rangle = -(ij), \tag{6.257}$$

$$[i\xi_{j}^{+}] = \frac{[ij]}{(jj)}, (6.258)$$

$$[\xi_i^+, \xi_j^+] = \frac{[ij]}{(ii)(jj)}, \tag{6.259}$$

$$p_i \cdot \xi_j^+ = \frac{1}{2} \langle i\xi_j \rangle [i\xi_j] = -\frac{1}{2} \frac{(ij)[ij]}{(jj)}.$$
(6.260)

Now we can simplify (6.255). The first line (dropping the prefactor $-8g'\delta$) gives:

$$\begin{split} &\sum_{\text{perms}} E_1^2 \left(E_2 \langle 1\xi_1 \rangle [\xi_1 \xi_2] \langle \xi_2 3 \rangle [31] (p_2 \cdot \xi_3^+) \right) \\ &= \frac{1}{8E_1 E_2 E_3} \sum_{\text{perms}} E_1^2 \left(E_2 (-(11)) [12] (32) [31] \left(-\frac{1}{2} (23) [23] \right) \right) \\ &= \frac{1}{8E_1 E_2 E_3} \sum_{\text{perms}} E_1^3 E_2 [12] (32) [31] (23) [23] \\ &= \frac{1}{2} \sum_{\text{perms}} E_1^2 E_2 [12] [23] [31] = \frac{1}{2} [12] [23] [31] \sum_{\text{cyc}} E_1^2 (E_2 - E_3) \,. \end{split}$$

Meanwhile, the second line of (6.255) (again dropping the prefactor $-8g'\delta$) gives:

$$\begin{split} &\sum_{\text{perms}} E_1^2 \left(E_2 \langle 1\xi_1 \rangle [\xi_1 2] \langle 2\xi_3 \rangle [\xi_3 1] (p_3 \cdot \xi_2^+) \right) \\ &= \frac{1}{2} \frac{1}{8E_1 E_2 E_3} \sum_{\text{perms}} E_1^2 \left(E_2 (-(11)) [12] (-(23)) [31] (-(32)) [32] \right) \\ &= \frac{1}{8E_1 E_2 E_3} \sum_{\text{perms}} E_1^3 E_2 [12] (23) [31] (32) [23] \\ &= \frac{1}{2} \sum_{\text{perms}} E_1^2 E_2 [12] [23] [31] = \frac{1}{2} [12] [23] [31] \sum_{\text{cyc}} E_1^2 \left(E_2 - E_3 \right). \end{split}$$

We see that the two contributions are exactly the same. In conclusion, we get

$$S_{3\to0}^{+} = -8g'\delta^{(4)}\left(\sum p_{i}^{\mu}\right)[12][23][31]\sum_{cyc}E_{1}^{2}\left(E_{2}-E_{3}\right)$$
$$= 8g'\delta^{(4)}\left(\sum p_{i}^{\mu}\right)[12][23][31](E_{1}-E_{2})(E_{2}-E_{3})(E_{3}-E_{1}).$$
(6.261)

The analogue of (6.255) for all-minus helicities is

$$S_{3\to0}^{-} = 8g'\delta^{(4)} \left(\sum p_i^{\mu}\right) E_1^2 E_2 \left\{ \left([1\xi_1] \langle \xi_1 \xi_2 \rangle [\xi_2 3] \langle 31 \rangle \right) (p_2 \cdot \xi_3^{-}) + \left([1\xi_1] \langle \xi_1 2 \rangle [2\xi_3] \langle \xi_3 1 \rangle \right) (p_3 \cdot \xi_2^{-}) \right\} + 5 \text{ perm-s.}$$
(6.262)

We choose reference spinors similarly as before,

$$\xi_i^- = \frac{\lambda_i(\epsilon, \lambda_i^T)}{(ii)}.$$
(6.263)

Then

$$\langle i\xi_j^- \rangle = \langle ij \rangle,$$
 (6.264)

$$[i\xi_j^-] = -\frac{(ji)}{(jj)}, (6.265)$$

$$\langle \xi_i^-, \xi_j^- \rangle = \langle ij \rangle, \tag{6.266}$$

$$p_i \cdot \xi_j^- = \frac{1}{2} \langle i\xi_j \rangle [i\xi_j] = -\frac{1}{2} \frac{\langle ij \rangle (ji)}{(jj)}.$$
(6.267)

The first line of (6.262), after dropping the prefactor $8g'\delta$, gives

$$\sum_{\text{perms}} E_1^2 \left(E_2[1\xi_1] \langle \xi_1 \xi_2 \rangle [\xi_2 3] \langle 31 \rangle (p_2 \cdot \xi_3^-) \right)$$

$$= \frac{1}{2} \sum_{\text{perms}} E_1^2 \left(E_2 \cdot (-1) \cdot \langle 12 \rangle \left(-\frac{(23)}{(22)} \right) \langle 31 \rangle \frac{-\langle 23 \rangle (32)}{(33)} \right)$$

$$= \frac{1}{2} \sum_{\text{perms}} \frac{E_1^2 E_2}{4 E_2 E_3} \langle 12 \rangle \langle 23 \rangle \langle 31 \rangle (-(23)) (32) = -\frac{1}{2} \langle 12 \rangle \langle 23 \rangle \langle 31 \rangle \sum_{cyc} E_1^2 (E_2 - E_3).$$

The second line of (6.262) yields

$$\sum_{\text{perms}} E_1^2 \left(E_2[1\xi_1] \langle \xi_1 2 \rangle [2\xi_3] \langle \xi_3 1 \rangle (p_3 \cdot \xi_2^-) \right) = \frac{1}{2} \sum_{\text{perms}} E_1^2 \left(E_2 \cdot (-1) \cdot \langle 12 \rangle \frac{(32)}{(33)} \langle 31 \rangle \frac{-\langle 32 \rangle (23)}{(22)} \right)$$
$$= -\frac{1}{2} \sum_{\text{perms}} \frac{E_1^2 E_2}{4 E_2 E_3} \langle 12 \rangle \langle 23 \rangle \langle 31 \rangle (23) (32) = -\frac{1}{2} \langle 12 \rangle \langle 23 \rangle \langle 31 \rangle \sum_{\text{cyc}} E_1^2 (E_2 - E_3).$$

So

$$S_{3\to0}^{-} = -8g'\delta^{(4)} \left(\sum p_i^{\mu}\right) \langle 12\rangle \langle 23\rangle \langle 31\rangle \sum_{cyc} E_1^2 (E_2 - E_3) \\ = 8g'\delta^{(4)} \left(\sum p_i^{\mu}\right) \langle 12\rangle \langle 23\rangle \langle 31\rangle (E_1 - E_2)(E_2 - E_3)(E_3 - E_1).$$
(6.268)

Comparing (6.261) and (6.268) with (6.52), we see that the amplitude due to $\epsilon^{\mu\nu\rho\sigma}\ddot{F}_{\mu\nu}\dot{F}_{\rho\alpha}F_{\sigma}^{\alpha}$ is parity-odd, as expected from the presence of the ϵ tensor.

6.7.5 Boost-breaking massless QED

In this Appendix we provide Lagrangians for the boost-breaking versions of massless QED we derived using the four-particle test in Section 6.4.2. In the boost-invariant limit massless scalar QED is described by the Lagrangian

$$\mathcal{L} = \frac{1}{4}F_{\mu\nu}^2 + \frac{1}{2}D^{\mu}\phi D_{\mu}\phi^*$$
(6.269)

where the covariant derivative is as usual $D_{\mu}\phi = \partial_{\mu}\phi - ie\phi A_{\mu}$. This gives rise to the standard kinetic terms plus cubic and quartic vertices. The Lagrangian is invariant under the gauge symmetry

$$\phi \to e^{ie\alpha(x)}\phi, \qquad A_{\mu} \to A_{\mu} + \partial_{\mu}\alpha(x).$$
 (6.270)

By choosing the basis $\phi = \phi^1 + i\phi^2$ the anti-symmetric nature of the cubic vertices is manifest and the three-particle amplitude has $F_{ab} = \epsilon_{ab}$ in (6.114) and (6.115). Now to realise the function of energy in the amplitude we need to add time derivatives to (6.269). We saw that in the boost-breaking case we have $F_{ab} = \epsilon_{ab}F(E_1 + E_2)$ and since $E_1 + E_2 = -E_3$ we can add time derivatives to the vector only, and we find that the correct Lagrangian is given by

$$\mathcal{L} = \frac{1}{4} F_{\mu\nu}^2 + \frac{1}{2} \hat{D}^{\mu} \phi \hat{D}_{\mu} \phi^*$$
(6.271)

where we have defined the new boost-breaking covariant derivative

$$\hat{D}_{\mu}\phi = \partial_{\mu}\phi - ie\phi\hat{\partial}_{t}A_{\mu}, \qquad (6.272)$$

in terms of the derivative operator

$$\hat{\partial}_t = a_1 \partial_t + a_2 \partial_t^2 + a_3 \partial_t^3 + \dots$$
(6.273)

In comparison to the boost-invariant theory, this theory also has a gauge symmetry given by

$$\phi \to e^{ie\hat{\partial}_t\beta(x)}\phi, \qquad A_\mu \to A_\mu + \partial_\mu\beta(x).$$
 (6.274)

If we again write $\phi = \phi^1 + i\phi^2$ we see that

$$\mathcal{L} \supset ie\epsilon_{ab}\phi^a\partial^\mu\phi^b\hat{\partial}_t A_\mu,\tag{6.275}$$

and these cubic vertices give rise to our three-particle amplitudes. We therefore have a consistent boost-breaking theory of massless scalar QED.

For S = 1/2 the story is a simple generalisation of the above discussion. In the boost-invariant limit, massless fermionic QED is described by the Lagrangian

$$\mathcal{L} = \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i \bar{\psi} \gamma^{\mu} D_{\mu} \psi, \qquad (6.276)$$

where ψ is a four-component Dirac spinor³³, γ^{μ} are the gamma matrices and $D_{\mu} = \partial_{\mu} + ieA_{\mu}$. This Lagrangian is invariant under the U(1) gauge symmetry

$$\psi \to e^{-ie\alpha(x)}\psi, \qquad A_{\mu} \to A_{\mu} + \partial_{\mu}\alpha(x).$$
 (6.277)

Guided by the scalar case, we can instead define a new covariant derivative as

$$\hat{D}_{\mu} = \partial_{\mu} + ie\hat{\partial}_t A_{\mu}, \tag{6.278}$$

and if we replace D_{μ} by \hat{D}_{μ} in (6.276) then we find a consistent boost-breaking theory of massless fermionic QED invariant under the gauge symmetry

$$\psi \to e^{-ie\partial_t \beta(x)}\psi, \qquad A_\mu \to A_\mu + \partial_\mu \beta(x).$$
 (6.279)

Again this theory gives rise to our boost-breaking amplitudes derived in Section 6.4.2.

6.7.6 More details on the inconsistency of $\dot{\phi}^3$ coupled to gravity

In this appendix we consider a self-interacting scalar minimally coupled to $h_{\mu\nu}$ in Minkowski space and directly compute the $\mathcal{A}_4(1^0, 2^0, 3^0, 4^{+2})$ amplitude due to scalar exchange, showing that the final result is gauge invariant only in the absence of Lorentz-violating interactions. Thus the aim is to provide further clarity on why an interaction of the form $\dot{\phi}^3$ is inconsistent. We take the graviton self-interactions and the minimal coupling between the scalar and the graviton to be Poincaré invariant and consider a Lagrangian of the form

$$\mathcal{L} = \mathcal{L}_{EH} + \frac{1}{2} (\partial \phi)^2 - \frac{1}{\sqrt{2}M_{\rm pl}} h^{\mu\nu} \partial_\mu \phi \partial_\nu \phi + \mathcal{L}_\phi, \qquad (6.280)$$

³³Recall that a Dirac spinor is not a irreducible respresentation of the Lorentz group. It is really comprised of two 2-component spinors reflecting the fact that we need two S = 1/2 particles each with $\pm 1/2$ helicities.

where \mathcal{L}_{EH} contains the quadratic and cubic terms in the canonically normalised graviton fluctuation $h_{\mu\nu}$ arising from expanding $\sqrt{-g}R$ around Minkowski space and \mathcal{L}_{ϕ} contains cubic self-interactions for the scalar with an unspecified number of time derivatives (all Lorentzian derivatives can be removed by field redefinitions). The results of this appendix will therefore capture $\dot{\phi}^3$ but also a more general class of self-interactions where the on-shell three-scalar amplitude is $\mathcal{A}_3(1^0, 2^0, 3^0) = F(E_1, E_2, E_3)$ where F is a symmetric polynomial.

First consider the s-channel of the $\mathcal{A}_4(1^0, 2^0, 3^0, 4^{+2})$ amplitude. Up to unimportant $\mathcal{O}(1)$ factors and inverse powers of $M_{\rm pl}$, we have

$$\mathcal{A}_{4}^{s}(1^{0}, 2^{0}, 3^{0}, 4^{+2}) = \frac{F(E_{1}, E_{2})}{s} \epsilon^{+}_{\mu\nu}(p_{4}) p_{3}^{\mu}(p_{3}^{\nu} + p_{4}^{\nu}) = \frac{F(E_{1}, E_{2})}{s} \epsilon^{+}_{\mu\nu}(p_{4}) p_{3}^{\mu} p_{3}^{\nu}, \qquad (6.281)$$

where we have used the fact that the graviton's on-shell polarisation tensor is transverse and have used energy conservation to eliminate the energy of the exchanged scalar particle. The t and u channel expressions are

$$\mathcal{A}_{4}^{s}(1^{0}, 2^{0}, 3^{0}, 4^{+2}) = \frac{F(E_{1}, E_{3})}{t} \epsilon^{+}_{\mu\nu}(p_{4}) p_{2}^{\mu} p_{2}^{\nu}, \qquad (6.282)$$

$$\mathcal{A}_{4}^{s}(1^{0}, 2^{0}, 3^{0}, 4^{+2}) = \frac{F(E_{2}, E_{3})}{u} \epsilon_{\mu\nu}^{+}(p_{4}) p_{1}^{\mu} p_{1}^{\nu}.$$
(6.283)

Now we can write these expressions in the spinor helicity formalism using

$$4\epsilon^{+}_{\mu\nu}(p_4)p^{\mu}_i p^{\mu}_i = e^{+}_{\alpha\dot{\alpha}}(p_4)e^{+}_{\beta\dot{\beta}}(p_4)\lambda^{\alpha}_i\tilde{\lambda}^{\dot{\alpha}}_i\lambda^{\beta}_i\tilde{\alpha}^{\dot{\beta}}_i = \left(\frac{\langle\eta i\rangle[4i]}{\langle\eta4\rangle}\right)^2.$$
(6.284)

Now we have infinitely many choices for the reference spinor η , but it is sufficient to consider only three options, $\eta = 1, 2, 3$, so that η corresponds to a spinor of one of the particles other than the graviton. The three choices for each channel yield (again dropping unimportant common factors)

$$A_{4}^{s}(1^{0}, 2^{0}, 3^{0}, 4^{+2}) = F_{12} \left(\langle 12 \rangle [14] [24] \right)^{2} \times \begin{cases} \frac{1}{su^{2}} & \eta = 1 \\ \frac{1}{st^{2}} & \eta = 2 \\ 0 & \eta = 3 \end{cases}$$
(6.285)

$$A_{4}^{t}(1^{0}, 2^{0}, 3^{0}, 4^{+2}) = F_{13} \left(\langle 12 \rangle [14] [24] \right)^{2} \times \begin{cases} \frac{1}{tu^{2}} & \eta = 1 \\ 0 & \eta = 2 \\ \frac{1}{ts^{2}} & \eta = 3 \end{cases}$$
(6.286)
$$A_{4}^{u}(1^{0}, 2^{0}, 3^{0}, 4^{+2}) = F_{23} \left(\langle 12 \rangle [14] [24] \right)^{2} \times \begin{cases} 0 & \eta = 1 \\ \frac{1}{ut^{2}} & \eta = 2 \\ \frac{1}{us^{2}} & \eta = 3 \end{cases}$$
(6.287)

where we have introduced the shorthand $F(E_i, E_j) = F_{ij}$. Using s + t + u = 0, we can therefore write the full amplitude as

$$A_4(1^0, 2^0, 3^0, 4^{+2}) = -\left(\langle 12 \rangle [14] [24]\right)^2 \times \begin{cases} \frac{F_{12}}{stu} + \frac{F_{12} - F_{13}}{tu^2} & \eta = 1\\ \frac{F_{23}}{stu} + \frac{F_{23} - F_{12}}{st^2} & \eta = 2\\ \frac{F_{13}}{stu} + \frac{F_{13} - F_{23}}{us^2} & \eta = 3. \end{cases}$$
(6.288)

For general boost-breaking scalar self-interactions, $F_{12} \neq F_{13}$ and so on. Hence we see that the above amplitude could change as different choices for the unphysical reference spinor are made. This certainly indicates an inconsistency. Demanding that the amplitude is the same for each choice of reference spinor leads to the constraints

$$F_{12} = F_{13} = F_{23} \,. \tag{6.289}$$

This is only solved by F = constant for generic energies, and so the three-particle amplitude for a scalar coupled to gravity must be Poincaré invariant.

Chapter 7

Bootstrap via BCFW Momentum Shifts

7.1 The Britto-Cachazo-Feng-Witten (BCFW) method

As we have seen above, in the consistent factorization approach we have to write down an ansatz that is guided by the helicity scaling and the expected mass dimension of the four-particle amplitude. While this process can be generalized and applied to other cases, in this subsection I will describe an alternative method, reviewed in [32, 83], which automates the process of bootstrapping three-particle amplitudes and constructing higher-point ones. This method also relies on a purely on-shell description and is known, after the names of its authors [159, 160], as the Britto-Cachazo-Feng-Witten (BCFW) recursion.

In the simplest implementation of the BCFW recursion, one considers a deformation of two spinors $\lambda^{(i)}$ and $\tilde{\lambda}^{(j)}$, promoting them both to functions of a complex variable z:

$$\lambda^{(i)}(z) = \lambda^{(i)} + z\lambda^{(j)}, \qquad (7.1)$$

$$\tilde{\lambda}^{(j)}(z) = \tilde{\lambda}^{(j)} - z\tilde{\lambda}^{(i)}.$$
(7.2)

The deformed spinors induce a change on the amplitude:

$$\mathcal{A} \mapsto A^{(i,j)}(z) \,. \tag{7.3}$$

The deformation is consistent with the on-shell kinematics and with the conservation of momentum,

$$\delta\left(\sum_{a} p_{a,\alpha\dot{\alpha}}\right) = z\lambda_{\alpha}^{(j)}\tilde{\lambda}_{\dot{\alpha}}^{(i)} - z\lambda_{\alpha}^{(j)}\tilde{\lambda}_{\dot{\alpha}}^{(i)} = 0, \qquad (7.4)$$

so the amplitude $A^{(i,j)}(z)$ is still a valid physical amplitude. We will now restrict attention to the tree-level amplitude, which is a rational function of the kinematic variables so that $A^{(i,j)}(z)$ is a meromorphic function with no branch cuts. We can use Cauchy's Theorem to write

$$A(0) = \operatorname{Res}\left[\frac{A(z)}{z}, z = 0\right] = -\sum_{I} \operatorname{Res}\left[\frac{A(z)}{z}, z = z_{I}\right] - \sum_{I} \operatorname{Res}\left[\frac{A(z)}{z}, z = \infty\right], \quad (7.5)$$

where we can define

$$-\operatorname{Res}\left[\frac{\mathcal{A}(z)}{z}, z=\infty\right] = B_{\infty}.$$
(7.6)

Note that $B_{\infty} = 0$ if $A(z) \to 0$ as $|z| \to \infty$.

Each of the residues at finite $z_I \neq 0$ corresponds to a physical pole of the amplitude A(z), i.e. a factorization channel. Let us determine the location of these residues. Consider a single exchange channel such that cutting the exchange leg would separate a diagram into a subset I of the set of external particles and its complement I^c . The process can thus be written as $(p_a)_{a \in I} \rightarrow (p_a)_{a \notin I}$. If $P_I(z)$ is the momentum of the exchanged virtual particle, then

$$P_I(z) = \sum_{a \in I} p_a(z). \tag{7.7}$$

By definition, the exchanged particle is on-shell when $z = z_I$,

$$P_I^2(z_I) = 0. (7.8)$$

If the deformation is applied to $\lambda^{(i)}$ and $\tilde{\lambda}^{(j)}$ (we denote this deformation as (i, j)), there are two distinct cases to consider:

- If particles i and j are on the same side of the exchange diagram (i, j ∈ I or i, j ∉ I), then P_I(z) is independent of z and P²_I(z_I) = 0 has no solutions (or else it is trivial, but we can arrange the external kinematics in such a way as to eliminate this possibility).
- If particles i and j are on opposite sides of the exchange diagram (i ∈ I, j ∉ I or i ∉ I, j ∈ I), then P_I²(z_I) = 0 will have exactly one solution. Let's focus on the case i ∈ I, j ∉ I and let q^μ

be the four-vector with spinors $(\lambda^{(j)}, \tilde{\lambda}^{(i)})$. We have

$$P_I^2(z) = (P_I(0) + zq)^2 = P_I^2(0) + 2z \left(\sum_{k \in I} p_k\right)^{\mu} q_{\mu} = P_I^2(0) + z \left(\sum_{k \in I} \langle kj \rangle [ki]\right).$$
(7.9)

Hence

$$z_I = -\frac{P_I^2(0)}{\sum\limits_{k \in I} \langle kj \rangle [ki]} \,. \tag{7.10}$$

We can now compute the residues on the right-hand side of (7.5). The residue of A(z)/z at $z = z_I$ can be obtained by noting that

$$\lim_{z \to z_I} P_I^2(z) A(z) = A_L(z_I) A_R(z_I) , \qquad (7.11)$$

where A_L, A_R are the constituent sub-amplitudes. Hence

$$\operatorname{Res}\left[\frac{A(z)}{z}, z = z_{I}\right] = \operatorname{Res}\left[\frac{A_{L}(z)A_{R}(z)}{zP_{I}^{2}(z)}, z = z_{I}\right] = \frac{A_{L}(z_{I})A_{R}(z_{I})}{z_{I}}\operatorname{Res}\left[\frac{1}{P_{I}^{2}(z)}, z = z_{I}\right]$$
$$= \frac{A_{L}(z_{I})A_{R}(z_{I})}{z_{I}}\left(\frac{dP_{I}^{2}(z_{I})}{dz_{I}}\right)^{-1} = \frac{A_{L}(z_{I})A_{R}(z_{I})}{z_{I}\sum_{k\in I}\langle kj\rangle[ki]} = -\frac{A_{L}(z_{I})A_{R}(z_{I})}{P_{I}^{2}(0)}.$$
(7.12)

Therefore, we obtain a recursion relation for the amplitude,

$$A^{(i,j)}(0) = \sum_{I} \frac{A_L(z_I)A_R(z_I)}{P_I^2} + B_{\infty}$$
(7.13)

where z_I are given by (7.10). This relation can be used to construct tree-level, n-particle amplitudes from basic building blocks, such as the three-particle amplitudes [29, 159, 160]. In this brief review, I will focus on the simplest case n = 4.

7.2 Lorentz-invariant four-particle amplitudes

We will now restrict the analysis to four-particle amplitudes. In this case, (7.10) takes a simple form. For example, for the (1,3) deformation, the *s*-channel residue is located at

$$z_{\{1,2\}} = -\frac{s}{\langle 23 \rangle [21]} = \frac{\langle 12 \rangle}{\langle 23 \rangle}.$$
(7.14)

The sum over residues in the equation

$$A^{(i,j)}(0) = \sum_{I} \frac{A_L(z_I)A_R(z_I)}{P_I^2} + B_{\infty}^{(i,j)}, \qquad (7.15)$$

will generally receive contributions from two distinct channels. Note that there is no contribution from the channel in which particles i and j are on the same side of the exchange diagram.

Let us consider the interactions of a unique spin S massless particle. We will need the threeparticle amplitudes for helicity configurations (+S, +S, -S),

$$\mathcal{A}_3(1^{+S}, 2_b^{+S}, 3_c^{-S}) = g\left(\frac{[12]^3}{[23][31]}\right)^S$$
(7.16)

We take the interactions to be parity even, so the mostly-minus amplitudes are

$$\mathcal{A}_3(1^{-S}, 2^{-S}, 3^{+S}) = g\left(\frac{\langle 12\rangle^3}{\langle 23\rangle\langle 31\rangle}\right)^S.$$
(7.17)

We shall proceed as follows:

- 1. Consider a specific four-particle amplitude $\mathcal{A}_4(1_a^{h_1}, 2_b^{h_2}, 3_c^{h_3}, 4_d^{h_4})$.
- 2. Choose an appropriate deformation (i, j) acting on the spinors $\lambda^{(i)}$ and $\tilde{\lambda}^{(j)}$.
- 3. Find the locations of all the poles, z_I .
- 4. Check that $B_{\infty}^{(i,j)}$ in (7.15) vanishes.
- 5. Use (7.15) to calculate $A^{(i,j)}(z=0) \equiv \mathcal{A}_4$.
- 6. Choose another pair of spinors i' and j' and repeat the steps 2-5 to find an alternative formula for \mathcal{A}_4 . Ideally, equating this formula with the previous one will yield constraints on the three-particle amplitudes and allow us to construct \mathcal{A}_4 .

Consider, therefore, the amplitude $\mathcal{A}_4(1^{+S}, 2^{+S}, 3^{-S}, 4^{-S})$. Generally speaking, and as we will later see, in order keep the behaviour of A(z) under control as $z \to \infty$ (and thus make $B_{\infty} = 0$), it is a good idea to deform the holomorphic spinor λ of a positive helicity particle and the anti-holomorphic spinor $\tilde{\lambda}$ of a negative helicity particle. First, we choose i = 1, j = 3. Two factorization channels (*s* and *u*, see Fig. 7.1), with residues located at z_s and z_u , contribute to the right hand side of (7.15),



Figure 7.1: The three factorization channels of the tree-level amplitude $\mathcal{A}_4(1^{+S}, 2^{+S}, 3^{-S}, 4^{-S})$.

$$A^{(1,3)}(0) = \frac{1}{s} \left(g \frac{[12]^{3S}}{[2I_s]^S [I_s1]^S} \right) \left(g \frac{\langle 34 \rangle^{3S}}{\langle 4I'_s \rangle^S \langle I'_s 3 \rangle^S} \right) + \frac{1}{u} \left(g \frac{[1I_u]^{3S}}{[I_u 4]^S [41]^S} \right) \left(g \frac{\langle 3I'_u \rangle^{3S}}{\langle I'_u 2 \rangle^S \langle 23 \rangle^S} \right) + B_{\infty}(7.18)$$

In the above, we have written I_s , I_u for the intermediate particle, which is treated as incoming into the left-hand side vertex,

$$p_{I_s} = -(p_1(z_s) + p_2), \qquad (7.19)$$

while I' is assumed to be the conjugate particle incoming into the right-hand side vertex,

$$p_{I'_s} = -(p_3(z_s) + p_4) = p_1(z_s) + p_2.$$
(7.20)

Note that $p_{I'_s} = -p_{I_s}$. The spinors of I and I' momenta can be related by [2]

$$|I'\rangle = |I\rangle, \quad |I'] = -|I]. \tag{7.21}$$

Given this, we can rewrite (7.18) in terms of I_s and I_u alone:

$$A^{(1,3)}(0) = \frac{1}{s}g^2 \frac{[12]^{3S} \langle 34 \rangle^{3S}}{[2I_s]^S [I_s1]^S \langle 4I_s \rangle^S \langle I_s3 \rangle^S} + \frac{1}{u}g^2 \frac{[1I_u]^{3S} \langle 3I_u \rangle^{3S}}{[I_u4]^S [41]^S \langle I_u2 \rangle^S \langle 23 \rangle^S} + B_{\infty}.$$
 (7.22)

Using $p_{I_s} = -(p_1(z_s) + p_2) = p_3(z_s) + p_4$ and $p_{I_u} = -(p_1(z_u) + p_4) = p_2 + p_3(z_u)$, we get

$$\begin{split} A^{(1,3)}(0) &= \frac{1}{s}g^2 \left(-\frac{[12]^3 \langle 34 \rangle^3}{[2\hat{1}] \left(\langle 3|(\hat{1}+2)|1] \right) \langle 4\hat{1} \rangle} \right)^S + \frac{1}{u}g^2 \left(\frac{([1|(\hat{1}+4)|3\rangle)^3}{[\hat{1}4][41] \langle \hat{1}2 \rangle \langle 23 \rangle} \right)^S + B^{(1,3)}_{\infty} \\ &= \frac{1}{s}g^2 \left(\frac{[12]^3 \langle 34 \rangle^3}{t \frac{\langle 34 \rangle}{\langle 23 \rangle} [12] \langle 23 \rangle} \right)^S + \frac{1}{u}g^2 \left(\frac{(\langle 34 \rangle [41])^3}{t \frac{\langle 23 \rangle}{\langle 34 \rangle} [41] \langle 23 \rangle} \right)^S + B^{(1,3)}_{\infty} \end{split}$$

$$= g^{2} \frac{1}{t^{S}} \left(\frac{1}{s} + \frac{1}{u}\right) [12]^{2S} \langle 34 \rangle^{2S} + B_{\infty}^{(1,3)}$$

$$= g^{2} \frac{1}{t^{S}} \left(\frac{-t}{su}\right) [12]^{2S} \langle 34 \rangle^{2S} + B_{\infty}^{(1,3)}$$

$$= -g^{2} \frac{[12]^{2S} \langle 34 \rangle^{2S}}{st^{S-1}u} + B_{\infty}^{(1,3)}.$$
(7.23)

Let us now calculate $\mathcal{A}_4(1^{+S}, 2^{+S}, 3^{-S}, 4^{-S})$ using the deformation of spinors λ_1 and $\tilde{\lambda}_4$. This time, the two residues on the right of (7.15) correspond to *s* and *t* channels (Fig. 7.1). We have

$$A^{(1,4)}(0) = \frac{1}{s} \left(g \frac{[12]^{3S}}{[2I_s]^S [I_s 1]^S} \right) \left(g \frac{\langle 34 \rangle^{3S}}{\langle 4I'_s \rangle^S \langle I'_s 3 \rangle^S} \right) + \frac{1}{t} \left(g \frac{[1I_t]^{3S}}{[3I_t]^S [31]^S} \right) \left(g \frac{\langle I'_t 4 \rangle^{3S}}{\langle 42 \rangle^S \langle 4I'_t \rangle^S} \right) + B^{(1,4)}_{\infty}$$
(7.24)

$$= \frac{1}{s}g^2 \frac{1}{u^S} [12]^{2S} \langle 34 \rangle^{2S} + \frac{1}{t}g^2 \frac{1}{u^S} [12]^{2S} \langle 34 \rangle^{2S} + B_{\infty}^{(1,4)}$$
(7.25)

$$=g^{2}\left(\frac{s+t}{u^{S}}\right)\frac{[12]^{2S}\langle 34\rangle^{2S}}{st} + B_{\infty}^{(1,4)}$$
(7.26)

$$= -g^2 \frac{[12]^{2S} \langle 34 \rangle^{2S}}{stu^{S-1}} + B_{\infty}^{(1,4)} \,. \tag{7.27}$$

Since $A^{(1,3)}(0)$ and $A^{(1,4)}(0)$ should both be equal to the same, undeformed amplitude A_4 , we can now equate (7.23) with (7.27),

$$-g^{2} \frac{[12]^{2S} \langle 34 \rangle^{2S}}{st^{S-1}u} + B_{\infty}^{(1,3)} = -g^{2} \frac{[12]^{2S} \langle 34 \rangle^{2S}}{stu^{S-1}} + B_{\infty}^{(1,4)} \,. \tag{7.28}$$

The boundary terms $B_{\infty}^{(i,j)}$ can be constrained by estimating the large z behaviour of the deformed amplitudes. As $z \to \infty$, the tree-level Feynman diagram contribution to $A^{(1,3)}(z)$ can be estimated as follows. First, note that the exchange diagram contains two powers of three-particle amplitudes \mathcal{A}_3 , each of which has bare mass dimension S. Each three-particle amplitude is therefore quadratic in the momenta and proportional to the product of all polarization tensors of the external particles. In the case of the (1,3) deformation, the polarization tensors $e^{+2}(\mathbf{k}_1)$ and $e^{-2}(\mathbf{k}_3)$ are proportional to $(\lambda_1)^{-S} \sim z^{-S}$ and $(\tilde{\lambda}_3)^{-S} \sim z^{-S}$, while all others are independent of z. The four-particle amplitude also contains a factor of $1/P_{ab}^2$ (the propagator) which is proportional to z^0 or to z^{-1} as $z \to \infty$. Collecting all the factors of z together, we get, in the s and u channels,

$$(z^{2S} \text{ from the momenta}) \times (z^{-2S} \text{ from the pol. tensors}) \times (z^{-1} \text{ from the propagator}) \sim z^{-1}.$$

While in the t channel, one of the constituent three-particle amplitudes will be independent of z and so will be the propagator,

$$(z^{S} \text{ from the momenta}) \times (z^{-2S} \text{ from the pol. tensors}) \times (z^{0} \text{ from the propagator}) \sim z^{-S}$$

Therefore, the contribution from an exchange diagram will behave as z^{-1} or z^{-S} as $z \to \infty$. This contribution to $A^{(1,3)}(z)$ will vanish as $|z| \to \infty$ provided that S > 0.

Let us now consider the contact contribution. This will again be proportional to $e^{+2}(\mathbf{k}_1)$ and $e^{-2}(\mathbf{k}_3)$, which will contribute an overall factor of z^{-2S} . If the quartic vertex has bare mass dimension (the number of derivatives) not greater than 2S-1, the amplitude is under control as $z \to \infty$. However, if the number of derivatives in that vertex is greater than 2S - 1, then the boundary term $B_{\infty}^{(1,3)}$ can be non-zero and the four-particle amplitude cannot be fully known just from three-particle amplitudes. The situation here is similar to that of the previous subsection: the four-particle amplitude can be constructed from the three-particle ones provided that the number of derivatives in the quartic vertex does not exceed 2S - 1.

An analogous argument for the (1, 4) deformation shows that $B_{\infty}^{(1,4)}$ also vanishes if the number of derivatives in the quartic vertex does not exceed 3. Thus, we have

$$-g^{2} \frac{[12]^{2S} \langle 34 \rangle^{2S}}{st^{S-1} u} = -g^{2} \frac{[12]^{2S} \langle 34 \rangle^{2S}}{stu^{S-1}} \,. \tag{7.29}$$

So (assuming $g \neq 0$),

$$t^{S-2} = u^{S-2} \,. \tag{7.30}$$

This equation is satisfied if and only if S = 2. Therefore, we can now conclude that the *S*-derivative interaction (which dominates in the IR) between spin S > 0 particles may only exist if S = 2, i.e. only if the particles are equivalent to gravitons. In that case we can construct the four-particle amplitude $\mathcal{A}_4(1^{+2}, 2^{+2}, 3^{-2}, 4^{-2})$ (under the assumption that the quartic interaction has at most 3 derivatives),

$$\mathcal{A}_4(1^{+2}, 2^{+2}, 3^{-2}, 4^{-2}) = -g^2 \frac{[12]^4 \langle 34 \rangle^4}{stu} \,. \tag{7.31}$$

This construction can then be extended to higher-point amplitudes. Therefore, under the assumption

that GR is a two-derivative theory, all tree-level amplitudes can be constructed in a straightforward manner from the three-particle amplitudes and depend only on one coupling $g = \frac{1}{M_{\text{Pl}}}$. On the level of the action, this is a consequence of the diffeomorphism symmetry of GR, which allows for only one two-derivative, gauge invariant scalar, namely $\sqrt{-gR}$. We note, however, that the Lagrangian description, unlike the on-shell techniques, does not allow for an easy derivation of n-particle amplitudes.

As we saw in this section, the BCFW method heavily relies on controlled behaviour of the amplitude in the limit of large momentum. Asymptotic behaviour $A(z) \sim i(z^0)$ guarantees that the boundary term vanishes, thus allowing for derivation of a four-particle amplitude from a three-particle one. It is interesting how the dependence of the deformed amplitude on z can be controlled in the case of boost-violating theories. This is the subject of the next section, published as [3].

7.3 The Boostless Bootstrap and BCFW Momentum Shifts

In this section, we show that the BCFW method can be applied in a boost-violating setting, despite possible divergence of the deformed amplitude at large z. The section is based on [3].

7.3.1 Constructibility criterion

In its original formulation, the above described method is reserved for *constructible* theories for which $B^{(i,j)}(z)$ vanishes. In this case the singular parts of undeformed amplitudes can be compared with one another and the full four-particle amplitude is determined by the three-particle ones. Since $B^{(i,j)}(z)$ is regular, it is sufficient to prove that the amplitude tends to zero as $z \to \infty$. This is usually a non-trivial matter, necessitating a reference to the Lagrangian and a detailed counting of powers of momenta. Fortunately, many theories describing nature are constructible including, most notably, YM [159, 160] and GR [161] (scalar field theories are not constructible in the sense described above. This has lead to new, interesting momentum shifts and on-shell recursion relations being derived for scalar theories with non-linearly realised symmetries [225, 226]).

However, for the boost-breaking amplitudes of interest here, it is unlikely that $B^{(i,j)}(z)$ would vanish, since the unknown functions of energies will in general contribute positive powers of z to the tree-level amplitude. Indeed, for both particles *i* and *j*, the deformation of their energies is linear in z and so the divergence at large z gets worse as additional powers of energy are included:

$$E_i(z) = E_i(0) + \frac{z}{2} (\bar{\sigma}^0)^{\dot{\alpha}\alpha} \lambda_{\alpha}^{(j)} \tilde{\lambda}_{\dot{\alpha}}^{(i)}, \qquad (7.32)$$

$$E_j(z) = E_j(0) - \frac{z}{2} (\bar{\sigma}^0)^{\dot{\alpha}\alpha} \lambda_{\alpha}^{(j)} \tilde{\lambda}_{\dot{\alpha}}^{(i)}.$$
(7.33)

The BCFW method can still be useful for non-constructible theories, however. One possibility is to introduce a distinction between *accessible* and *inaccessible* singularities of $\mathcal{A}_4^{(i,j)}(z)$, along the lines of [227]. We say a singularity is *accessible* via a deformation of momenta *i* and *j* if this singularity is approached as $z \to z_*$ for some z_* . Otherwise we say it is *inaccessible*. The regular term $B^{(i,j)}(z)$, by definition, cannot have any singularities in the *z*-plane and therefore cannot contribute to any residues of the accessible singularities of $\mathcal{A}_4^{(i,j)}(z)$. But it may exhibit inaccessible singularities. As an illustration of this distinction, consider a single scalar theory which is famously non-constructible. In the absence of additional global charges, the three-particle amplitude is a non-zero constant, $\mathcal{A}_3 = g$, and so we have

$$\mathcal{A}_{4}^{(1,2)}(0) = g^{2}\left(\frac{1}{t} + \frac{1}{u}\right) + B^{(1,2)}(0), \qquad (7.34)$$

$$\mathcal{A}_{4}^{(1,4)}(0) = g^{2}\left(\frac{1}{s} + \frac{1}{t}\right) + B^{(1,4)}(0).$$
(7.35)

The consistency condition $\mathcal{A}_4^{(1,2)}(0) = \mathcal{A}_4^{(1,4)}(0)$ can be satisfied by choosing $B^{(1,2)}(z) = \frac{g^2}{s}$ and $B^{(1,4)}(z) = \frac{g^2}{u}$, since these two functions do not have any accessible singularities with regards to their own deformations.

In the following section we will constrain boost-breaking amplitudes using the fact that the regular term $B^{(i,j)}(z)$ does not have any accessible singularities. We will see that for spinning particles, we can derive the highly non-trivial constraints first found in [2] (Chapter 6).

7.3.2 Constraining three-particle interactions

In this section we will constrain three-particle interactions for theories of a single spin-S particle with integer S. We will derive the constraints first presented in [2] (Chapter 6). We also checked that the BCFW techniques allow us to recover other results in of Chapter 6, namely those of (gravitational) Compton scattering and the full analysis for a scalar or a photon coupled to gravity. Those calculations

contain only minor differences compared with what is presented below so we omit the details in favour of brevity. We remind the reader that we do not impose boost invariance, but only demand that the free theory is Poincaré invariant, with the on-shell condition $E^2 - \mathbf{p}^2 = 0$ for each particle, and that boost violations enter the action only through time derivatives.

Consider the amplitude $\mathcal{A}_4(1^{+S}2^{-S}3^{+S}4^{-S})$, where superscripts denote the helicities of incoming particles of some integer spin S. We will impose matching conditions between deformations (1,2) and (1,4). First consider (1,2). Using the expressions for $A_3(1^{+S}, 2^{+S}, 3^{-S})$ and $A_3(1^{-S}, 2^{-S}, 3^{+S})$ given in (6.48), and²

$$p_1(z_t) + p_3 = \frac{[13]}{[14]} \lambda^{(3)} \tilde{\lambda}^{(4)},$$
 (7.36)

$$p_1(z_u) + p_4 = \frac{[14]}{[13]} \lambda^{(4)} \tilde{\lambda}^{(3)}, \qquad (7.37)$$

to eliminate all copies of $\lambda^{(I)}$ and $\tilde{\lambda}^{(I)}$, which are the spinors associated with the exchanged particle³, we find

$$\mathcal{A}_{4}^{(1,2)}(0) = B^{(1,2)}(0) + \left(\frac{1}{t}F_{\hat{1},3}F_{\hat{2},4} + \frac{1}{u}F_{\hat{1},-\hat{1}-4}F_{\hat{2},-\hat{2}-3}\right) \left(\frac{[13]^{2}\langle 24\rangle^{2}}{s}\right)^{S}.$$
(7.38)

In the *u*-channel we have summed over the two possibilities for the helicity configuration of the exchanged particle but given (7.37), only one of these is non-zero. To keep formulae compact, here we have introduced subscripts to the F's to denote their arguments e.g. $F(E_i, E_j) \equiv F_{i,j}$ and $F(E_i, E_j + E_k) \equiv F_{i,j+k}$. Again, hats denote deformed objects evaluated at the appropriate points e.g. in the 1/t coefficient, $F_{1,3} \equiv F(\hat{E}_1(z_t), E_3)$ where $\hat{E}_1(z_t)$ is the deformed energy of particle 1 evaluated at $z = z_t$. Likewise, in the 1/u coefficient, hatted energies are evaluated at $z = z_u$. We have also removed the H/AH superscripts since the functions are identical, due to parity, up to an inconsequential overall sign (see Chapter 6).

¹These amplitudes arise from the leading order couplings. Higher-dimension operators give rise to the $A_3(1^{+S}, 2^{+S}, 3^{+S})$, $A_3(1^{-S}, 2^{-S}, 3^{-S})$ amplitudes but we don't consider these here. We refer the reader to [2] for a discussion on these amplitudes.

²Here we have assumed [13] and [14] are non-zero, and therefore t = 0 and u = 0 are approached as $\langle 13 \rangle = 0$ and $\langle 14 \rangle = 0$ respectively (or as [24] = 0 and [23] = 0 respectively, by momentum conservation). ³For example, in the *t*-channel we set $\lambda^{(I)} = \alpha \lambda^{(3)}$ and $\tilde{\lambda}^{(I)} = \beta \tilde{\lambda}^4$ with $\alpha \beta = \frac{[13]}{[14]}$. When computing the residue, α

and β only appear in the product $\alpha\beta$.

Now, we can also write

$$\mathcal{A}_{4}^{(1,2)}(0) = \tilde{B}^{(1,2)}(0) + \left(\frac{1}{t}F_{1,3}F_{2,4} + \frac{1}{u}F_{1,-1-4}F_{2,-2-3}\right) \left(\frac{[13]^2 \langle 24 \rangle^2}{s}\right)^S,$$
(7.39)

where here we dropped the hat above all the energies, which indicates that the expression is evaluated at their *undeformed* values. This can be justified as follows. We assume that the F's can be Taylor expanded around the undeformed energies. The deformed energies are

$$\hat{E}_{1}(z_{t}) = E_{1} - \frac{t}{2\langle 23\rangle[13]} (\bar{\sigma}^{0})^{\dot{\alpha}\alpha} \lambda_{\alpha}^{(2)} \tilde{\lambda}_{\dot{\alpha}}^{(1)},$$
(7.40)

$$\hat{E}_2(z_t) = E_2 + \frac{t}{2\langle 23\rangle[13]} (\bar{\sigma}^0)^{\dot{\alpha}\alpha} \lambda_{\alpha}^{(2)} \tilde{\lambda}_{\dot{\alpha}}^{(1)}, \qquad (7.41)$$

with similar expressions evaluated at $z = z_u$. For the class of Lagrangians considered in this paper, energies appear in $F_{a,b}$ only with non-negative powers, and each factor of an energy is generated by a time derivative acting on the field. In the above, we see that potential new singularities generated by the deformed energies are all *inaccessible*, as they correspond to the vanishing of $\langle 23 \rangle$ or [13], but these do not depend on z. Moreover, only the leading term in the Taylor expansion will exhibit accessible singularities, since in all subleading terms t and u will be cancelled out. We can therefore simply absorb all subleading terms into $B^{(1,2)}$, thus introducing $\tilde{B}^{(1,2)}$ that still does not contain any terms singular in t or u. Although it could become singular in some kinematic configurations, especially at s = 0, that is not a problem, because this singularity is *inaccessible* and we only demand that $\tilde{B}^{(1,2)}$, for those configurations for which it can be defined, does not have any singularities as a function of z.

We now play the same game for the (1, 4) deformation which amounts to interchanging particles 2 and 4. We have

$$\mathcal{A}_{4}^{(1,4)}(0) = \tilde{B}^{(1,4)}(0) + \left(\frac{1}{t}F_{1,3}F_{4,2} + \frac{1}{s}F_{1,-1-2}F_{4,-4-3}\right) \left(\frac{[13]^2 \langle 24 \rangle^2}{u}\right)^S.$$
(7.42)

We discussed the S = 0 case earlier where we showed that equating $\mathcal{A}_4^{(1,2)}(0)$ and $\mathcal{A}_4^{(1,4)}(0)$ requires us to make certain choices for the boundary terms. Let us now consider S > 0 with S an integer. We see that $\mathcal{A}_{4}^{(1,2)}(0)$ in (7.39) contains terms proportional to $1/(ts^S)$ and $1/(us^S)$ which are both singular in more than one Mandelstam variable and thus cannot be accounted for or modified by $\tilde{B}^{(1,2)}(0)$ nor $\tilde{B}^{(1,4)}(0)$. A similar observation applies to $\mathcal{A}_{4}^{(1,4)}(0)$ in (7.42). Thus, by matching the amplitudes we find the necessary condition

$$\frac{a}{s^{S}t} + \frac{b}{s^{S}u} = \frac{c}{u^{S}t} + \frac{d}{u^{S}s},$$
(7.43)

where

$$a = F_{1,3}F_{2,4}, (7.44)$$

$$b = F_{1,-1-4}F_{2,-2-3}, (7.45)$$

$$c = F_{1,3}F_{4,2}, (7.46)$$

$$d = F_{1,-1-2}F_{4,-4-3}. (7.47)$$

Recalling that s + t + u = 0, this constraint, given that it must be valid for all kinematics, is equivalent to

$$au^{S} - b(s+u)u^{S-1} - cs^{S} + d(s+u)s^{S-1} = 0.$$
(7.48)

For S = 1 we therefore have a = (b - d) = -c, or equivalently,

$$F_{1,3}F_{2,4} - F_{1,-1-4}F_{2,-2-3} + F_{1,-1-2}F_{4,-4-3} = 0, (7.49)$$

which is simply an alternative form of (6.97). Assuming that the F's are polynomials, in Chapter 6 we showed that the only solution to this system is $F \equiv 0$ once we impose that the S = 1 functions are alternating polynomials as dictated by Bose symmetry⁴. We therefore see that the leading order three-particle interactions for three-photons must vanish, as is the case for Poincaré invariant theories. For S = 2 we require a = b = c = d, or equivalently,

$$F_{1,3}F_{2,4} = F_{1,-1-4}F_{2,-2-3} = F_{1,-1-2}F_{4,-4-3},$$
(7.50)

⁴The spinor helicity parts of the S = 1 three-particle amplitudes are odd under the exchange of identical particles so if the amplitudes are to be even by Bose symmetry, the *F*'s must be alternating.

which gives rise to the constraints (6.105) - (6.107) once we use the fact that the S = 2 functions are symmetric in their arguments by Bose symmetry⁵ (this also makes the a = c constraint trival). In Chapter 6 that the only solution to this system is F = constant and so again the three-particle interactions are reduced to their Poincaré invariant form, but this time the amplitudes are non-zero and are those of GR. Finally, for S > 2, it is simple to see that a = b = c = d = 0 is required and therefore there are no consistent three-particle interactions for these massless, higher-spin particles even when boosts are broken, as we also concluded in the previous chapter.

7.3.3 Summary

In Chapter 6 (published as [2]), the singular parts of four-particle amplitudes were bootstrapped by demanding that they factorise into a product of on-shell three-particle amplitudes on simple poles. In that work, consistent factorisation was implemented directly without making use of BCFW momentum shifts. In this short section, we have shown that the same results can be derived by using BCFW shifts to automate consistent factorisation. We presented full details for the illustrative cases of single spin-S particle amplitudes but have also checked that the procedure produces the expected results for Compton scattering, and its gravitational analogue, as well as for scalars or photons coupled minimally to gravity. For single spin-S particles, the boostless bootstrap teaches us that the leading three-particle couplings for a photon must vanish, the leading three-particle couplings for a graviton must be those of GR, while massless higher-spinning particles do not self-interact. For photon Compton scattering, boost-breaking interactions between the photon, scalars and spin-1/2 fermions are allowed and can be described by Lagrangians with generalised boost-breaking gauge redundancies. For gravitational Compton scattering, all couplings must reduce to their boost-invariant counter-parts with universal couplings of all particles to gravity. Finally, scalars and photons that are minimally coupled to gravity are forced to have Poincaré invariant self-interactions (constant or vanishing, respectively for the scalar and the photon). For full details see [2].

Although the theories we have considered are not *a priori* constructible, in the sense that the boundary terms do not necessarily vanish at large z, we have still been able to use BCFW shifts to constrain the three-particle couplings contributing to particle exchange. This does mean, of course,

⁵For S = 2, the spinor parts of the three-particle amplitudes are even under the exchange of identical particles and so the *F*'s are symmetric polynomials.

that the three-particle amplitudes themselves do not fully fix the four-particle ones. Indeed, all of the four-particle amplitudes we have constructed are defined up to the presence of "contact" terms that are regular for all kinematic configurations. It would be very interesting to investigate the possibility of using generalised momentum shifts, possibly along the lines of [228], to recursively derive exact higher-point amplitudes even if only for a subset of boost-breaking theories. It would also be very interesting to investigate the generalised on-shell recursion relations introduced in [229], where boundary terms are fixed with additional knowledge of a subset of the zeros of the deformed amplitude, in our boost-breaking setting.

Chapter 8

Bootstrapping Large Graviton non-Gaussianities

Abstract

Gravitational interferometers and cosmological observations of the cosmic microwave background offer us the prospect to probe the laws of gravity in the primordial universe. To study and interpret these datasets we need to know the possible graviton non-Gaussianities. To this end, we derive the most general tree-level three-point functions (bispectra) for a massless graviton *to all orders in derivatives*, assuming scale invariance. Instead of working with explicit Lagrangians, we take a bootstrap approach and obtain our results using the recently derived constraints from unitarity, locality and the choice of vacuum. Since we make no assumptions about de Sitter boosts, our results capture the phenomenology of large classes of models such as the effective field theory of inflation and solid inflation. We present formulae for the infinite number of parity-even bispectra. Remarkably, for parity-odd bispectra, we show that unitarity allows for only a handful of possible shapes: three for graviton-graviton, three for scalar-graviton-graviton and one for scalar-scalar-graviton, which we bootstrap explicitly. These parity-odd non-Gaussianities can be large, for example in solid inflation, and therefore constitute a concrete and well-motivated target for future observations.

8.1 Introduction

Being the only force that stubbornly refuses to be described at arbitrarily high energies within the dominant framework of quantum field theory, gravity is a prominent testing ground for our understanding of fundamental physics. Ideas from string theory, the study of black holes and gauge-gravity duality suggest that the field-theoretic gravitons that appear to describe low-energy phenomena very well, most likely don't provide the right language to discuss non-perturbative and high-energy aspects of quantum gravity. Given how difficult it is to establish what gravity *is*, a useful approach to the problem is to ask the related question: *What can gravity be?* For example, given the framework of quantum mechanics as we know it, what different descriptions of gravity can be formulated that are mathematically and physically consistent?

Concrete and quantitative progress in this direction has been achieved for quantum fields on flat spacetime, e.g. via the derivation of positivity bounds that constrain effective field theories admitting standard and consistent UV-completions. To understand and model cosmology, and in particular inflation, dark energy and dark matter, we would like to use these bounds as a compass pointing us in the direction of the most promising consistent theories. However, the set of consistent theories of dynamical gravity is different in flat and cosmological spacetimes. Concrete examples of this difference include a theory of interacting massless spin 3/2 particles, which is given by supergravity in flat space, but is not known in de Sitter; or the theory of a scalar coupled to gravity with boostbreaking interactions, which is easily written down in cosmological spacetimes, as in many realistic models of inflation and dark energy, but which is inconsistent in flat spacetime as can be shown by examining amplitude factorization [2]. At the same time, new probes of gravity have just become available through the observation of gravitational waves at interferometers, and there is a substantial international effort and a well-kindled hope to detect a cosmological background of gravitational waves from the primordial universe. In light of these considerations, it is highly desirable to study the consistency of effective field theories of gravity directly on the cosmological spacetimes where we want to use them for phenomenology.

In this work, we are interested in constraining the possible phenomenological descriptions of gravity around a (quasi) de Sitter spacetime, with an eye towards applications to inflation. To this end,

we focus on the natural observables of this system: cosmological correlators, namely the expectation values of the product of fields in the space-like asymptotic future, which we will call the (conformal) boundary. Given a concrete model, such observables can be computed in perturbation theory using the in-in formalism. However, since we don't know what the "right" model is, we will follow a different approach, which is inspired by parallel progress in the study of amplitudes [30, 31, 28]. In particular, we aim to derive all possible correlators that are compatible with fundamental principles such as symmetry, unitarity and locality. This model-independent approach goes under the name of the cosmological bootstrap and has received growing attention in recent years [12–14, 33, 92, 168, 177–186, 230–239].

We will focus on the simplest non-trivial correlators of massless spin-2 fields, a.k.a. gravitons, and massless scalars, namely three-point functions or bispectra. An important previous result is that of [171], where, assuming invariance under the full isometry group of de Sitter, it was shown that for gravitons only three cubic cosmological wavefunctions are allowed, and of those only the two parity-even ones can lead to a non-vanishing bispectrum [240]. Several additional results can be derived in this setup using conformal Ward identities, as done for example in [171, 175, 241, 115, 116, 90], and parity-odd correlators in CFT's were recently discussed in [242, 243]. While some of these results are remarkable because they are non-perturbative in nature, we are faced with the issue that de Sitter boosts are actually broken in all cosmological models and, in particular, during inflation. Unlike the breaking of scale-invariance, *the breaking of boosts is in general not slow-roll suppressed and may be large*, as for example in so-called P-of-X models (a.k.a. "k-inflation" [244]), where the Lagrangian is an arbitrary function of the kinetic term, or more general models captured by the effective field theory of inflation [147, 136]. In fact, as emphasized in [168], the breaking of de Sitter boosts is a necessary condition to have phenomenologically large non-Gaussianities.

Therefore, to make contact with cosmological observations, in this work we will weaken the assumption of full de Sitter invariance and instead assume only the symmetries that have been observed in primordial perturbations, namely statistical homogeneity, isotropy and (approximate) scale invariance. In particular, we will allow for arbitrary breaking of de Sitter boosts. The price to pay for this smaller set of isometries is that we have to work in perturbation theory and we will restrict ourselves to tree-level.

Progress in understanding boost-breaking gravitational interactions has been achieved using effective field theories and the Lagrangian approach in a series of recent papers [245–250]. This approach is quite general and intuitive but its computational complexity grows quickly as one considers operators with an increasing number of derivatives. To overcome this difficulty, here we will instead follow the "boostless" cosmological bootstrap approach proposed in [231, 33], which partially builds upon results in [171, 183, 177, 2, 12] and is reviewed in Section 8.2. Our approach leverages the powerful constraints of fundamental principles such as unitarity, locality and the choice of vacuum and allows us to bootstrap all tree-level graviton bispectra to *any order in derivatives*, as well as all parity-odd mixed bispectra. At the end of our derivation we will see how the bootstrap results can be understood in the familiar Lagrangian language (see Section 8.5).

Our main results are summarized below:

- Unitarity and the choice of the Bunch-Davies vacuum highly restrict the allowed set of parity-odd correlators. In particular, for massless scalars and gravitons and to all orders in derivatives, there is only a finite number of tree-level correlators. In contrast, the number of possible wavefunction coefficients and Lagrangian interactions grows without bound as one increases the number of derivatives in the effective field theory expansion. In more detail, a contact parity-odd correlator can only arise when there is a logarithmic IR-divergence in the associated wavefunction coefficient. In turn this may only happen when 2n_{∂η} + n_{∂i} ≤ 3, where n_{∂η} and n_{∂i} are respectively the number of time and space derivatives in the parity-odd interaction¹. This explains on general grounds why parity-odd correlators where found to vanish in the scale-invariant limit in a number of explicit calculations [240, 251, 249].
- We computed all tree-level graviton bispectra to any order in derivatives, assuming in particular scale-invariance and massless gravitons. There are infinitely many *parity-even* graviton bispectra B₃. For example, for the choice of all plus helicities these are given by the symmetrized products of three factors

$$e_3^3 B_3^{+++}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \mathbf{SH}_{+++} \sum_{\text{permutations}} h_\alpha(k_1, k_2, k_3) \psi_3^{\text{trimmed}}(k_1, k_2, k_3) \,. \tag{8.1}$$

The first factor SH_{+++} includes the spinor helicities and provides the correct little-group scaling. It is given by

¹This is valid for any contract n-point function and assumes that there is at most one time derivative per field. Interactions with more than one time derivative can always be re-written in terms of those with at most one time derivative using the equations of motion.

$$\mathbf{SH}_{+++} = \frac{[12]^2 [23]^2 [31]^2}{e_3^2} = -8e_{ij}^+(\mathbf{k}_1)e_{jk}^+(\mathbf{k}_2)e_{ik}^+(\mathbf{k}_3)$$
(8.2)

$$= -\frac{k_T^3 \left(8e_3 - 4k_T e_2 + k_T^3\right)}{e_3^2}, \qquad (8.3)$$

where [ij] is the square-bracket product of helicity spinors, e_{ij} are polarization tensors, and k_T , e_2 and e_3 are the elementary symmetric polynomials defined in (8.12). The second factor h_{α} roughly accounts for the contractions between spatial derivatives and polarization tensors and can be any one of the following four possibilities

$$h_0 = 1, \quad h_2 = k_2 k_3, \quad h_4 = I_1^2 I_2 I_3, \quad h_6 = I_1^2 I_2^2 I_3^2,$$
 (8.4)

where

$$I_a \equiv (k_T - 2k_a) = k_b + k_c - k_a$$
 $a \neq b \neq c.$ (8.5)

For parity-odd interactions there are a further five possibilities for h_{α} . Finally, the third factor is the "trimmed" wavefunction ψ_3^{trimmed} , which roughly accounts for the conformal time integrals of mode functions, time derivatives and spatial derivatives contracted with each other. This can be any of the infinitely-many rational-function solutions of the manifestly local test, $\partial_{k_a}\psi_3^{\text{trimmed}} = 0$ at $k_a = 0$ (see (8.39)), which are conveniently organized in terms of the increasing order of the polynomial in the numerator, roughly corresponding to the derivative expansion of an effective field theory. For concreteness, the first few explicit bispectra are given in (8.156) through (8.174). The bispectra corresponding to other helicity choices can be derived from the all-plus bispectrum as discussed in Section 8.4.2.

• Remarkably, *there are only three parity-odd graviton bispectra at tree level to all orders in derivatives*. These are explicitly found to be

$$\begin{split} B_3^{+++} &= \frac{g_{1,1}}{e_3^3} \mathrm{SH}_{+++} k_T \left(k_T^2 - 2e_2 \right) \,, \\ B_3^{++-} &= \frac{g_{1,1}}{e_3^3} \mathrm{SH}_{++-} I_3 \left(k_T^2 - 2e_2 \right) \,, \\ B_3^{+++} &= \frac{g_{1,2}}{e_3^3} \mathrm{SH}_{+++} \left(-3e_3 + k_T e_2 \right) \,, \end{split}$$

$$\begin{split} B_3^{++-} &= \frac{g_{1,2}}{e_3^3} \mathrm{SH}_{++-} \left[(k_1 + k_2) \left(k_1 k_2 + k_3^2 \right) - (k_1^2 + k_2^2) k_3 \right] \,, \\ B_3^{+++} &= \frac{g_{3,3}}{e_3^3} \mathrm{SH}_{+++} I_1 I_2 I_3 \,, \\ B_3^{++-} &= \frac{g_{3,3}}{e_3^3} \mathrm{SH}_{++-} I_1 I_2 k_T \,, \end{split}$$

where the $g_{\alpha,p}$ are arbitrary real coupling constants whose indices denote respectively the number α of spatial momenta contracted with polarization tensors and the total number of derivatives p in the associated interaction. The remaining two helicity configurations, namely - - - and - - +, can be obtained via a parity transformation, while keeping in mind the odd-parity of the above bispectra. In the effective field theory of inflation only one specific combination of these three shapes can appear and it must be accompanied by a parity-odd correction to the free theory. In this case, the final parity-odd graviton bispectrum must be small, and in particular much smaller than the standard parity-even contribution from General Relativity (GR) computed in [106]. By contrast, all three shapes above can appear in a general model of solid inflation [130], without any modification to the free theory and with arbitrarily large amplitudes. Hence, these three parity-odd graviton bispectra are an important target for non-Gaussian searches in the graviton sector. Their shapes are plotted in Figure 8.5. In solid inflation they should be accompanied by correlated scalar-scalar-graviton and scalar-gravitongraviton bispectra with larger signal-to-noise ratios (see Section 8.5.4).

• We show that *there are only three parity-odd scalar-graviton-graviton bispectra and one scalar-scalar-graviton bispectrum at tree level to all orders in derivatives*, assuming scale invariance and manifest locality. These are given by

$$B_3^{00+} = \frac{h_{3,3}}{e_3^3} \frac{[13]^2 [23]^2}{k_3^2 [12]^2} I_3^2 k_3 , \qquad (8.6)$$

$$B_3^{0++} = \frac{[23]^4}{k_2^2 k_3^2 e_3^3} [q_{1,1}(k_2+k_3)k_1^2 + q_{1,2,a}(k_2^3+k_3^3) + q_{1,2,b}(k_2k_3^2+k_3k_2^2)], \qquad (8.7)$$

where $h_{3,3}$ and $q_{\alpha,p}$ are arbitrary coupling constants. Notice, however, that for scalars nonmanifestly local interactions do arise in GR. We show in Section 8.5 that the above scalarscalar-graviton bispectrum can be large in solid inflation, but not in the effective field theory of inflation, and can be the leading observational signal.

The rest of this work is organized as follows. In Section 8.2, we review the framework and tools

used to bootstrap correlators in general scale-invariant and boost-breaking theories, and in particular the boostless bootstrap rules, the constraints of unitarity in the form of the Cosmological Optical Theorem and associated cutting rules, the constraints from locality on massless fields in the form of the Manifestly Local Test, and finally the spinor helicity formalism for spinning cosmological correlators. The expert reader might skip directly to Section 8.3, where we derive a very general consequence of unitarity for tree-level contact correlators that implies that to all orders in derivatives there is only a small and finite number of non-vanishing parity-odd correlators. Then in Section 8.4 we present formulae for all graviton bispectra to any order in the derivative expansion and show that there are only three non-vanishing parity-odd bispectra, and infinitely many parity-even ones. In Section 8.5, we show that the parity-odd bispectra can indeed arise in realistic models such as solid inflation, and study how they are constrained in the effective field theory of inflation. We also discuss their detectability by studying the associated signal-to-noise ratio. We conclude in Section 8.6 with an outlook on future research directions.

Notation and conventions Throughout we will work with the mostly positive metric signature (-+++) and we define the three-dimensional Fourier transformation as

$$f(\mathbf{x}) = \int \frac{d^3 \mathbf{k}}{(2\pi)^3} f(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x}) \equiv \int_{\mathbf{k}} f(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x}), \qquad (8.8)$$

$$f(\mathbf{k}) = \int d^3 \mathbf{x} f(\mathbf{x}) \exp(-i\mathbf{k} \cdot \mathbf{x}) \equiv \int_{\mathbf{x}} f(\mathbf{x}) \exp(-i\mathbf{k} \cdot \mathbf{x}).$$
(8.9)

We use bold letters to refer to vectors, e.g. x for spatial co-ordinates and k for spatial momenta, and we write the magnitude of a vector as $k \equiv |\mathbf{k}|$. We will sometimes refer to these objects as "energies" even though there is no time-translation symmetry in cosmology. We will use $i, j, k, \ldots = 1, 2, 3$ to label the components of SO(3) vectors, and $a, b, c = 1, \ldots, n$ to label the *n* external fields. For wavefunction coefficients and cosmological correlators we use ψ_n and B_n respectively:

$$\psi_n(\mathbf{k}_1, \dots, \mathbf{k}_n) \equiv \psi'_n(\mathbf{k}_1, \dots, \mathbf{k}_n)(2\pi)^3 \delta^3 \left(\sum \mathbf{k}_a\right) , \qquad (8.10)$$
$$\langle \mathcal{O}(\mathbf{k}_1) \dots \mathcal{O}(\mathbf{k}_n) \rangle \equiv \langle \mathcal{O}(\mathbf{k}_1) \dots \mathcal{O}(\mathbf{k}_n) \rangle'(2\pi)^3 \delta^3 \left(\sum \mathbf{k}_a\right)$$
$$\equiv B_n(\mathbf{k}_1, \dots, \mathbf{k}_n) (2\pi)^3 \delta^3 \left(\sum \mathbf{k}_a\right) , \qquad (8.11)$$

and we will drop the primes on ψ_n when no confusion arises. We will also use a prime to denote a derivative with respect to the conformal time e.g. $\phi' = \partial_\eta \phi$. We will often encounter polynomials that are symmetric in three variables, for example, for the + + + correlator. We write these in terms of the elementary symmetric polynomials (ESP):

$$k_T = k_1 + k_2 + k_3 \,, \tag{8.12}$$

$$e_2 = k_1 k_2 + k_1 k_3 + k_2 k_3 , \qquad (8.13)$$

$$e_3 = k_1 k_2 k_3 \,. \tag{8.14}$$

8.2 Bootstrap techniques from symmetries, locality and unitarity

In this section, we define the objects that we will be bootstrapping, namely wavefunction coefficients appearing in the wavefunction of the universe and the associated cosmological correlators. In this part of the chapter we also review bootstrap techniques that have been recently developed in the context of boost-breaking interactions. We outline how symmetries, locality and unitarity can be directly imposed on cosmological observables thanks to a set of Boostless Bootstrap Rules [231], a Manifestly Local Test [33] and the Cosmological Optical Theorem [12–14, 238]. Finally, we review the cosmological spinor helicity formalism that we will use to succinctly present graviton bispectra.

8.2.1 The wavefunction of the universe and cosmological correlators

Let's start by reviewing the computation of the wavefunction of the universe Ψ and defining wavefunction coefficients ψ_n which will be our primary objects of interest. We will also remind the reader how correlation functions are extracted from knowledge of the wavefunction.

We take the background geometry to be that of rigid de Sitter (dS) spacetime which we write as²

$$ds^{2} = a^{2}(\eta)(-d\eta^{2} + d\mathbf{x}^{2}), \qquad a(\eta) = -\frac{1}{\eta H},$$
(8.15)

where the conformal time coordinate $\eta \in (-\infty, 0)$ and H is the constant Hubble parameter which we will often set to unity. This background geometry is an excellent approximation of an inflationary solution, and considering quantum fields fluctuating on this rigid background allows us to compute

²These are the so-called Poincaré or flat-slicing coordinates and cover half of the maximally extended de Sitter spacetime. This spacetime is the one relevant for the discussion of cosmological observations.

 $\phi(\eta, \mathbf{x})$ and the transverse, traceless massless graviton $\gamma_{ij}(\eta, \mathbf{x})$. When our results apply to both scalars and graviton, especially in Section 8.3, we will use $\varphi(\eta, \mathbf{x})$ with any SO(3) indices suppressed.

The free action of a massless scalar is

$$S_{\phi,\text{free}} = \int d\eta d^3 \mathbf{x} \, a^2(\eta) \frac{1}{2} \left[\phi'^2 - c_s^2 \partial_i \phi \partial_i \phi \right] \,, \tag{8.16}$$

where we have allowed for an arbitrary, constant speed of sound c_s which signals the fact we are allowing for dS boosts to be spontaneously or explicitly broken³. Working in momentum space, we write the quantum free field operator as

$$\hat{\phi}(\eta, \mathbf{k}) = \phi^{-}(\eta, k)a_{\mathbf{k}} + \phi^{+}(\eta, k)a_{-\mathbf{k}}^{\dagger}, \qquad (8.17)$$

where the mode functions $\phi^{\pm}(\eta, k)$ correspond to solutions of the free classical equation of motion and are given by

$$\phi^{\pm}(\eta, k) = \frac{H}{\sqrt{2c_s^3 k^3}} (1 \mp i c_s k \eta) e^{\pm i c_s k \eta} .$$
(8.18)

The mode functions for graviton fluctuations take the same form as (8.18) (with $c_s = 1$) with the addition of polarisation tensors $e_{ij}^h(\mathbf{k})$, with $h = \pm 2$, as required by little group scaling. This is because for each polarisation mode the equation of motion is that of a massless scalar. The polarisation tensors satisfy the following conditions:

$$e_{ii}^{h}(\mathbf{k}) = k^{i} e_{ij}^{h}(\mathbf{k}) = 0$$
 (transverse and traceless), (8.19)

$$e_{ij}^{h}(\mathbf{k}) = e_{ji}^{h}(\mathbf{k})$$
 (symmetric), (8.20)

$$e_{ij}^{h}(\mathbf{k})e_{jk}^{h}(\mathbf{k}) = 0 \qquad (\text{lightlike}), \qquad (8.21)$$

$$e_{ij}^{h}(\mathbf{k})e_{ij}^{h'}(\mathbf{k})^{*} = 4\delta_{hh'} \qquad (\text{normalization}), \qquad (8.22)$$

³When the speed of sound differs from the speed of light appearing in the metric, $c_s \neq 1$, the sound cone is not invariant under de Sitter boosts, a fact which can be simply seen in the flat-space limit, where de Sitter boosts reduce to Lorentz boosts.

$$e_{ij}^{h}(\mathbf{k})^{*} = e_{ij}^{h}(-\mathbf{k}) \qquad (\gamma_{ij}(x) \text{ is real}). \qquad (8.23)$$

As we explained in the introduction, we are interested in scenarios where dS boosts are broken since we know that these symmetries could not have been exact in the early universe, and large non-Gaussianities are associated with a large breaking of boosts [168]. We take the remaining symmetries of the dS group to be exact: spatial translations, spatial rotations and dilations. A general interaction vertex with n fields, scalars and gravitons, therefore takes the schematic form

$$S_{\rm int} = \int d\eta d^3 \mathbf{x} \, a(\eta)^{4-N_{\rm deriv}} \partial^{N_{\rm deriv}} \varphi^n \,, \tag{8.24}$$

where ∂ stands for either time derivatives ∂_{η} or spatial derivatives ∂_i , and N_{deriv} is the total number of derivatives. Spatial derivatives and the graviton's indices are contracted with the SO(3) invariant objects δ_{ij} and ϵ_{ijk} and the overall number of scale factors is dictated by scale invariance.

We now turn to the wavefunction of the universe which we denote as Ψ . We are interested in this wavefunction evaluated at the end of inflation or alternatively on the late-time boundary of dS space, at a conformal time which we denote as η_0 . Ultimately we will take $\eta_0 \rightarrow 0$. To illustrate the wavefunction of the universe method, let us focus on a single massless scalar ϕ . The generalisation to gravitons simply requires the addition of SO(3) indices where appropriate. We refer the reader to [12, 252–254, 181] for further details. At late-times, the wavefunction has an expansion in the late-time value of the scalar, $\phi(\mathbf{k}) \equiv \phi(\eta_0, \mathbf{k})$, given by

$$\Psi[\eta_0, \phi(\mathbf{k})] = \exp\left[-\sum_{n=2}^{\infty} \frac{1}{n!} \int_{\mathbf{k}_1, \dots, \mathbf{k}_n} \psi_n(\mathbf{k}_1 \dots \mathbf{k}_n) \phi(\mathbf{k}_1) \dots \phi(\mathbf{k}_n)\right], \quad (8.25)$$

where we have written the exponent as an expansion in powers of the field multiplied by the *wave*function coefficients $\psi_n(\mathbf{k}_1 \dots \mathbf{k}_n)$ which contain the dynamical information about the bulk processes. Invariance of the theory under spatial translations ensures that the $\psi_n(\mathbf{k}_1 \dots \mathbf{k}_n)$ always contain a momentum conserving delta function and so we can write

$$\psi_n(\mathbf{k}_1\dots\mathbf{k}_n) = \psi'_n(\mathbf{k}_1\dots\mathbf{k}_n)(2\pi)^3\delta^3(\mathbf{k}_1+\dots+\mathbf{k}_n).$$
(8.26)

We will often drop the prime even when we do not explicitly include the delta function. At weak
coupling, we can compute the leading contribution to the wavefunction using the saddle-point approximation where the wavefunction is completely fixed by the value of the action evaluated on classical solutions:

$$\Psi[\eta_0, \phi(\mathbf{k})] \approx e^{iS_{\rm cl}[\phi(\mathbf{k})]}.$$
(8.27)

Traditionally, one computes $S_{cl}[\phi(\mathbf{k})]$ in perturbation theory using Feynman diagrams which involve bulk interaction vertices, bulk-boundary propagators $K(\eta, k)$ and bulk-bulk propagators $G(\eta, \eta', k)$. If we denote the scalar's free equation of motion as $\mathcal{O}(\eta, k)\phi = 0$, then these propagators satisfy

$$\mathcal{O}(\eta, k)K(\eta, k) = 0, \tag{8.28}$$

$$\mathcal{O}(\eta, k)G(\eta, \eta', k) = -\delta(\eta - \eta'), \tag{8.29}$$

with boundary conditions

$$\lim_{\eta \to \eta_0} K(\eta, k) = 1, \qquad \qquad \lim_{\eta \to -\infty(1-i\epsilon)} K(\eta, k) = 0 \qquad (8.30)$$

$$\lim_{\eta,\eta'\to\eta_0} G(\eta,\eta',k) = 0, \qquad \qquad \lim_{\eta,\eta'\to-\infty(1-i\epsilon)} G(\eta,\eta',k) = 0.$$
(8.31)

We can then write both propagators in terms of the positive and negative frequency mode functions as

$$K(k,\eta) = \frac{\phi_k^+(\eta)}{\phi_k^+(\eta_0)},$$

$$G(p,\eta,\eta') = i \left[\theta(\eta - \eta') \left(\phi_p^+(\eta')\phi_p^-(\eta) - \frac{\phi_p^-(\eta_0)}{\phi_p^+(\eta_0)} \phi_p^+(\eta)\phi_p^+(\eta') \right) + (\eta \leftrightarrow \eta') \right]$$

$$= iP(p) \left[\theta(\eta - \eta') \frac{\phi_p^+(\eta')}{\phi_p^+(\eta_0)} \left(\frac{\phi_p^-(\eta)}{\phi_p^-(\eta_0)} - \frac{\phi_p^+(\eta)}{\phi_p^+(\eta_0)} \right) + (\eta \leftrightarrow \eta') \right],$$
(8.32)
(8.32)

where P(p) is the power spectrum of ϕ and we have introduced the notation $\phi_k(\eta) \equiv \phi(\eta, k)$ to shorten the expressions. In deriving these expressions we have imposed the Bunch-Davies vacuum state as an initial condition which is the assumption that at very early times the mode functions are those of the flat-space theory. Physically this is because at very high energies the modes do not feel the expansion of the universe.

Now to extract the wavefunction coefficients one follows the following Feynman rules. For a



Figure 8.1: Contact diagram for n external fields

contact diagram like the one shown in Figure 8.1, we insert an overall factor of (-i) and perform a single time integral where the integrand is a product of the coupling parameter, the n bulk-boundary propagators and their derivatives (as dictated by the interaction vertex), and an appropriate number of scale factors (as dictated by scale invariance). Time derivatives act on the bulk-boundary propagators whereas spatial derivatives simply bring down a factor of ik_i , as is the case for scattering amplitudes. We integrate from the far past at $\eta = -\infty(1 - i\epsilon)$ to the future boundary at $\eta = \eta_0$. This $i\epsilon$ prescription ensures that there is a short period of evolution in Euclidean time rather than Lorentzian time that dampens the exponential factors appearing in the integral, thereby projecting the theory onto the vacuum state [255, 106]. In analogy to scattering amplitudes, we finally sum over all possible permutations. For an exchange diagram like the one shown in Figure 8.2 we now have two time integrals, one for each vertex. The vertices contribute n and m powers of the bulk-boundary propagators, possibly time-differentiated as dictated by the interaction vertices, while the internal line requires us to include one bulk-bulk propagator, which may also be differentiated with respect to time. The number of scale factors is fixed by scale invariance and as for contact diagrams we sum over all possible permutations. The generalisation of these rules to more complicated tree diagrams is simple, with a time integral for each local vertex. See Appendix A of [33] for more details and examples.

As an example, for a massless scalar with a $\frac{a(\eta)}{3!}\phi'^3$ self-interaction in the bulk, the three-point wavefunction coefficient is given by



Figure 8.2: Single exchange diagram for n + m external fields

$$\psi_{\phi'^{3}}(k_{1},k_{2},k_{3}) = -i \int d\eta \, a(\eta) \, K'(k_{1},\eta) \, K'(k_{2},\eta) \, K'(k_{3},\eta) \,, \tag{8.34}$$

while the s-channel four-point exchange diagram is given by

$$\psi^{s}_{\phi'^{3}\times\phi'^{3}} = -i \int d\eta' \, d\eta \, \left(a(\eta) K'(k_{1},\eta) K'(k_{2},\eta) \right) \, \partial_{\eta} \partial_{\eta}' G(s,\eta,\eta') \, \left(a(\eta') K'(k_{3},\eta') K'(k_{4},\eta') \right),$$
(8.35)

where $s = |\mathbf{k}_1 + \mathbf{k}_2|$ is the "energy" of the internal line and we have suppressed the integration limits. This traditional computational process can be complicated due to the (nested) time integrals that have to be performed, which may obscure the origin of analytic properties of the final answer. In this chapter we will usually avoid computing time integrals altogether and instead fix the final form of the wavefunction coefficients using symmetries, locality and unitarity, only computing explicit time integrals to verify that all parity-odd bispectra can be generated in solid inflation (Section 8.5.2). In general, the wavefunction is a complex function of the kinematics and η_0 , since we are evaluating the action on complex field configurations, and we will use our bootstrap methods to construct both the real and imaginary parts.

With the wavefunction in hand, one can extract equal-time (late-time) expectation values using the usual quantum mechanics formula. We have

$$\langle \phi(\mathbf{k}_1) \dots \phi(\mathbf{k}_n) \rangle = \frac{\int \mathcal{D}\phi \ \Psi \Psi^* \ \phi(\mathbf{k}_1) \dots \phi(\mathbf{k}_n)}{\int \mathcal{D}\phi \ \Psi \Psi^*}, \qquad (8.36)$$

for an *n*-point function of scalars. Here $\mathcal{D}\phi$ is the functional measure on a fixed time slice. Correlators are therefore fixed via the bulk dynamics through the probability distribution $\Psi\Psi^*$. We will use this equation in Section 8.3 to derive some general results for cosmological correlators arising from unitary time evolution in the bulk.

8.2.2 Boostless Bootstrap Rules

We now turn to reviewing bootstrap techniques for efficient computation of late-time wavefunctions/correlators. In [231] a set of *Boostless Bootstrap Rules* was introduced that enables one to write down general structures for the three-point functions of massless scalars and gravitons without assuming full dS symmetries. In total, six rules were introduced, each based on the following principles:

- Rule 1: Spatial translations, spatial rotations and scale invariance,
- Rule 2: Tree-level approximation for wavefunctions and correlators in dS,
- Rule 3: High-energy boundary condition in the form of an amplitude limit,
- Rule 4: Bose statistics for wavefunctions/correlators of external bosons,
- Rule 5: Bunch-Davies initial vacuum state,
- Rule 6: Soft theorems.

For the curvature perturbation in inflation each of these six rules are necessary to bootstrap the bispectrum [231], however for gravitons and spectator scalars that are the primary interest in this chapter, rules 3 and 6 are not required and are replaced by the *Manifestly Local Test* of [33] which we will review in the following subsection. Before doing so let us first review the other rules (1, 2, 4, 5) and refer the reader to [231] for further details on all rules.

• Rule 1: Spatial translations, spatial rotations and scale invariance. These symmetries ensure that wavefunction coefficients can be written as a product of a *polarisation factor*, which is an SO(3) invariant function of polarisation tensors and spatial momenta, multiplied by a *trimmed wavefunction coefficient* which is only a function of the energies:

$$\psi_n = \sum_{\text{contractions}} (\text{polarization factor}) \times (\text{trimmed wavefunction coefficient}).$$
 (8.37)

We take all coefficients appearing in the polarisation factor to be real and therefore include any factors of *i* that might appear when converting to momentum space, or simply as part of the Feynman rules, in the trimmed part which we will denote as $\tilde{\psi}_n$. We denote the total number of spatial momenta appearing in the polarisation factor as α . For the bispectrum of massless gravitons which is our primary interest in this chapter, we have

$$\psi_{3} = \sum_{\text{contractions}} \left[e^{h_{1}}(\mathbf{k}_{1}) e^{h_{2}}(\mathbf{k}_{2}) e^{h_{3}}(\mathbf{k}_{3}) \mathbf{k}_{1}^{\alpha_{1}} \mathbf{k}_{2}^{\alpha_{2}} \mathbf{k}_{3}^{\alpha_{3}} \right] \psi_{3}^{\text{trimmed}},$$
(8.38)

with $\alpha_1 + \alpha_2 + \alpha_3 = \alpha$. Here we have already stripped off the ever-present momentum conserving delta function that is a consequence of spatial momentum conservation. Furthermore, scale invariance ensures that for all n we have $\psi_n \sim k^3$ which cancels the scaling of the threedimensional delta function thereby ensuring invariance of Ψ . If one also includes dS boosts as a symmetry, the trimmed wavefunction coefficients for gravitons are very constrained [171]. In this chapter we are interested in boost-breaking scenarios and so will not impose invariance under dS boosts.

- Rule 2: Tree-level approximation for wavefunctions/correlators in dS. This rule simply imposes that the bispectrum is a rational function of the external kinematics up to possible logarithmic terms. Such logs will indeed be captured by our bootstrap analysis. Our focus in this chapter will be at tree-level but progress is now also being made on using bootstrap techniques at loop-level [14, 233, 235].
- Rule 4: Bose statistics for wavefunctions/correlators of external bosons. This rule enforces invariance under permutations of the momenta of identical fields.
- Rule 5: Bunch-Davies initial vacuum state. The assumption of a Bunch-Davies initial state enforces that the only allowed poles for contact diagrams are in the total energy $k_T = \sum_{a=1}^{n} k_a$. The degree of the leading k_T pole is given by $p = 1 + \sum_A (\Delta_A - 4)$ where the sum is over all vertices appearing in a given diagram and Δ_A is their mass dimension [231]. We only have one type of pole since the integrands appearing in the bulk formalism only depend on the positive frequency modes. For excited initial states both positive and negative frequency modes can contribute leading to so-called flattened singularities, see e.g. [256, 257] for the phenomenology of such poles. It is also interesting to note that the residue of the leading order

 k_T poles contain the flat-space scattering amplitude for the same process [171, 12, 172].

These four rules will play an important role in our ability to bootstrap graviton bispectra in Section 8.4.

8.2.3 Manifestly Local Test

In [33] a condition, referred to as the *Manifestly Local Test* (MLT), was introduced that must be satisfied by both contact and exchange *n*-point wavefunction coefficients of massless scalars and gravitons with manifestly local interactions. Manifestly local interactions are those with only positive powers of derivatives, i.e. without inverse Laplacians; this is a natural locality condition for gravitons and spectator scalars in dS at cubic order in perturbations [231]. Manifest locality can be violated upon integrating out the non-dynamical modes in a gravitational theory, so such a violation is a feature of the self-interactions of the inflationary curvature perturbation [106] as well as gravitons at quartic and higher order in the fields. The MLT was used in [33] to bootstrap bispectra of the Goldstone mode in the Effective Field Theory of Inflation [136] to all orders in derivatives, and used in conjunction with partial energy recursion relations to bootstrap inflationary trispectra (see also [183] for a use of energy shifts for the flat-space wavefunction). The MLT was also recently employed in [243]. The MLT offers a conceptually simple yet very powerful bootstrap technique and will be a central feature of this work.

The MLT takes the form

$$\frac{\partial}{\partial k_c}\psi_n(k_1,...,k_n;\{p\};\{\mathbf{k}\})\Big|_{k_c=0} = 0, \qquad \forall \, c = 1,\ldots,n\,,$$
(8.39)

where k_a are the energies of the external fields, $\{p\}$ collectively denotes the energies of possible exchange fields while $\{k\}$ collectively denotes a possible dependence of *n*-point functions on spatial momenta and polarisation tensors. We will also often also use $\{k\}$ to collectively denote the external energies. The derivative with respect to one of the external energies is taken while keeping all other variables fixed and this condition must be met for all external energies if they are those of a massless scalar or a graviton in de Sitter. Two complementary derivations of the MLT were given in [33]. The first arises from demanding that exchange diagrams have the appropriate singularities while the second comes directly from the bulk representation of such n-point functions. We refer the reader to [33] for details of the first method while reviewing the second here.

The computation of tree-level diagrams in the bulk formalism reduces to nested time integrals of the following schematic form

$$\psi_n(\{k\};\{p\};\{\mathbf{k}\}) \sim \int \left(\prod_A^V d\eta_A F_A\right) \left(\prod_a^n \partial_\eta^\# K_\phi(k_a)\right) \left(\prod_m^I \partial_\eta^\# G(p_m)\right), \quad (8.40)$$

where the F_A 's denote the momentum dependence due to the spatial derivatives and polarisation tensors in the V vertices, each vertex representing a contact interaction placed at the conformal time η_A . We have included a bulk-boundary propagator for each external field and have allowed for an arbitrary number of time derivatives acting on these propagators. Finally, we have allowed for Iinternal bulk-bulk propagators G, possibly with time derivatives. Now we differentiate the above expression with respect to one of the external energies. This derivative acts only on the bulk-boundary propagator associated to this energy, because F_A depend only on the spatial momenta and polarisations while $G(p_m)$ depend only on energies of internal legs. Assuming that η integrals and $\frac{\partial}{\partial k_c}$ commute, we have

$$\frac{\partial}{\partial k_c} \psi_n \Big|_{k_c=0} \sim \int \left(\prod_A^V d\eta_A F_A \right) \left(\prod_{a \neq c} \partial_\eta^\# K_\phi(k_a) \right) \left(\partial_\eta^\# \left(\frac{\partial}{\partial k_c} K_\phi(k_c) \right) \Big|_{k_c=0} \right) \left(\prod_m^I \partial_\eta^\# G(p_m) \right)$$
(8.41)

The bulk-boundary propagator for a massless graviton is the same as for a massless scalar up to the presence of a polarisation tensor. In both cases, we have

$$\frac{d}{dk}K(\eta,k) = \frac{d}{dk}\left((1-ik\eta)e^{ik\eta}\right) = k\eta^2 e^{ik\eta},$$
(8.42)

which vanishes at k = 0. It follows that (8.41) must vanish. We emphasise that we have not assumed anything about the form of the ψ_n , so the MLT holds for contact and exchange diagrams, even those with IR-divergences: it follows from a simple property of the bulk-boundary propagators, namely that $\frac{d}{dk}K(\eta, k)$ vanishes at k = 0. In fact, this property also holds in slow roll inflation, for both massless gravitons and massless scalars, and therefore the MLT (8.39) is applicable in that case as well. The main obstacle to extending all of our results beyond exact scale invariance is therefore not the MLT itself, but the assumption of scale invariance (Rule 1), which allows us to write down a simple ansatz for the wavefunction coefficient before applying the MLT (as will be shown in detail in Section 8.4). We will return to the prospect of employing the MLT to construct slow-roll corrections in the future.

The MLT, in conjuction with the bootstrap rules from the previous section, can be used to find *all* consistent, tree-level, contact wavefunction coefficients for massless scalars and gravitons in de Sitter. Let us present a constructive proof of this claim. As a first step, we find an exhaustive list of polarization factors (see (8.37)), which covers all possible contractions of tensor indices. Then we write down an ansatz for ψ_n^{trimmed} , consistent with rules 2 and 5 (rule 4 is automatically satisfied once we sum over the permutations). Any such ansatz can be written in the form of a bulk integral

$$\psi_n^{\text{trimmed}} \sim \int d\eta f(k_a, \mathbf{k}_a, \mathbf{k}_b; \eta) e^{ik_T \eta},$$
(8.43)

where $f(k_a, \mathbf{k}_a, \mathbf{k}_b; \eta)$ is a polynomial in the energies k_a and the scalar products $\mathbf{k}_a, \mathbf{k}_b$, with appropriate factors of η as required by scale invariance. The exponential factor contributes the needed poles in k_T , and these are the only possible poles, as dictated by rules 2 and 5. The IR divergences, which are of the form η_0^{-m} or $\log(-k_T\eta_0)$, are fully accounted for by those terms in f that have negative powers of η .

The final ingredient is the MLT, which imposes the following constraints on f:

$$\frac{\partial f}{\partial k_a}\Big|_{k_a=0} + i\eta f\Big|_{k_a=0} = 0.$$
(8.44)

It is easy to see that any such polynomial (assuming scale invariance) can be written as

$$f(k_a, \mathbf{k}_a, \mathbf{k}_b; \eta) = (1 - ik_1\eta)g(k_2, \dots, k_n, \mathbf{k}_a, \mathbf{k}_b; \eta) + k_1^2h(k_a, \mathbf{k}_a, \mathbf{k}_b; \eta), \qquad (8.45)$$

where g, h are polynomials satisfying

$$\frac{\partial g}{\partial k_a}\Big|_{k_a=0} + i\eta g\Big|_{k_a=0} = 0, \quad a \neq 1,$$
(8.46)

$$\frac{\partial h}{\partial k_a}\Big|_{k_a=0} + i\eta h\Big|_{k_a=0} = 0, \quad a \neq 1.$$
(8.47)

Then, we can repeat the decomposition (8.45), albeit now for g and h. By iterating over a = 1, 2, ..., n, we can arrive at a general form of $f(k_a, \mathbf{k}_a, \mathbf{k}_b; \eta)$:

$$f(k_a, \mathbf{k}_a, \mathbf{k}_b; \eta) = \sum_{S \subset \mathbb{Z}_n} \left(\prod_{j \notin S} \left(1 - ik_j \eta \right) \prod_{j \in S} \left(k_j^2 \right) h_S \left(k_{a \in S}, \mathbf{k}_a, \mathbf{k}_b; \eta \right) \right),$$
(8.48)

where h_S are polynomials in the $k_a \in S$ and the scalar products $\mathbf{k}_a.\mathbf{k}_b$. The sum is taken over all subsets S of the set $\mathbb{Z}_n := \{1, 2, ..., n\}$. It will now be sufficient to show that any term of the above sum can be produced by some linear combination of functions constructed from bulk-boundary propagators. In fact, we can focus on the case where h_S is a monomial, since any polynomial is just a linear combination of those. If this monomial includes factors of $\mathbf{k}_a.\mathbf{k}_b$, we can generate them from the Lagrangian by writing pairs of spatial derivatives contracted with each other, so from now on, let us assume for simplicity that h_S is a monomial that does not include such factors. Reinstating powers of η as required by scale invariance, we are thus looking for a functional of bulk propagators that would generate

$$\psi_n^{\text{trimmed}} \sim \int d\eta \prod_{j \notin S} \left(1 - ik_j \eta\right) \prod_{j \in S} \left(k_j^{2+n_j}\right) \eta^{\alpha + \sum_{j \in S} n_j + 2|S| - 4} e^{ik_T \eta}, \quad (8.49)$$

for some arbitrary $n_j \ge 0$; α is the energy dimension of the polarization factor. The linear combination we are looking for is, up to an overall constant,

$$\eta^{\alpha-4} \prod_{j \notin S} K(k_j, \eta) \prod_{j \in S} \left(K_{2+n_j}(k_j, \eta) \right) , \qquad (8.50)$$

where $K(k_i, \eta)$ is the usual bulk-boundary propagator, and

$$K_2(k,\eta) \equiv \eta \partial_\eta K(k,\eta) = k^2 \eta^2 e^{ik\eta}, \qquad (8.51)$$

$$K_3(k,\eta) \equiv -i\left(\eta^2 \partial_\eta^2 K(k,\eta) - \eta \partial_\eta K(k,\eta)\right) = k^3 \eta^3 e^{ik\eta},\tag{8.52}$$

$$K_{n+2}(k,\eta) \equiv k^2 \eta^2 K_n(k,\eta) \quad \text{for } n \ge 2.$$
(8.53)

Each of these functions can be obtained from the massless bulk-boundary propagators by applying time derivatives, Laplacians $(k^2 \leftrightarrow -\nabla^2)$ and taking linear combinations. Recall that we can introduce the dependence on $\mathbf{k}_a \cdot \mathbf{k}_b$ by introducing pairs of spatial derivatives, followed by taking linear combinations again to account for terms with distinct dependencies on $\mathbf{k}_a \cdot \mathbf{k}_b$. Therefore, any integral of the form (8.45) can be generated by a linear combination of products of bulk-boundary propagators, their time derivatives, factors of $a(\eta)^2 k_a^2$ and by pairs of spatial derivatives contracted with each another. This entails that any solution to the MLT corresponds to a combination of some manifestly local operators.

8.2.4 Cosmological Optical Theorem

The final bootstrap tool we are going to review is the Cosmological Optical Theorem (COT) [12] which is a consequence of unitary time evolution in the bulk. It was shown in [12] that if the wavefunction of the universe is normalised at time η then it only remains normalised at time η' if contact and exchange wavefunction coefficients satisfy some simple yet powerful relations. Assuming a Bunch-Davies initial condition, the bulk-boundary propagator of fields of general mass and spin on any FLRW spacetime satisfies (see [13] for a proof and a discussion of the related technical assumptions)

$$K^*(-k^*,\eta) = K(k,\eta), \qquad k \in \mathbb{C},$$
(8.54)

from which one can derive the COT for contact diagrams [12]

Disc
$$[i\psi_n(k_1,...,k_n; \{\mathbf{k}\})] = i [\psi_n(k_1,...,k_n; \{\mathbf{k}\}) + \psi_n^*(-k_1^*,...,-k_n^*; \{-\mathbf{k}\})] = 0,$$
 (8.55)

which must be satisfied by any contact *n*-point function arising from unitary evolution in the bulk spacetime. Note that all spatial momenta in the second term get a minus sign, $\mathbf{k} \to -\mathbf{k}$, and all energies are analytically continued. One is usually interested in real values of the energies *k*, and so in the following we will drop the complex conjugation. This notation is unambiguous as long as one adopts the prescription that *all negative energies are approached from the lower-half complex plane*. For scalars it is clear from (8.55) how the second term should be computed but for spinning fields the presence of polarisation tensors introduces slight complications which were addressed in [13]. Ultimately any polarisation factors appear as a common factor in this contact COT since e.g. $e_{ij}^{h}(\mathbf{k})^* = e_{ij}^{h}(-\mathbf{k})$. The COT is therefore not constraining the polarisation factor (which is constrained by symmetry), rather it is constraining the trimmed part of the wavefunction that in the bulk representation arises from performing the bulk time integrals. This of course makes sense as the COT is indeed a consequence of unitary time evolution. For our purposes in this chapter the COT for contact diagrams is enough and we will use it in Section 8.3 to derive some general results about cosmological correlators, but the consequences of unitarity for exchange diagrams are also known [12, 13] and were used extensively in [33] to bootstrap inflationary trispectra. The COT for exchange diagrams relates the discontinuity of an exchange diagram to products of the contributing sub-diagrams, multiplied by the power spectrum of the exchanged field. It is reminiscent of the factorisation theorem for scattering amplitudes. A complementary derivation of the COT was given in [238] where the consequences of excited initial states were also considered. The COT was also extended to general FLRW spacetimes in [13] and to loop level in the form of cutting rules in [14], see also [92] for a recent discussion of cosmological cuts. Unitarity constraints on cosmological observables were also recently studied in [233, 236, 235]. See [258–260] for analogous statements in anti-de Sitter (AdS) space.

8.2.5 Cosmological spinor helicity formalism

In this chapter we are primarily concerned with bootstrapping graviton bispectra and just as is the case for scattering amplitudes, wavefunctions/correlators of spinning fields are most compactly presented using spinors rather than polarisation tensors. We end this section by reviewing the cosmological spinor helicity formalism and refer to the reader to [171, 181] for other presentations.

The spinor helicity formalism is most useful when we have null momenta as is the case for massless on-shell particles in flat-space and has been used extensively in that setting. In our cosmological setting the spatial momentum k is not null, but we can define a null four-component object $k_{\mu} = (k, \mathbf{k})$, with $k = |\mathbf{k}|$, which we can express as the outer product of two spinors via

$$k_{\alpha\dot{\alpha}} = \sigma^{\mu}_{\alpha\dot{\alpha}} k_{\mu} = \lambda_{\alpha} \dot{\lambda}_{\dot{\alpha}} \,, \tag{8.56}$$

where $\sigma^{\mu} = (\mathbf{1}, \boldsymbol{\sigma})$ and $\boldsymbol{\sigma}$ are the Pauli matrices. Using the relation $\sigma^{\mu}_{\alpha\dot{\alpha}}\bar{\sigma}^{\dot{\beta}\beta}_{\mu} = 2\delta^{\beta}_{\alpha}\delta^{\dot{\beta}}_{\dot{\alpha}}$ (we follow the conventions used in [205]), where $\bar{\sigma}^{\mu} = (\mathbf{1}, -\boldsymbol{\sigma})$, the inverse of (8.56) is

$$k_{\mu} = \frac{1}{2} \bar{\sigma}_{\mu}^{\dot{\alpha}\alpha} k_{\alpha\dot{\alpha}} \,. \tag{8.57}$$

A little group transformation by definition should leave this four-momentum invariant, so we can model this transformation as $\lambda \to t\lambda$, $\tilde{\lambda} \to t^{-1}\tilde{\lambda}$ where each external field transforms with a different constant $t \in \mathbb{C}$. These very simple helicity transformations allow us to easily extract an overall dependence of a wavefunction/correlator on the spinors given some helicity configuration for the external fields, and is one of the primary virtues of the spinor helicity formalism. As usual, dotted and un-dotted indices are raised and lowered by $\epsilon_{\dot{\alpha}\dot{\beta}}$ and $\epsilon_{\alpha\beta}$ respectively e.g. $\tilde{\lambda}_{\dot{\alpha}} = \epsilon_{\dot{\alpha}\dot{\beta}}\tilde{\lambda}^{\dot{\beta}}$, $\lambda_{\alpha} = \epsilon_{\alpha\beta}\lambda^{\beta}$.

Now for objects with three external fields, conservation of spatial momentum $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$ leads to

$$\sum_{a=1}^{3} \lambda_{\alpha}^{(a)} \tilde{\lambda}_{\dot{\alpha}}^{(a)} = k_T(\sigma_0)_{\alpha \dot{\alpha}} \quad \text{and} \quad \langle ab \rangle [ab] = k_T(k_T - 2k_c) \equiv k_T I_c \quad \text{for} \quad a \neq b \neq c \,, \quad (8.58)$$

where we have introduced

$$I_a \equiv \left(k_T - 2k_a\right),\tag{8.59}$$

and we recall that

$$\langle ab \rangle = \epsilon^{\alpha\beta} \lambda^{(a)}_{\alpha} \lambda^{(b)}_{\beta} , \quad [ab] = \epsilon^{\dot{\alpha}\dot{\beta}} \tilde{\lambda}^{(a)}_{\dot{\alpha}} \tilde{\lambda}^{(b)}_{\dot{\beta}} .$$
(8.60)

We remind the reader that the above spinors are not Grassmanian, so these angle and square brackets are anti-symmetric due to the anti-symmetric nature of the epsilon tensors. For scattering amplitudes one also has time translation invariance, which implies $k_T = 0$. In this case the above relations reduce to the usual flat-space ones, see e.g. [28]. Now to construct SO(3) invariant objects we can use (8.60) but can also contract dotted and un-dotted indices using $(\bar{\sigma}^0)^{\dot{\alpha}\alpha}$ [171, 2]:

$$(ab) = (\sigma^0)_{\alpha\dot{\alpha}}\lambda^{(a)}_{\alpha}\tilde{\lambda}^{(b)}_{\dot{\alpha}}, \qquad (8.61)$$

with $(aa) = 2k_a$. We can use (8.58) to obtain an expression for (ab) with $a \neq b$ i.e. the off-diagonal components. We have

$$(ab)[ac] = I_b[bc] \quad \text{for} \quad a \neq b \neq c ,$$
 (8.62)

$$(ab)\langle bc \rangle = I_a \langle ac \rangle \quad \text{for} \quad a \neq b \neq c \,,$$

$$(8.63)$$

and therefore a general three-point function is a function of the angle brackets, the square brackets and the energies.

For spinning fields, we will find it necessary to write polarisation tensors in terms of spinors. The transverse and traceless graviton polarisation tensors $e^{\pm}_{\mu\nu}$ are given by $e^{\pm}_{\mu}e^{\pm}_{\nu}$, where e^{\pm}_{μ} is the polarisation vector for a spin-1 particle of the same momentum. We therefore only need an expression for e_{μ}^{\pm} in the spinor helicity formalism. The form of the polarisation vectors follows from the fact that they must be lightlike, orthogonal to the corresponding momentum, and carry the appropriate helicity weight. We have (see e.g. [28, 2])

$$e_{\alpha\dot{\alpha}}^{+} = 2 \frac{\mu_{\alpha}\tilde{\lambda}_{\dot{\alpha}}}{\langle\mu\lambda\rangle}, \qquad e_{\alpha\dot{\alpha}}^{-} = 2 \frac{\lambda_{\alpha}\tilde{\mu}_{\dot{\alpha}}}{[\mu\lambda]},$$
(8.64)

for generic reference spinors μ_{α} and $\tilde{\mu}_{\dot{\alpha}}$. For scattering amplitudes in flat-space these reference spinors represent the redundancy in defining massless spinning fields as a representation of the Lorentz group, but for cosmology we can make a choice to eliminate this redundancy [171]. Indeed, we can use our freedom to mix dotted and undotted indices to choose

$$\mu_{\alpha} = (\sigma^0)_{\alpha\dot{\alpha}} \tilde{\lambda}^{\dot{\alpha}}, \qquad \tilde{\mu}_{\dot{\alpha}} = (\sigma^0)_{\alpha\dot{\alpha}} \lambda^{\alpha}, \qquad (8.65)$$

which makes the zero component of the polarisation vectors vanish. We can therefore write

$$e^{+}_{\alpha\dot{\alpha}} = \frac{(\sigma^{0})_{\alpha\dot{\beta}}\tilde{\lambda}^{\beta}\tilde{\lambda}_{\dot{\alpha}}}{k}, \qquad e^{-}_{\alpha\dot{\alpha}} = \frac{(\sigma^{0})_{\beta\dot{\alpha}}\lambda^{\beta}\lambda_{\alpha}}{k}, \qquad (8.66)$$

which has the correct normalisation. Under a helicity transformation we have $e^+ \rightarrow t^{-2}e^+$ and $e^- \rightarrow t^2 e^-$, as expected.

With these relations at hand, we can easily convert any SO(3) invariant object containing spatial momenta and polarisation vectors into the spinor helicity formalism using the necessary σ and ϵ identities which are given in [205]. We present a complete list of distinct contractions of SO(3)indices for a massless graviton in Appendix 8.7.1. We will use these relations extensively in Section 8.4.1 where we study the tensor structures for the graviton bispectrum.

8.3 Unitarity constraints on *n*-point cosmological correlators

In this section we are going to use the Cosmological Optical Theorem (COT) for contact diagrams to derive some general results about the form of cosmological correlators. Recall that with the wavefunction of the universe at hand, one can compute expectation values via Eq. (8.36), i.e.

$$\langle \varphi(\mathbf{k}_1) \dots \varphi(\mathbf{k}_n) \rangle = \frac{\int \mathcal{D}\varphi \ \Psi \Psi^* \ \varphi(\mathbf{k}_1) \dots \varphi(\mathbf{k}_n)}{\int \mathcal{D}\varphi \ \Psi \Psi^*},$$
 (8.67)

where in the weak coupling approximation we are using here, the late-time wavefunction is given by

$$\Psi[\eta_0,\varphi(\mathbf{k})] = \exp\left[-\sum_{n=2}^{\infty} \frac{1}{n!} \int_{\mathbf{k}_1,\dots,\mathbf{k}_n} \psi_n(\{k\};\{\mathbf{k}\})\varphi(\mathbf{k}_1)\dots\varphi(\mathbf{k}_n)\right].$$
(8.68)

Here we have made a distinction between the dependence of the wavefunction coefficients on the set of spatial momenta {k} and their norms {k}, since in general we will work away from the physical configuration and treat {k} and {k} as independent objects, for reasons that will become clear. We have not included a possible dependence on internal energies {p} since our focus in this section is on *contact diagrams*. We are going to use the COT to constrain the form of the probability distribution $\Psi\Psi^*$. Here and throughout this section we use $\varphi(\mathbf{k})$ to schematically denote scalars and gravitons, with SO(3) indices suppressed, and each of these fields satisfies $\varphi(\mathbf{k}) = \varphi(-\mathbf{k})^*$ which follows directly from (8.17), (8.18) and (8.23). Now from this perturbative expression for the wavefunction, we have

$$-\log(\Psi\Psi^*) = \left(\sum_{n=2}^{\infty} \frac{1}{n!} \int_{\mathbf{k}_1,\dots,\mathbf{k}_n} \psi_n(\{k\}; \{\mathbf{k}\})\varphi(\mathbf{k}_1)\dots\varphi(\mathbf{k}_n)\right) + \left(\sum_{n=2}^{\infty} \frac{1}{n!} \int_{\mathbf{k}_1,\dots,\mathbf{k}_n} \psi_n(\{k\}; \{\mathbf{k}\})\varphi(\mathbf{k}_1)\dots\varphi(\mathbf{k}_n)\right)^*$$
(8.69)
$$= \left(\sum_{n=2}^{\infty} \frac{1}{n!} \int_{\mathbf{k}_1,\dots,\mathbf{k}_n} \psi_n(\{k\}; \{\mathbf{k}\})\varphi(\mathbf{k}_1)\dots\varphi(\mathbf{k}_n)\right) + \left(\sum_{n=2}^{\infty} \frac{1}{n!} \int_{\mathbf{k}_1,\dots,\mathbf{k}_n} \psi_n^*(\{k\}; \{\mathbf{k}\})\varphi(-\mathbf{k}_1)\dots\varphi(-\mathbf{k}_n)\right).$$
(8.70)

If we change the integration variables on the final line by sending $\{k\} \to \{-k\}$ we have

$$-\log(\Psi\Psi^*) = \sum_{n=2}^{\infty} \frac{1}{n!} \int_{\mathbf{k}_1,\dots,\mathbf{k}_n} [\psi_n(\{k\};\{\mathbf{k}\}) + \psi_n^*(\{k\};\{-\mathbf{k}\})]\varphi(\mathbf{k}_1)\dots\varphi(\mathbf{k}_n).$$
(8.71)

It follows from Gaussian integral formulae that the resulting correlators arising from these contact diagrams, in perturbation theory, are given by

$$B_n^{\text{contact}}(\{k\};\{\mathbf{k}\}) = -\frac{\psi_n'(\{k\};\{\mathbf{k}\}) + \psi_n'^*(\{k\};-\{\mathbf{k}\})}{\prod_{a=1}^n 2\operatorname{Re}\psi_2'(k_a)},$$
(8.72)

where in deriving this expression we kept only terms linear in the coupling constants. For parity-even interactions of scalars and gravitons, the numerator is simply 2Re ψ'_n in which case our expression matches the one that usually appears in the literature.

Let's now use the contact COT to constrain B_n^{contact} . As we reviewed above, unitary time evolution in the bulk inflationary spacetime and the choice of the Bunch-Davies vacuum imply that [12]

$$\psi_n(\{k\};\{\mathbf{k}\}) + \psi_n^*(\{-k\};-\{\mathbf{k}\}) = 0.$$
(8.73)

By directly comparing (8.72) and (8.73), we conclude that

Any contribution to the wavefunction of the universe that is invariant under $\{k\} \rightarrow \{-k\}$, which is a flip in the sign of all external energies, does not contribute to the contact correlator.

What are the implications of this observation? To answer this question we need to look more closely at the form of ψ_n . After stripping away the polarization factor in ψ , see (8.38), the remaining trimmed wavefunction ψ^{trimmed} for a contact interaction can have the following structures:

1. The trimmed wavefunction may be a rational functions of $\{k\}$,

$$\psi_n^{\text{trimmed}} \supset \frac{\text{Poly}_{3-\alpha+q}(\{k\})}{\text{Poly}_a(\{k\})}, \tag{8.74}$$

where the subscripts indicate the degrees of the polynomials and the combination $3 - \alpha + q$ is fixed by scale invariance such that $\psi_n \sim k^3$. If we further impose locality and the Bunch-Davies vacuum as in the bootstrap Rule 5 then the denominator must be k_T to some power, but we will not use this fact in the following.

If α is even, this trimmed wavefunction contains an overall odd number of energies and therefore is not invariant under $\{k\} \rightarrow \{-k\}$, whereas if α is odd, the trimmed wavefunction contains an overall even number of energies and so is invariant under $\{k\} \rightarrow \{-k\}$. So rational terms in the wavefunction can only contribute to the correlator if the polarisation factor has an even number of spatial momenta, which for scalars and gravitons implies parity-even. Conversely, parity-odd interactions of scalars and gravitons have an odd number of derivatives, which are contracted with a Levi-Civita tensor, and the contribution of their rational part to the correlator must vanish. This observation explains why k_T poles were never found in the in-in computation of parity-odd graviton bispectra in the effective theory of inflation performed in [247]: they are simply incompatible with unitarity.

2. The trimmed wavefunction may have logarithmic IR-divergences,

$$\psi_n^{\text{trimmed}} \supset \text{Poly}_{3-\alpha}(\{k\}) \log(-k_T \eta_0) \qquad 3-\alpha \ge 0, \tag{8.75}$$

where again the degree of the polynomial that multiplies the log is fixed by scale invariance. We cannot have any poles multiplying the log and so we need $3 - \alpha \ge 0^4$. Such logs can arise from relevant operators in the bulk at tree-level but are also a common feature of loop corrections [261, 262].

These logs break the $\{k\} \rightarrow \{-k\}$ symmetry for both even and odd α , so they can in principle contribute to the correlator. Unitarity in the form of the contact COT tells us that these logs do not appear on their own but rather always appear in the combination [12]

$$\log(-k_T\eta_0) + \frac{i\pi}{2}, \qquad (8.76)$$

multiplied by a *real* function of $\{k\}$, and possibly a polarisation factor (which also has real coefficients). Indeed, if we consider a wavefunction coefficient of the schematic form

$$\psi_n^{\text{trimmed}} \sim \mathbf{k}^{\alpha} e^{\beta}(\mathbf{k}) [A \log(-k_T \eta_0) + B], \qquad (8.77)$$

where we have allowed for β polarisation structures, a complex polynomial A and a complex rational function B, then the COT (8.73) tells us that (recall that the polarisation factor becomes a common factor on the LHS of the COT)

$$A\log(-k_T\eta_0) + B - A^*[\log(-k_T\eta_0) + i\pi] - B^* = 0.$$
(8.78)

We therefore conclude that Im(A) = 0, $Im(B) = \frac{A\pi}{2}$ while Re(B) is unconstrained and would actually contribute to the rational part of the wavefunction covered above in point 1. It then

⁴This fact can be quite easily seen from the bulk representation and the corresponding time integrals one must perform. We don't have a better "bootstrap" reason but it would be interesting to find one. We note that if the interactions violate manifest locality, there can be poles multiplying the log as they can come from inverse Laplacians.

follows from (8.72) that for even α only the log contributes to the correlator and not the $i\pi$ piece, whereas for odd α the $i\pi$ piece contributes to the correlator but the log does not. For parity-odd interactions of scalars and gravitons, which necessarily have an odd α , we therefore conclude again that the singular part of the wavefunction does not contribute to the correlator. Indeed the parity-odd contributions to the graviton bispectrum computed in [247] come from this $\frac{i\pi}{2}$ part of the wavefunction.

 The trimmed wavefunction may have a polynomial IR-divergence 1/η₀^q with q ≥ 1 as η₀ → 0. These terms may not have any singularity as k_T → 0 because there we recover scattering amplitudes which, by time translation invariance, must be time independent. Scale invariance then tells us that

$$\psi_n^{\text{trimmed}} \supset \sum_{q=1}^3 \frac{\text{Poly}_{3-\alpha-q}(\{k\})}{\eta_0^q} \qquad 3-\alpha-q \ge 0.$$
(8.79)

Now we observe that we need $\alpha + q$ to be even in order to break the $\{k\} \rightarrow \{-k\}$ symmetry, while the MLT can only be satisfied if $3 - \alpha - q \ge 2$ or $3 - \alpha - q = 0$. These two conditions imply that $3 - \alpha - q \ge 3$, which contradicts the fact that $q \ge 1$. Thus, a combination of the COT and MLT leads us to conclude that $\eta_0 = 0$ poles cannot contribute to cosmological correlators arising from manifestly-local bulk interactions⁵.

We have therefore seen that parity-odd contact correlators of scalars and gravitons do not contain any total-energy singularities: the only part of the trimmed wavefunction that survives when we compute parity-odd correlators is finite or vanishing as $k_T \rightarrow 0$. These contributions arise from the polynomial function of $\{k\}$ that multiplies $\log(-k_T\eta_0) + i\pi/2$ in the wavefunction and can only appear when the overall number of derivatives in bulk interactions is relatively small, which we will make precise in Section 8.4. This is consistent with the observation that the parity-odd Weyl-cubed vertex yields a vanishing bispectrum in dS space [171, 240, 251]. In this case there are too many derivatives for a logarithm to appear in the wavefunction. Related observations about the consequences of unitarity cuts were recently made in [92]. We summarise these results in Table 8.1 and remind the reader that the above discussion applies to contact diagrams, as relevant for this work. In Section 8.4 we provide

⁵Although here our proof was outlined in D = 4 spacetime dimensions, a generalised version of the MLT [263] applies in all other dimensions and with this generalised MLT and the COT, one can show that $\eta_0 = 0$ poles never appear in correlators. We thank Harry Goodhew for discussions on this point.

a full analysis of the form of the wavefunction for graviton cubic interactions and one can then use the results of this section to extract the contributions to the bispectra.

Before proceeding we would like to comment on what happens for tree-level contributions to the wavefunction that are not contact but include some exchange interaction (a bulk-bulk propagator in the bulk representation). In that case, two things change: (i) the expression for the correlator in terms of wavefunction coefficients in (8.72) acquires additional contributions and (ii) the right-hand side of the Cosmological Optical Theorem (COT) does not vanish anymore [12]. Notice that both of these additional contributions are not singular as $k_T \rightarrow 0$. Hence, one can still conclude that any term in the wavefunction that is invariant under $\{k\} \rightarrow \{-k\}$ cannot contribute to the part of the correlator that is singular as $k_T \rightarrow 0$. Unfortunately, the wavefunction coefficients can become quite complicated for general exchange diagrams and we did not find a simple rule to establish when ψ_n^{trimmed} is invariant under $\{k\} \rightarrow \{-k\}$.

$$\begin{array}{c|ccc} k_T \text{ poles } & \log(-k_T\eta_0) + \frac{i\pi}{2} & \eta_0 \text{ poles} \\ \hline \text{even } \alpha & \checkmark & \checkmark & (\text{only the log}) & \checkmark \\ \text{odd } \alpha & \bigstar & \checkmark & (\text{only the } i\pi) & \checkmark \end{array}$$

Table 8.1: In this table we indicate which parts of the trimmed wavefunction, arising from contact diagrams, can contribute to cosmological correlators and which cannot. Here α is the number of spatial derivatives contracted with polarizations tensors, as defined in (8.38), and these results apply for three spatial dimensions, d = 3.

8.4 Bootstrapping *all* graviton bispectra

In this section we bootstrap boost-breaking graviton bispectra at tree-level. We detail the general method that allows one to extract bispectra for any helicity configuration, and up to any desired order in derivatives. Throughout we employ the Boostless Bootstrap Rules and Manifestly Local Test, which were both reviewed in Section 8.2.

8.4.1 Polarisation factors

It is the presence of spin-2 polarization tensors that distinguishes graviton bispectra from any other. As we reviewed in Section 8.2, we write a general three-point wavefunction coefficient in terms of a polarisation factor multiplied by a "trimmed" wavefunction coefficient ψ_3^{trimmed} which is an SO(3) scalar. We have [231]

$$\psi_{3}^{h_{1},h_{2},h_{3}}(\mathbf{k}_{1},\mathbf{k}_{2},\mathbf{k}_{3}) = \sum_{\text{contractions}} \left[e^{h_{1}}(\mathbf{k}_{1})e^{h_{2}}(\mathbf{k}_{2})e^{h_{3}}(\mathbf{k}_{3})\mathbf{k}_{1}^{\alpha_{1}}\mathbf{k}_{2}^{\alpha_{2}}\mathbf{k}_{3}^{\alpha_{3}} \right] \psi_{3}^{\text{trimmed}}(k_{1},k_{2},k_{3}),$$
(8.80)

where $h_a = \pm 2$ are the helicities of the external fields, and we remind the reader that we define the total number of *spatial* momenta as $\alpha = \alpha_1 + \alpha_2 + \alpha_3$. Here index contractions between the momenta and polarization tensors are left implicit, and indeed our first goal is to construct all of the possible polarisation factors. As we explained in Section 8.2, the trimmed wavefunction is constrained by the Manifestly Local Test (MLT) [33] and the Cosmological Optical Theorem (COT) [12], and so with the polarisation factors at hand, we will solve the MLT and obtain the complete three-point functions.

We first note that we can restrict our attention to $\alpha \leq 7$. This is because in order to construct an SO(3)-invariant object, we need to contract momenta with one of

$$e_{i_1i_2}^{h_1} e_{i_3i_4}^{h_2} e_{i_5i_6}^{h_3} \quad \text{or} \quad \epsilon_{i_1i_2i_3} e_{i_4i_5}^{h_1} e_{i_6i_7}^{h_2} e_{i_8i_9}^{h_3} \,, \tag{8.81}$$

where the presence of a Levi-Civita tensor tells us that the resulting graviton bispectrum will violate parity. All remaining contractions are made with δ_{ij} and from now on we omit the dependence of polarization tensors on momenta for simplicity of notation. Now, it is straightforward to see that α can be at most 6 in the parity-even case, with all six polarisation indices contracted with momenta, and 7 in the parity-odd case since we can have at most two spatial momenta contracted with the Levi-Civita tensor due to momentum conservation. We will deal with the parity-even and parity-odd cases separately.

As is the case for scattering amplitudes, graviton bispectra are most compactly presented using the spinor helicity formalism rather than polarisation tensors. Indeed, this was the view advocated in [171] and is the route we will follow in this chapterÅ virtue of the spinor helicity formalism is that it can easily highlight possible degeneracies that could be hidden when using polarization tensors. Unfortunately, we do not have the means to construct the full structure of all allowed polarisation factors directly using spinors, so the approach we will take is to write down all possible polarisation

factors in terms of polarisation tensors, with potential degeneracies still present, and to then convert these expressions into the spinor helicity formalism, where all degeneracies are manifest and can be easily eliminated.

We initially focus on the + + + helicity configuration, and in the following subsection we will show how to easily obtain the polarisation factors for all the other helicity configurations (+ + -, - - + and - - -) from this + + + building block. The helicity scaling of the external fields tells us that all + + + polarisation factors must contain

$$[12]^2 [23]^2 [31]^2, (8.82)$$

as an overall factor. This is the same factor that appears in three-point scattering amplitudes of massless gravitons [28, 2] and is unique for this helicity configuration. The symmetries of the wavefunction then ensure that this can only be multiplied by SO(3) invariant quantities that are simply functions of the three external energies. As explained in Section 8.2.5, whenever we convert a polarisation tensor into an expression with spinor brackets, we gain two powers of the corresponding energy in the denominator of the wavefunction. It is therefore not merely (8.82) that appears as an overall factor, but actually the *dimensionless* quantity

$$SH_{+++} = \frac{[12]^2 [23]^2 [31]^2}{e_3^2}, \qquad (8.83)$$

where $e_3 = k_1 k_2 k_3$ is the third elementary symmetric polynomial. The above factor is ever-present. The information about the specific contraction is contained in an additional factor which is a function of the energies and which we denote as $h_{\alpha}(k_1, k_2, k_3)$. This is always a polynomial of degree α . Finally, this product can be multiplied by the trimmed wavefunction, which in the bulk representation arises from bulk time integrals. This general form is true before we sum over all possible permutations, so the final form of the three-point function is

$$\psi_3^{+++}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \frac{[12]^2 [23]^2 [31]^2}{e_3^2} \sum_{\text{permutations}} h_\alpha(k_1, k_2, k_3) \psi_3^{\text{trimmed}}(k_1, k_2, k_3) , \quad (8.84)$$

where the sum over permutations ensures that the final expression is invariant under the exchange of any two external fields and their momenta, as dictated by Bose symmetry. In Appendix 8.7.1 we

257

construct all possible polarisation factors using polarisation tensors. With repeated use of (8.62), and recalling the definition of $I_a = k_T - 2k_a$, we find the following general structures for the + + + polarisation factors:

$$h_0 = 1,$$
 (8.85)

$$h_1 = ik_1 \text{ and perms}, \qquad (8.86)$$

$$h_2 = k_1^2$$
 and perms, $k_1 k_2$ and perms, (8.87)

$$h_3 = ik_1^3$$
 and perms, $ik_1^2k_2$ and perms, $ik_1k_2k_3$, (8.88)

$$h_4 = I_1^2 I_2 I_3 \text{ and perms},$$
 (8.89)

$$h_5 = iI_1^3 I_2 I_3$$
 and perms, $iI_1^2 I_2^2 I_3$ and perms, (8.90)

$$h_6 = I_1^2 I_2^2 I_3^2 \,, \tag{8.91}$$

$$h_7 = i I_1^3 I_2^2 I_3^2$$
 and perms, (8.92)

where in some cases we have only presented one of the possible permutations, but we should keep in mind that one needs to sum over permutations in the final expression. For odd α we have included overall factors of *i* which arise from the Levi-Civita tensor as shown in Appendix 8.7.1. Note that, if we only use spinor helicity variables, we do not have the means to derive the full form of the polarisation factors: for example, we did not find a good reason why a term like I_1^7 would be prohibited in the case of $\alpha = 7$. This was the main rationale for invoking polarization tensors in our argument, although it would be very interesting to derive the above list of structures, and to understand why some terms are not permitted, directly using spinors.

As we have explained in Sections 8.2 and 8.3, the general form of the trimmed wavefunction can be fixed by a set of Boostless Bootstrap Rules [231]. A combination of symmetries (including scale invariance), a weak-coupling approximation and Bunch-Davies initial conditions, ensures that the trimmed part of the wavefunction takes the form

$$\psi_{3}^{\text{trimmed}}(k_{1},k_{2},k_{3}) = \frac{\text{Poly}_{3+p-\alpha}(k_{1},k_{2},k_{3})}{k_{T}^{p}} + \text{Poly}_{3-\alpha}(k_{1},k_{2},k_{3})\log(-k_{T}\eta_{0}) + \frac{\text{Poly}_{2-\alpha}(k_{1},k_{2},k_{3})}{\eta_{0}} + \frac{\text{Poly}_{1-\alpha}(k_{1},k_{2},k_{3})}{\eta_{0}^{2}} + \frac{\text{Poly}_{-\alpha}(k_{1},k_{2},k_{3})}{\eta_{0}^{3}},$$
(8.93)

where we remind the reader that the degree of these complex polynomials is indicated by the subscripts. For those terms that diverge as $\eta_0 \rightarrow 0$, we have strong restrictions on the allowed values of α : a $1/\eta_0^q$ singularity can only arise for $\alpha \leq 3 - q$, which also justifies truncating the expansion at q = 3. The above general form of the trimmed wavefunction is then further constrained by the MLT, which must be satisfied for all external energies. Note that we impose the MLT *before* we sum over permutations in (8.84), and so in that formula, each $\psi_3^{\text{trimmed}}(k_1, k_2, k_3)$ is a solution to the MLT. The general recipe for constructing a + + wavefunction coefficient is therefore the following:

- 1. Write down the spinor helicity factor SH_{+++} and multiply it by one of the above choices for $h_{\alpha}(k_1, k_2, k_3)$.
- 2. Multiply this polarisation factor by a trimmed wavefunction coefficient of the form (8.93) where the polynomials in this ansatz have been constrained by the MLT (8.39). Note that for computational purposes it is useful to choose the permutation symmetry of this trimmed part to be the same as that of the polarisation factor. For example, if the polarisation factor is symmetric in the exchange of k_2 and k_3 then the trimmed part should be too, while if the polarisation factor has no symmetry then the trimmed part shouldn't either.
- 3. Use the COT (8.55) to deduce if unitarity demands real or imaginary coefficients.
- 4. Finally, sum over the remaining permutations such that the final wavefunction coefficient is fully symmetric, as dictated by Bose symmetry (Rule 4 of [231]).
- 5. To extract the corresponding three-point correlators, we use the results of Section 8.3. For even α we take the rational and log terms, with real coefficients, and divide by the appropriate powers of the power spectrum. For odd α , we take the log part and simply replace the log with $i\pi/2$ such that we have some polynomial multiplied by a polarisation factor. Finally, we divide by the appropriate powers of the power spectrum. In both cases the result is real since for even α the polarisation factor is real, and is multiplied by a real function of the energies, while for odd α the polarisation factor is imaginary but it is multiplied by an imaginary function of the energies.

8.4.2 +++ to rule them all

Before we constrain these + + + wavefunctions further, let us first show how we can obtain the + + -, - - + and - - - helicity configurations if $h_{\alpha}(k_1, k_2, k_3)$ and $\psi_3^{\text{trimmed}}(k_1, k_2, k_3)$ are known. It might be tempting to go back to the beginning, i.e. to the polarization tensors, and derive the spinor helicity form of tensor structures independently for each configuration. However, this is not necessary as the spinor variables can do most of the work for us. Let us first construct the + + - tensor structures in spinor helicity variables directly from the + + + ones. Flipping the helicity of the third graviton is equivalent to sending its energy from k_3 to $-k_3$ while keeping its momentum fixed. Under this transformation, the spinors transform according to [2]

$$\tilde{\lambda} \mapsto i(\lambda_2, -\lambda_1) , \quad \lambda \mapsto i(-\tilde{\lambda}_2, \tilde{\lambda}_1) .$$
 (8.94)

Using the definitions of the various brackets given in Section 8.2.5, we then have

$$[13] \mapsto -i(31),$$
 (8.95)

$$[23] \mapsto -i(32) ,$$
 (8.96)

from which it follows that

$$SH_{+++} \mapsto \frac{[12]^2}{e_3^2} (31)^2 (32)^2 = \frac{[12]^6}{[23]^2 [31]^2} \frac{I_1^2 I_2^2}{e_3^2} \equiv SH_{++-} .$$
(8.97)

So all + + - wavefunction coefficients are multiplied by this common factor of SH₊₊₋. Note the square brackets are completely fixed by the helicities of the external fields and are the same as for amplitudes [2, 28], while the ever-present $I_1^2 I_2^2$ factor in the numerator is required for the absence of divergences. Indeed, consider the following argument: with the help of (8.58), the spinor helicity factor $\frac{[12]^6}{[23]^2[31]^2}$ can be rewritten as

$$\frac{[12]^6 \langle 23 \rangle^2 \langle 31 \rangle^2}{k_T^4 I_1^2 I_2^2}.$$
(8.98)

If the momenta are allowed to be complex, then I_1 can be taken to zero while keeping k_T , I_2 and the numerator finite. Such a divergence is forbidden and therefore we should include two factors of I_1 in the numerator to cancel it out. The absence of a divergence as I_2 is taken to zero similarly demands

that we should include two factors of I_2 . This argument can be easily generalised to other helicities to show that in general the bispectrum of any three fields with helicities h_a for a = 1, 2, 3 has to contain the following factor (for $H \equiv h_1 + h_2 + h_3 \ge 0$)

$$SH_{h_1,h_2,h_3} = \frac{[12]^{d_3}[23]^{d_1}[31]^{d_2}}{\prod_{a=1}^3 k_a^{|h_a|}} \prod_{b=1}^3 I_b^{\max[0,-d_b]}, \qquad (8.99)$$

where

$$d_a \equiv h_b + h_c - h_a = H - 2h_a \quad (a \neq b \neq c).$$
 (8.100)

The scaling dimension of the spinor helicity factor SH_{h_1,h_2,h_3} is max $\{0, -d_1, -d_2, -d_3\}$. The wavefunction coefficient then takes the form

$$\psi_3^{h_1,h_2,h_3}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = \mathbf{SH}_{h_1,h_2,h_3} \times P_m(k_1,k_2,k_3)$$
(8.101)

where P_m is a rational function of the energies (possibly also including $\log(-k_T\eta_0)$ multiplied by a polynomial) and m is its scaling dimension.

To extract the + + - wavefunction, then, we take SH₊₊₋ and multiply it by $h_{\alpha}(k_1, k_2, -k_3)$ and by $\psi_3^{\text{trimmed}}(k_1, k_2, k_3)$. Note that only in h_{α} is the sign of k_3 flipped. Indeed, the structure of h_{α} is fixed by the form of the polarisation factor which certainly depends on the helicity configuration, whereas $\psi_3^{\text{trimmed}}(k_1, k_2, k_3)$ is a product of time integrals in the bulk formalism and is therefore independent of the helicity configuration of the external fields. Therefore, the + + - wavefunction coefficients are given by

$$\psi_3^{++-}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \frac{[12]^6}{[23]^2 [31]^2} \frac{I_1^2 I_2^2}{e_3^2} \sum_{\text{permutations}} h_\alpha(k_1, k_2, -k_3) \psi_3^{\text{trimmed}}(k_1, k_2, k_3) \,. \quad (8.102)$$

The recipe we outlined above for the + + + configuration is then easily applied to this + + - case, with the symmetries of $\psi_3^{\text{trimmed}}(k_1, k_2, k_3)$ fixed by $h_\alpha(k_1, k_2, -k_3)$ and with the final sum over permutations ensuring that the final wavefunction is symmetric under the exchange of k_1 and k_2 , as dictated by Bose symmetry.

Finally, the --+ and --- wavefunction coefficients are then obtained directly from the ++and +++ ones respectively, by sending $k_a \mapsto -k_a$ for a = 1, 2, 3. This corresponds to all square brackets changing into (minus) angle brackets, such that

$$\mathrm{SH}_{+++} \mapsto \frac{\langle 12 \rangle^2 \langle 23 \rangle^2 \langle 31 \rangle^2}{e_2^2} \equiv \mathrm{SH}_{---},$$
 (8.103a)

$$\mathrm{SH}_{++-} \mapsto \frac{\langle 12 \rangle^6}{\langle 23 \rangle^2 \langle 31 \rangle^2} \frac{I_1^2 I_2^2}{e_3^2} \equiv \mathrm{SH}_{--+} \,.$$
 (8.103b)

Under $k_a \mapsto -k_a$, we have $h_{\alpha}(k_1, k_2, k_3) \mapsto (-1)^{\alpha} h(k_1, k_2, k_3)$, while $\psi_3^{\text{trimmed}}(k_1, k_2, k_3)$ is again taken to be unchanged.

In conclusion, with knowledge of the building blocks of the + + + wavefunction coefficients, one can easily compute wavefunction coefficients for other helicity configurations. We note that our ability to do this is due to fact that time translations are no longer a symmetry in cosmology and therefore square, angle and round brackets are related as shown in Section 8.2.5. For scattering amplitudes, where time translations are a symmetry, one cannot simply map between different configurations in this way. As a very non-trivial check of this procedure, we verified that the + + + wavefunction coefficient arising from a parity-even Weyl³ vertex in the bulk gives rise to a vanishing + + - coefficient, as it should [171].

8.4.3 A further simplification of the polarisation factors

Now given that $h_{\alpha}(k_1, k_2, k_3)$ must be multiplied by a solution to the MLT, we can actually further simplify the structures given in (8.85) to (8.92). The general h_{α} in (8.84) is given by an arbitrary linear combination of polynomials listed in (8.85)-(8.92), as well as all their permutations, for each α . However, now we will show that we may consider only a few special h_{α} and still obtain fully general wavefunction coefficients. We give an explicit argument for $\alpha = 2$, but a closely analogous argument works for any α .

We have already established that $h_2(k_1, k_2, k_3) = \sum_a n_a k_a^2 + \sum_a m_a k_a k_{a+1}$, where n_a, m_b are arbitrary numerical coefficients. We then have (recall that $\tilde{\psi}_3$ is a shorthand notation for ψ_3^{trimmed}):

$$\frac{\psi_{3}^{+++}(\mathbf{k}_{1},\mathbf{k}_{2},\mathbf{k}_{3})}{\mathrm{SH}_{+++}} = \sum_{\sigma \in S_{3}} \sum_{a} \left(n_{a}k_{\sigma(a)}^{2} + m_{a}k_{\sigma(a)}k_{\sigma(a+1)} \right) (\tilde{\psi}_{3} \circ \sigma)(k_{1,2,3})
= \sum_{a} \sum_{\sigma \in S_{3}} \left(n_{\sigma^{-1}(a)}k_{a}^{2} + m_{\sigma^{-1}(a)}k_{a}k_{a+1} \right) (\tilde{\psi}_{3} \circ \sigma)(k_{1,2,3})
= \sum_{a} \left(k_{a}^{2} \sum_{\sigma \in S_{3}} n_{\sigma^{-1}(a)} (\tilde{\psi}_{3} \circ \sigma)(k_{1,2,3}) + k_{a}k_{a+1} \sum_{\sigma \in S_{3}} m_{\sigma^{-1}(a)} (\tilde{\psi}_{3} \circ \sigma)(k_{1,2,3}) \right)^{(8.104)}
= \sum_{a} \left(k_{a}^{2} \sum_{\sigma \in S_{3}} n_{\sigma^{-1}(a)} (\tilde{\psi}_{3} \circ \sigma)(k_{1,2,3}) + k_{a}k_{a+1} \sum_{\sigma \in S_{3}} m_{\sigma^{-1}(a)} (\tilde{\psi}_{3} \circ \sigma)(k_{1,2,3}) \right)^{(8.104)}
= \sum_{cyclic} k_{1}^{2} f_{(23)}(k_{1},k_{2},k_{3}) + \sum_{cyclic} k_{1}k_{2}g_{(12)}(k_{1},k_{2},k_{3}),$$

where $f_{(23)}$ and $g_{(12)}$ are linear combinations of trimmed wavefunction coefficients, and therefore they must take the form given in (8.93) and satisfy the MLT. Moreover, we employ the notation that a function of the three external energies is symmetric under the exchange of energies indicated in a subscript e.g. $f_{(23)}$ is symmetric under the exchange of k_2 and k_3 , while $f_{(123)}$ would be fully symmetric. An analogous argument can be used to show that $\psi_3^{++-}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ can be simplified in the same way. More precisely, we have

$$\frac{\psi_3^{++-}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)}{\mathrm{SH}_{++-}} = \sum_{\text{cyclic}} k_1^2 f_{(23)}(k_1, k_2, k_3) + k_1 k_2 g_{(12)}(k_1, k_2, k_3) \\ - k_2 k_3 g_{(23)}(k_2, k_3, k_1) - k_3 k_1 g_{(31)}(k_3, k_1, k_2).$$
(8.105)

Thus, we see that we can take $h_2(k_a)$ to be a linear combination of k_1^2 and k_1k_2 and still get a fully general $\alpha = 2$ solution. Moreover, we note that all solutions constructed from $h_2(k_a) = k_1^2$ can also be constructed using the $\alpha = 0$ polarization factor $h_0(k_a) = 1$. This is because, if $f_{(23)}(k_1, k_2, k_3)$ satisfies the MLT, then $k_1^2 f_{(23)}(k_1, k_2, k_3)$ must satisfy it too, so wavefunction coefficients of the form

$$\frac{\psi_3^{+++}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)}{\mathrm{SH}_{+++}} = \sum_{\text{cyclic}} k_1^2 f_{(23)}(k_1, k_2, k_3), \qquad (8.106)$$

$$\frac{\psi_3^{++-}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)}{\mathrm{SH}_{++-}} = \sum_{\text{cyclic}} k_1^2 f_{(23)}(k_1, k_2, k_3), \qquad (8.107)$$

are already accounted for and contained in the MLT solutions for polarisation factors with $\alpha = 0$. Assuming we construct solutions iteratively with increasing α , so that $\alpha = 0$ wavefunction coefficients have already been constructed, for $\alpha = 2$ we only need to consider $h_2(k_1, k_2, k_3) = k_1k_2$ to derive a complete set of such coefficients. One can proceed in a similar manner at each order in α by studying the different allowed h_{α} and asking if the resulting wavefunction coefficients have already been captured by lower order solutions in α . We find that to construct fully general wavefunction coefficients, it is sufficient to consider the following polarisation factors for the + + + helicity configuration:

$$h_0 = 1$$
, (8.108)

$$h_1 = ik_1,$$
 (8.109)

$$h_2 = k_2 k_3 \,, \tag{8.110}$$

$$h_3 = i I_1 I_2 I_3 \,, \tag{8.111}$$

$$h_4 = I_1^2 I_2 I_3 \,, \tag{8.112}$$

$$h_{5a,b} = iI_1^3 I_2 I_3, iI_1 I_2^2 I_3^2, \qquad (8.113)$$

$$h_6 = I_1^2 I_2^2 I_3^2 \,, \tag{8.114}$$

$$h_7 = i I_1^3 I_2^2 I_3^2 \,. \tag{8.115}$$

Therefore, at each order in α we have to consider a single polarisation factor, apart from $\alpha = 5$ where there are two possible structures. Note that in all cases we can write the polarisation factor in such a way that it is symmetric in the kinematical data of two out of the three external fields, which we take to be fields 2 and 3. We can now follow the recipe outlined above, and constrain the remaining part of the wavefunction coefficients with the MLT.

8.4.4 Constraining the trimmed wavefunction

We now turn to the final piece of the puzzle, which requires us to solve the MLT (8.39) to constrain (8.93) and therefore construct the trimmed part of the wavefunction coefficients. By writing out the allowed form of the polynomials in this ansatz we have

$$\psi^{\text{trimmed}}(k_1, k_2, k_3) = \frac{1}{k_T^p} \sum_{l+m+n=3+p-\alpha} c_{lmn} k_1^l k_2^m k_3^n + \log\left(-k_T \eta_0\right) \sum_{l+m+n=3-\alpha} d_{lmn} k_1^l k_2^m k_3^n + \frac{1}{\eta_0} \sum_{l+m+n=2-\alpha} e_{lmn} k_1^l k_2^m k_3^n + \frac{1}{\eta_0^2} \sum_{l+m+n=1-\alpha} f_{lmn} k_1^l k_2^m k_3^n + \frac{1}{\eta_0^3} \sum_{l+m+n=-\alpha} g_{lmn} k_1^l k_2^m k_3^n ,$$

$$(8.116)$$

where $l, m, n \ge 0$ and we remind the reader that the sums are fixed by scale invariance. The following conditions are then necessary for the above ansatz to pass the MLT:

$$d_{1,n,r-1} = 0, (8.117)$$

$$d_{1,n,r-1} = 0, \qquad (8.117)$$

$$\sum_{m} {p \choose n-m} d_{0,m,3-\alpha-m} = p c_{0,n,p+r} - c_{1,n-1,p+r} - c_{1,n,p+r-1}, \qquad (8.118)$$

$$e_{1,n,r-2} = f_{1,n,r-3} = g_{1,n,r-4} = 0, \qquad (8.119)$$

with $r \equiv 3 - \alpha - n$; along with analogous conditions for all other permutations of indices. Note that the conditions that arise from the terms in the first line of (8.116) decouple from those in the second line.

Whenever a polynomial h_{α} has a symmetry under interchange of external labels, the trimmed wavefunction coefficient may also be assumed to have such a symmetry without loss of generality. This is because any non-symmetric part will be cancelled out after summing over all permutations indicated in (8.84), as we saw explicitly in the previous section in the $\alpha = 2$ case. Therefore, if $h_{\alpha}(k_1, k_2, k_3) = h_{\alpha}(k_1, k_3, k_2)$, then we have

÷

$$c_{lmn} = c_{lnm}, \qquad (8.120)$$

$$d_{lmn} = d_{lnm}, \qquad (8.121)$$

Moreover, if $h_{\alpha}(k_1, k_2, k_3)$ is completely symmetric, then we have

$$c_{lmn} = c_{lnm} = c_{mln} , \qquad (8.122)$$

$$d_{lmn} = d_{lnm} = d_{mln} ,$$
 (8.123)

As we saw above, in all cases h_{α} is symmetric in at least two external labels. We will now present the first few solutions for each α , considering even and odd α separately.

Parity-even interactions We begin with parity-even interactions which have even α .

 $\alpha = 0$ In this case we have $h_0 = 1$ and so the solution to the MLT must be fully symmetric. This case is actually exactly the same as the situation for three identical scalars which was covered in [33]. The following solutions are therefore the same as those found in that work. Given the symmetry, we present the results using the three elementary symmetric polynomials k_T , e_2 , e_3 . Up to p = 3 we have

$$\eta_0^{-1}: \quad \frac{i(k_T^2 - e_2)}{\eta_0} \,, \tag{8.124}$$

$$\eta_0^{-3}: \quad \frac{i}{\eta_0^3}, \tag{8.125}$$

$$p = 0: \quad 4e_3 - e_2k_T + (k_T^3 - 3k_Te_2 + 3e_3)\log(-k_T\eta_0), \quad k_T^3 - 3k_Te_2 + 3e_3, \qquad (8.126)$$

$$p = 2: \quad \frac{e_2 e_3 + e_2^2 k_T - 2e_3 k_T^2}{k_T^2}, \tag{8.127}$$

$$p = 3: \quad \frac{e_3^2}{k_T^3}, \tag{8.128}$$

where, as indicated, there are two possible solutions for p = 0.

Unitarity places the following additional constraints. The coefficients of $1/\eta_0$ and $1/\eta_0^3$ must be imaginary as consequence of the Cosmological Optical Theorem (COT), see Section 8.3. This has a nice interpretation in terms of the holographic language of (A)dS/CFT, along the lines of [106]. These two terms are bulk IR divergences and should be holographically renormalized as described in [264]. For the associated renormalization group flow to be unitary, these divergences should be imaginary, which is precisely what the COT ensures. Conversely, the COT says that the coefficient of the $1/\eta_0^2$ divergence must be real. This would correspond to a counterterm with imaginary coupling constant. It is quite intriguing that the MLT forbids precisely these terms and we will discuss this elsewhere. For p = 0 the MLT admits two solutions. The second one does not have a log and can satisfy the COT by itself with an arbitrary real coefficient. Conversely, the first solution, which contains a log, satisfies the COT only when it is combined with the second solution with a relative factor of $i\pi/2$ (see Section 8.3 or [12]), namely in the combination

$$\lambda \left[4e_3 - e_2k_T + (k_T^3 - 3k_T e_2 + 3e_3) \log(-k_T \eta_0) + i\frac{\pi}{2} \left(k_T^3 - 3k_T e_2 + 3e_3 \right) \right], \qquad (8.129)$$

for real λ .

There are no p = 1 solutions. This can simply be understood as follows. Recall that for cubic wavefunction coefficients, the degree p of the leading k_T pole equals the number of derivatives. Then the absence of solutions for p = 1 is related to the fact that there are no single derivative interactions one can write down (for $\alpha = 0$), other than a total time derivative. Wavefunction coefficients with p = 1 do arise in not-manifestly-local theories. Indeed the scalar bispectrum induced by gravity has p = 1, which is consistent with the above discussion because, after integrating out the non-dynamical parts of the metric, GR displays not-manifestly-local interactions. We refer the reader to [33] for more details on this case. Note that the COT fixes the coefficients of the terms rational in $\{k\}$ to be real.

 $\alpha = 2$ In this case, we have $h_2(k_1, k_2, k_3) = k_2 k_3$ without loss of generality, so we take the ansatz to also be symmetric in k_2 and k_3 . The leading solutions are then

$$\eta_0^{-1}: \quad \frac{i}{\eta_0} \,, \tag{8.130}$$

$$p = 2: \quad \frac{e_3 + e_2 k_T - k_T^3}{k_T^2}, \tag{8.131}$$

$$p = 3: \quad \frac{k_1^2 \left(k_1^2 + 3k_1k_{23} + 2(k_{23}^2 + k_2k_3)\right)}{k_T^3}, \quad (8.132)$$

$$p = 4: \quad \frac{k_2^2 k_3^2 \left(k_T + 3k_1\right)}{k_T^4}, \tag{8.133}$$

We see that only a simple
$$\eta_0 = 0$$
 pole is allowed with a constant and imaginary residue, by unitarity.
In terms of total-energy poles, the leading solution has a degree two pole which is related to the fact that such wavefunction coefficients arise from bulk vertices with at least two derivatives. Again, the COT demands that the coefficients of these rational terms are real.

÷

 $\alpha = 4$. Here we again have a single choice for the polarisation factor which is $h_4(k_1, k_2, k_3) = I_1^2 I_2 I_3$. This must be combined with an $\alpha = 4$ solution to the MLT, which is symmetric in k_2 and k_3 . Clearly no $\eta_0 = 0$ poles are allowed, and the leading solutions are

$$p = 4: \quad \frac{3e_3 + k_T e_2 + k_T^3}{k_T^4}, \tag{8.134}$$

$$p = 5: \quad \frac{k_1^2 \left(k_1^2 + 5k_1k_{23} + 4(k_{23}^2 + 3k_2k_3)\right)}{k_T^5}, \quad (8.135)$$

Again we see that the lowest possible total energy pole has degree 4.

 $\alpha = 6$. Finally, we have $h_6 = I_1^2 I_2^2 I_3^2$. This is fully symmetric, so we can present solutions to the MLT using the elementary symmetric polynomials. There are no $\eta_0 = 0$ poles and the leading solutions are

$$p = 6: \quad \frac{15e_3 + 3k_T e_2 + k_T^3}{k_T^6}, \tag{8.136}$$

$$p = 8: \quad \frac{7e_2e_3 + k_Te_2^2 - 2k_T^2e_3}{k_T^8}, \tag{8.137}$$

In each case, solutions with higher-order k_T poles can be easily computed. We see that an IR-divergent logarithm is only permitted for $\alpha = 0$, while IR-divergences in the form of $\eta_0 = 0$ poles can only arise for $\alpha = 0, 2$ and they always come with imaginary coefficients. In Section 8.4.5 we will use these solutions to write down the final form of the leading + + + and + + - bispectra.

Parity odd interactions We now turn to odd α which correspond to parity-odd interactions.

 $\alpha = 1$ In this case we have $h_1(k_1, k_2, k_3) = k_1$. This must be combined with an $\alpha = 1$ solution to the MLT, symmetric in k_2 and k_3 . The leading solutions are

$$\eta_0^{-2}: \quad \frac{1}{\eta_0^2} \,, \tag{8.138}$$

$$p = 0: \quad k_1^2, \ k_T^2 - 2e_2, \tag{8.139}$$

$$p = 1: \quad \frac{2e_3 - e_2k_T}{k_T} + (k_T^2 - 2e_2)\log(-k_T\eta_0), \quad (8.140)$$

$$p = 2: \quad \frac{k_1^2 \left(k_T (k_2 + k_3) + k_2 k_3\right)}{k_T^2} - k_1^2 \log(-k_T \eta_0), \quad (8.141)$$

$$p = 3: \quad \frac{-2e_3k_T^2 + 2e_3e_2 + k_Te_2^2}{k_T^3}, \tag{8.142}$$

÷

where, as indicated, there are two possible solutions for p = 0. We see that the only allowed $\eta_0 = 0$ pole is of degree two, as it should be for $\alpha = 1$ because of scale invariance. Interestingly, we also see that IR-divergent logarithms are also permitted but *only* in combination with total-energy poles. This is in contrast to even α where logarithms could contribute as the only singular term. The solutions with higher total-energy poles that are not shown here do not have logarithms.

Unitarity places the following additional constraints. All terms without logs can appear with real coefficients. The two solutions containing a log, namely p = 1 and p = 2, solve the Cosmological Optical Theorem (COT) only when accompanied by a corresponding p = 0 solution with a relative coefficient of $i\pi/2$, namely in the combinations

$$\frac{2e_3 - e_2k_T}{k_T} + (k_T^2 - 2e_2) \left[\log(-k_T\eta_0) + i\frac{\pi}{2} \right]$$
(8.143)

$$\frac{k_1^2 \left(k_T (k_2 + k_3) + k_2 k_3\right)}{k_T^2} - k_1^2 \left[\log(-k_T \eta_0) + i\frac{\pi}{2}\right],\tag{8.144}$$

with real overall coefficients. Notice that, since we are considering parity-odd interactions, it is only the imaginary part of these trimmed wavefunction coefficients, namely that proportional to $i\pi/2$, that contributes to the bispectrum.

 $\alpha = 3$ Here we can choose $h_3(k_1, k_2, k_3) = I_1 I_2 I_3$ and the solution to the MLT may be assumed to be fully symmetric. No $\eta_0 = 0$ poles are allowed, and the leading solutions are

$$p = 0: 1,$$
 (8.145)

$$p = 3: \quad \frac{2e_3 + e_2k_T}{k_T^3} - \log(-k_T\eta_0), \qquad (8.146)$$

$$p = 5: \quad \frac{4e_2e_3 + e_2^2k_T - 2e_3k_T^2}{k_T^5}, \qquad (8.147)$$

Again the higher order solutions do not contain logarithms, so only a single solution with such a IR-divergence is allowed in this case. As above, unitarity in the form of the Cosmological Optical

Theorem (COT) requires that the p = 3 term, which contains a log, must appear together with the (trivial) p = 0 solution in the combination

$$\frac{2e_3 + e_2k_T}{k_T^3} - \left[\log(-k_T\eta_0) + i\frac{\pi}{2}\right], \qquad (8.148)$$

with a real overall coefficient.

 $\alpha = 5$. In this penultimate case there are two choices for h_5 : $h_5(k_1, k_2, k_3) = I_1^3 I_2 I_3$ and $h_5(k_1, k_2, k_3) = I_1 I_2^2 I_3^2$. Both must be multiplied by a solution to the MLT that is symmetric in k_2 and k_3 . No $\eta_0 = 0$ poles or logarithmic terms are allowed, and the leading solutions are

$$p = 5: \quad \frac{8e_3 + 2k_T e_2 + k_T^3}{k_T^5}, \tag{8.149}$$

$$p = 6: \frac{k_1^2 \left(k_1^2 + 6k_1k_{23} + 5(k_{23}^2 + 4k_2k_3)\right)}{k_T^6}, \qquad (8.150)$$

both with real coefficients by unitarity.

 $\alpha = 7$. In this final case, we have $h_7(k_1, k_2, k_3) = I_1^3 I_2^2 I_3^2$ and so the solution to the MLT needs to be symmetric k_2 and k_3 . The leading solutions are

$$p = 7: \quad \frac{24e_3 + 4k_T e_2 + k_T^3}{k_T^7}, \tag{8.151}$$

$$p = 8: \quad \frac{k_1^2 \left(k_1^2 + 8k_1(k_2 + k_3) + 7((k_2 + k_3)^2 + 6k_2k_3)\right)}{k_T^8}, \quad (8.152)$$

Again, higher order solutions are easily found. As we have emphasised a number of times, only the coefficients of the logarithms contribute to the final bispectra for these parity-odd interactions. We have found only three solutions with logarithms which are also required to come alongside total-energy poles which will ultimately drop out from the correlator. It is important to stress that the fact we only have three logarithmic terms is true to all orders in derivatives. Indeed, all remaining solutions not explicitly shown above are purely rational. We can therefore extract the full form of parity-odd

graviton bispectra, to all orders in derivatives, from these MLT solutions. Given that there is only a single polarisation structure for $\alpha = 1, 3$, there are only three independent parity-odd graviton bispectra. We will discuss this further in Section 8.4.5 where we construct the final form of the correlators.

Contact reconstruction formula In this section we have derived wavefunction coefficients for graviton interactions without any reference to flat space. However, there also exists a well-defined relationship between wavefunction coefficients in de Sitter and scattering amplitudes in flat space: the residue of the leading total-energy pole of a wavefunction coefficient contains the flat space amplitude (see also [185, 92, 265] for additional relations between correlators and amplitudes). This was first noticed in [171, 172] and then an explicit formula was derived in [12]. For n external fields the relationship is

$$\psi_n = (p-1)!(iH)^{p-n-1} \frac{e_n A_n^{(p-n+3)}}{k_T^p} + \dots, \qquad (8.153)$$

where $e_n = \prod_{a=1}^n k_a$ is a product of the *n* energies and here we have re-inserted the factors of Hubble. The ellipsis denote terms with subleading total-energy poles and $A_n^{(p-n+3)}$ is the part of the corresponding scattering amplitude that contains the largest scaling in energy and momentum, which is of order p - n + 3. For n = 3 which is the primary focus of this work, this leading total-energy pole picks out that part of the amplitude that comes from operators with *p* derivatives.

One may go a step further and hope that with the knowledge of the scattering amplitude, as well as the form of the de Sitter mode functions, the full de Sitter wavefunction coefficient could be produced since it is the same bulk interaction vertex that gives rise to the amplitude and the wavefunction. As we have seen above, some knowledge of the de Sitter mode functions is contained in the MLT and indeed in a recent paper [266] solutions to the MLT were used to convert a contact flat space amplitude into a contact de Sitter wavefunction via a *contact reconstruction formula*:

$$\psi_n = (p-1)!(iH)^{p-n-1} \sum_{m=0}^n \sum_{\pi \in S_n} \frac{A_n^{(p-n+3)} |_{\{k_{\pi(j)}=0\}_{j=n-m+1}^n} \prod_{i=1}^{n-m} k_{\pi(i)}}{m!(n-m)!k_T^{p-m} \prod_{l=1}^m (p-l)},$$
(8.154)

where the sum $\sum_{\pi \in S_n}$ runs over the n! permutations π of $\{1, 2, \ldots, n\}$. For n = 3 this reconstruction

formula takes the following form

$$\psi_{3} = (p-1)!i^{p}H^{p-4} \left[\frac{A_{3}^{(p)}k_{1}k_{2}k_{3}}{k_{T}^{p}} + \frac{A_{3}^{(p)}\big|_{k_{1}=0}k_{2}k_{3} + A_{3}^{(p)}\big|_{k_{2}=0}k_{1}k_{3} + A_{3}^{(p)}\big|_{k_{3}=0}k_{1}k_{2}}{k_{T}^{p-1}(p-1)} + \frac{A_{3}^{(p)}\big|_{k_{2}=k_{3}=0}k_{1} + A_{3}^{(p)}\big|_{k_{1}=k_{3}=0}k_{2} + A_{3}^{(p)}\big|_{k_{1}=k_{2}=0}k_{3}}{k_{T}^{p-2}(p-1)(p-2)} + \frac{A_{3}^{(p)}\big|_{k_{1}=k_{2}=k_{3}=0}}{k_{T}^{p-3}(p-1)(p-2)(p-3)} \right].$$
(8.155)

This formula is valid for $p \ge 4$ where the time integrals in the bulk computation of these wavefunction coefficients do not produce logarithms or purely analytic terms. In this case (8.155) yields the full wavefunction. For $p \le 3$ the time integrals can yield such logarithms or analytic terms which are not captured, but in those cases the total-energy poles can still be computed using this formula; then one would need to write down an ansatz for the MLT solution and fix the additional terms that are ultimately required to satisfy the MLT. For more details we refer the reader to [266].

Instead of taking the route outlined in this chapter one could in principle use (8.155) to construct graviton bispectra. The *p*-derivative amplitude that we must input is simply given by taking one of the polarisation factors we classified in Appendix 8.7.1, multiplying this SO(3) invariant object by a polynomial in the energies of degree $(p - \alpha) > 0$, followed by summing over permutations [2]. The final sum over permutations is crucial since as can be seen from (8.155), the wavefunction coefficient will only have the correct Bose symmetry if the amplitude does. For $p \ge 4$ this procedure will generate all possible bispectra. Note that here we are advocating to use this contact reconstruction formula using polarisation tensors rather than the spinor helicity formalism since in $A_3^{(p)}$ the energy dependence needs to be from bulk time derivatives *only*. When the amplitude is written in terms of spinors, there is an energy dependence that has arisen from the polarisation factor itself rather than from bulk time derivatives, as we explained above. With the final result computed from (8.155), one can convert this expression into the spinor helicity formalism using the expressions given in Appendix 8.7.1. Above we have presented the leading order MLT solutions for each α , one can in principle use this reconstruction formula to generate *all* higher-order solutions.

8.4.5 The final form of graviton bispectra

With all of the ingredients at hand, we can now write down the final form of the wavefunction coefficients and extract the corresponding correlators. We will concentrate on the + + + and + + -

helicity configurations (since the other two are easily obtained from those by a parity transformation, with an extra – sign for odd α) and again work at each order in α treating the even and odd cases separately. Note that we classify the final form of the bispectra in terms of the leading pole of the MLT solutions presented in the previous subsection. Once we sum over permutations there can be cancellations meaning that the final form has a lower order pole. However, it is the solution to the MLT whose leading degree pole is generically equal to the number of derivatives in a corresponding bulk vertex. Each of the bispectra below can be multiplied by a real coupling which we denote as $g_{\alpha,p}$, and we absorb all O(1) factors that appear when we go from a wavefunction to correlator (c.f. (8.72)) into these couplings.

Parity-even interactions We begin with even α where both the rational parts and the logarithmic parts contribute to the correlator, as shown in Section 8.3.

 $\alpha = 0$ Since in this case we have $h_{\alpha} = 1$, both the final + + + and + + - bispectra are easily read off from the solutions to the MLT given above. We simply take the spinor helicity factors, multiply them by the MLT solutions and then divide by the power spectrum of each external field which contributes a factor of $1/e_3^3$. We have

$$p = 0: \quad e_3^3 B_3^{+++} = g_{0,0} \mathbf{SH}_{+++} [4e_3 - e_2 k_T + (k_T^3 - 3k_T e_2 + 3e_3)(\log(-k_T \eta_0/\mu)], \quad (8.156)$$

$$e_3^3 B_3^{++-} = g_{0,0} \mathbf{SH}_{++-} [4e_3 - e_2 k_T + (k_T^3 - 3k_T e_2 + 3e_3) \log(-k_T \eta_0/\mu)], \quad (8.157)$$

$$p = 2: \quad e_3^3 B_3^{+++} = g_{0,2} \mathrm{SH}_{+++} \frac{e_2 e_3 + e_2^2 k_T - 2e_3 k_T^2}{k_T^2}, \tag{8.158}$$

$$e_3^3 B_3^{++-} = g_{0,2} \mathbf{SH}_{++-} \frac{e_2 e_3 + e_2^2 k_T - 2e_3 k_T^2}{k_T^2}, \qquad (8.159)$$

$$p = 3: \quad e_3^3 B_3^{+++} = g_{0,3} \mathbf{SH}_{+++} \frac{e_3^2}{k_T^3}, \qquad (8.160)$$

$$e_3^3 B_3^{++-} = g_{0,3} \mathbf{SH}_{++-} \frac{e_3^2}{k_T^3}, \qquad (8.161)$$

This p = 0 bispectrum corresponds to a combination of a potential term in the bulk of the form γ_{ij}^3 and the contribution $k_T^3 - 3k_T e_2 + 3e_3$ which is the graviton version of the well studied local non-Gaussianity [267]. The local shape arises from taking the free theory for the massless graviton
and performing a field redefinition $\gamma_{ij} \rightarrow \gamma_{ij} + \gamma_{ik}\gamma_{kj}$. Such a redefinition does not alter the S-matrix and so its contribution to the wavefunction must be regular as $k_T \rightarrow 0$, which it is. The log piece is produced by the γ_{ij}^3 vertex which appears in Solid Inflation [130] and in the slow-roll limit it is the leading contribution from this interaction. The p = 3 bispectrum corresponds to that of a $\dot{\gamma}_{ij}^3$ vertex in the bulk which appears in the Effective Field Theory of Inflation (EFToI) [136], without corrections to the two-point function and with an independent coefficient [247]. We provide more details about these examples in Section 8.5.

 $\alpha = 2$ In this case the polarisation factor is not fully symmetric, so after we multiply it by a solution to the MLT, we need to symmetrize the result. We find

$$p = 2: \quad e_3^3 B_3^{+++} = g_{2,2} \mathsf{SH}_{+++} \frac{e_2(e_3 + e_2 k_T - k_T^3)}{k_T^2}, \tag{8.162}$$

$$e_3^3 B_3^{++-} = g_{2,2} \mathrm{SH}_{++-} \frac{(k_1 k_2 - k_2 k_3 - k_3 k_1)(e_3 + e_2 k_T - k_T^3)}{k_T^2}, \qquad (8.163)$$

$$p = 3: \quad e_3^3 B_3^{+++} = g_{2,3} \mathrm{SH}_{+++} \frac{e_3 (6e_3 + 2e_2 k_T + k_T^3)}{k_T^3}, \tag{8.164}$$

$$e_3^3 B_3^{++-} = g_{2,3} \mathbf{SH}_{++-} \frac{-e_3 (4e_3 + k_T (4e_2 + I_3^2 + 2I_3 k_T - k_T^2))}{2k_T^3}, \qquad (8.165)$$

Since GR is a two-derivative, parity-even theory, its bispectrum in de Sitter space must be contained within the solutions we have written up to this point. Indeed, if we first take $\mu = -k_T \eta_0 e^{-\tilde{g}_{0,0}/g_{0,0}}$, and then

$$g_{0,2} = 2\tilde{g}_{0,0} = -g_{2,2}, \qquad g_{0,0} = 0, \qquad (\text{GR tuning})$$
(8.166)

then both the + + + and + + - wavefunction coefficients are those of GR [171]. We remind the reader that on the total-energy poles we recover the amplitude, and in GR the + + + amplitude vanishes while the + + - amplitude does not. This tells us that in GR the + + + bispectrum should not have such a pole while the + + - one should have a degree-2 pole. If we take an arbitrary linear combination of these bispectra and demand that the + + + wavefunction does not have a total-energy pole, while the + + - has a non-zero total-energy pole, then the result is a linear combination of GR and the local non-Gaussianity. In [171] these conditions along with full de Sitter symmetry was

enough to uniquely pick out GR. Without some additional symmetry principle, we cannot set the coefficient of the local non-Gaussianity coupling to zero. Interestingly, this GR bispectrum is the leading order one in the EFToI [245, 247, 246]: the breaking of boosts is only felt at higher-order in derivatives.

 $\alpha = 4$ Again in this case the polarisation factor is not fully symmetric, so we take the solutions to the MLT and then symmetrise appropriately. We find

$$p = 4: \quad e_3^3 B_3^{+++} = g_{4,4} \mathbf{SH}_{+++} I_1 I_2 I_3 \frac{3e_3 + e_2 k_T + k_T^3}{k_T^3}, \tag{8.167}$$

$$e_3^3 B_3^{++-} = g_{4,4} \mathbf{SH}_{++-} I_1 I_2 I_3 \frac{3e_3 + e_2 k_T + k_T^3}{k_T^3}, \qquad (8.168)$$

$$p = 5: \quad e_3^3 B_3^{+++} = g_{4,5} \mathbf{SH}_{+++} I_1 I_2 I_3 \frac{24e_2 e_3 + 6e_2^2 k_T - 9e_3 k_T^2 + e_2 k_T^3}{k_T^5}, \tag{8.169}$$

$$e_{3}^{3}B_{3}^{++-} = g_{4,5}SH_{++-}I_{1}I_{2}(2k_{T})^{-4} \Big[12e_{2}^{2}k_{T} + 48e_{2}e_{3} + e_{2}k_{T} \left(3I_{3}^{2} - 2I_{3}k_{T} + k_{T}^{2} \right) \\ + 2e_{3} \left(6I_{3}^{2} - 6I_{3}k_{T} - 7k_{T}^{2} \right) + I_{3}^{2}k_{T}^{3} - k_{T}^{5} \Big],$$

$$(8.170)$$

÷

 $\alpha = 6$ Here we have $h_{\alpha} = I_1^2 I_2^2 I_3^2$, which is fully symmetric and no symmetrization is necessary when constructing the full bispectra. We have:

$$p = 6: \quad e_3^3 B_3^{+++} = g_{6,6} \mathbf{SH}_{+++} I_1^2 I_2^2 I_3^2 \frac{15e_3 + 3k_T e_2 + k_T^3}{k_T^6}, \tag{8.171}$$

$$e_3^3 B_3^{++-} = g_{6,6} \mathbf{SH}_{++-} I_1^2 I_2^2 \frac{15e_3 + 3k_T e_2 + k_T^3}{k_T^4}, \qquad (8.172)$$

$$p = 8: \quad e_3^3 B_3^{+++} = g_{6,8} \mathbf{SH}_{+++} I_1^2 I_2^2 I_3^2 \frac{7e_2 e_3 + k_T e_2^2 - 2k_T^2 e_3}{k_T^8}, \quad (8.173)$$

$$e_3^3 B_3^{++-} = g_{6,8} \mathrm{SH}_{++-} I_1^2 I_2^2 \frac{7e_2 e_3 + k_T e_2^2 - 2k_T^2 e_3}{k_T^6} , \qquad (8.174)$$

Parity-odd interactions We now turn to odd α where only the coefficient of the logarithm can contribute to the correlator. In all cases it must be multiplied by $i\pi/2$. We absorb the $\pi/2$ factor into the overall coupling $g_{\alpha,p}$ and then the additional factor of *i* combines with the factors of *i* appearing

÷

in each h_{α} , c.f. (8.115), to give real coefficients. As we showed above, to all orders in derivatives logarithms can only appear for $\alpha = 1, 3$ and give rise to a total of three solutions. Since in each case the logarithmic solutions to the MLT must always come with total-energy poles, we still classify these solutions by the corresponding p.

 $\alpha = 1$ In this case we need to multiply $h_1(k_1, k_2, k_3) = k_1$ by the appropriate solutions to the MLT and then symmetrise. We find

$$p = 1: \quad e_3^3 B_3^{+++} = g_{1,1} \mathrm{SH}_{+++} k_T \left(k_T^2 - 2e_2 \right) \,, \tag{8.175}$$

$$e_3^3 B_3^{++-} = g_{1,1} \mathbf{SH}_{++-} I_3 \left(k_T^2 - 2e_2 \right) , \qquad (8.176)$$

$$p = 2: \quad e_3^3 B_3^{+++} = g_{1,2} \mathbf{SH}_{+++} \left(-3e_3 + k_T e_2 \right) \,, \tag{8.177}$$

$$e_3^3 B_3^{++-} = g_{1,2} \mathbf{SH}_{++-} \left(k_1 (k_2^2 + k_3^2) + k_2 (k_1^2 + k_3^2) - k_3 (k_1^2 + k_2^2) \right) \,. \tag{8.178}$$

Possible operators that generate these bispectra are, respectively (up to constant factors),

$$a(\eta)^{-1}g_{1,1}\epsilon_{ijk}\gamma_{il}\gamma_{lm}\partial_j\gamma_{km},\tag{8.179}$$

$$a(\eta)^{-2}g_{1,2}\epsilon_{ijk}\gamma'_{il}\gamma_{lm}\partial_j\gamma_{km}.$$
(8.180)

 $\alpha = 3$ In this case we have $h_3(k_1, k_2, k_3) = I_1 I_2 I_3$ which is already symmetric. We then have a unique log term yielding

$$p = 3: \quad e_3^3 B_3^{+++} = g_{3,3} \mathbf{SH}_{+++} I_1 I_2 I_3 = g_{3,3} \mathbf{SH}_{+++} \left(-8e_3 + 4e_2 k_T - k_T^3 \right) \,, \tag{8.181}$$

$$e_3^3 B_3^{++-} = g_{3,3} \mathrm{SH}_{++-} I_1 I_2 k_T \,. \tag{8.182}$$

This can be generated, up to a constant factor, by the operator

$$a(\eta)^{-3}g_{3,3}\epsilon_{ijk}\partial_l\gamma_{in}\partial_m\gamma_{jl}\partial_n\gamma_{km}.$$
(8.183)

These bispectra correspond to those in computed in [247] where the three couplings were tuned as dictated by the symmetries of the EFToI. Although in that work the bispectra were presented using polarisation tensors, we have checked that they are all indeed captured by our expressions and provide details in Section 8.5.1. Such parity-odd interactions do not appear on their own in EFToI; rather,

they come with a correction to the two-point function [247]. We also discuss this further in Section 8.5.1. It is also worth pointing out that our results tell us that for bulk vertices with more than three derivatives, and therefore $p \ge 4$ degree poles in the solutions to the MLT, there are no contributions to the bispectra. Indeed, IR-divergent logarithms can only appear when

Condition for IR-divergent logs:
$$2n_{\partial_n} + n_{\partial_i} \leq 3$$
, (8.184)

where $n_{\partial_{\eta}}$ and n_{∂_i} are respectively the number of time and space derivatives in the parity-odd interaction. Note that here we assume that each field in the cubic vertex contains at most one time derivative which can always be guaranteed by using the equations of motion. This offers a complementary proof that the parity-odd Weyl³ vertex in de Sitter space leads to a vanishing bispectrum. Indeed, this is a six derivative vertex and therefore the corresponding wavefunction does not have logarithms and therefore the correlator vanishes. See [171, 251, 240] for further discussions.

Discussion So far in this chapter we have bootstrapped three-point wavefunction coefficients arising from tree-level and manifestly local bulk graviton self-interactions. We have made very minimal assumptions. We assumed the usual massless de Sitter mode functions, and assumed that the vertices are SO(3) and scale invariant. With this full catalogue at hand, one can now search for interesting subsets. Indeed, given a particular symmetry breaking pattern for inflation, as recently classified in [248], only some of these bispectra will be permitted and non-linear realisations of the broken symmetries could result in relations among the couplings $g_{\alpha,p}$, as is the case for GR. One could attempt to perform the classification at the level of the Lagrangian using an effective field theory approach, however the main message of the bootstrap approach is that the Lagrangian route might not be the most efficient. Rather, one would like to take this full catalogue of bispectra and use soft theorems to classify consistent subsets, or even better to use these objects as the building blocks of higher-point functions. We have learned from the S-matrix programme that gluing together three-point amplitudes to form consistent four-point ones can be very constraining [29, 2]. We expect this gluing procedure to also be very constraining for cosmology and plan to explore this in future work. For parity-even vertices in the EFToI, there is only a single operator at both cubic and quartic order in derivatives that does not modify the two-point function [247]. It would be interesting to rederive this result directly using bootstrap methods. In any case, in Section 8.5 we provide a discussion of how these bispectra could be classified by the EFToI [136] or as the leading contributions to the bispectra of Solid Inflation [130], following the approach of [248].

8.4.6 Parity-odd bispectra involving gravitons and scalars

Given that we have discovered very few possible parity-odd bispectra for three gravitons, let us provide a more complete analysis by also considering bispectra involving a scalar. As is well-known, the bispectrum for three scalars cannot break parity. This is easily seen given that there is no nonzero way to contract two independent momenta with an epsilon tensor, and so there are simply no parity-odd tensor structures in the absence of polarisation tensors. Let us therefore concentrate on scalar-scalar-graviton (B_3^{00+}) and scalar-graviton-graviton (B_3^{0++}) bispectra. As always, other helicity configurations can be extracted from these as explained in Section 8.4.2.

As we have done throughout this work, we concentrate on manifestly local interactions and so the solutions to the MLT that we have classified previously in this section can be used to construct trimmed wavefunctions when we also have scalars: the MLT applies to scalars and gravitons alike. Our results of Section 8.3 also apply and so for contact interactions there can be no singularities in the parity-odd bispectra. Our job to classify B_3^{00+} and B_3^{0++} is then a simple one: we first write down all possible parity-odd tensor structures, and then multiply these by a solution to the MLT. Here we will concentrate on the contributions to the bispectrum rather than the wavefunction and so the relevant part of the solution to the MLT is the coefficient of a logarithm, as we have explained in detail above. These logs can only occur for $\alpha = 1, 3$ and so we only need to consider these tensor structures.

Scalar-scalar-graviton First consider B_3^{00+} . In this case there is only a single parity-odd tensor structure, which has $\alpha = 3$:

$$\epsilon_{ijk} e^{h_3}_{im}(\mathbf{k}_3) k_1^j k_2^k k_1^m \,, \tag{8.185}$$

and permutations. The relevant solution to the MLT contains a log with a k-independent coefficient just as was the case for $\alpha = 3$ with three gravitons. If we multiply (8.185) by the appropriate MLT solution and sum over permutations, then the contribution to the bispectrum written in terms of spinor helicity variables is

$$e_3^3 B_3^{00+} = h_{3,3} \frac{[13]^2 [23]^2}{k_3^2 [12]^2} I_3^2 k_3 \,. \tag{8.186}$$

Regardless of the form of the symmetry breaking pattern, this is the *only* such parity-odd bispectrum when the bulk interactions are manifestly local, to all orders in derivatives.

Scalar-graviton-graviton Moving onto B_3^{0++} we find a single tensor structure for $\alpha = 1$ and two for $\alpha = 3$. Up to permutations we have

$$\alpha = 1 : \epsilon_{ijk} e_{im}^{h_2}(\mathbf{k}_2) e_{jm}^{h_3}(\mathbf{k}_3) k_k^2, \qquad (8.187)$$

$$\alpha = 3: \epsilon_{ijk} e_{im}^{h_2}(\mathbf{k}_2) e_{ml}^{h_3}(\mathbf{k}_3) k_j^2 k_k^3 k_l^1 \quad \text{and} \quad \epsilon_{ijk} e_{il}^{h_2}(\mathbf{k}_2) e_{jm}^{h_3}(\mathbf{k}_3) k_k^2 k_l^1 k_m^1 \,. \tag{8.188}$$

Now for $\alpha = 1$ we need to multiply this tensor structure by a degree-2 polynomial that arises from the coefficient of a log in a solution to the MLT. We find three such solutions: this tensor structure can be multiplied by k_1^2 , k_2^2 or k_3^2 with arbitrary coefficients. Each of the appropriate MLT solutions also have rational contributions with k_T poles: one solution has a simple pole while the other two solutions have k_T^{-2} poles. A complete basis is

$$e_3^3 B_3^{0++} = \frac{[23]^4}{k_2^2 k_3^2} \left[q_{1,1}(k_2+k_3)k_1^2 + q_{1,2,a}(k_2^3+k_3^3) + q_{1,2,b}(k_2k_3^2+k_3k_2^2) \right] .$$
(8.189)

Now for $\alpha = 3$ we find that when converted to spinor helicity variables, the two $\alpha = 3$ structures are equivalent and since they already scale as $\sim k^3$, the relevant part of the MLT solution is simply a constant multiplied by a log. We therefore have a single solution for the $\alpha = 3 B_3^{0++}$ bispectrum which turns out to be a linear combination of those from $\alpha = 1$ in (8.189). It follows that (8.189) is a complete list, to all orders in derivatives. From these bispectra we can also extract those for B_3^{0+-} . We have

$$e_3^3 B_3^{0+-} = \frac{I_2^4}{k_2^2 k_3^2} \frac{[12]^4}{[31]^4} \left[q_{1,1}(k_2 - k_3)k_1^2 + q_{1,2,a}(k_2^3 - k_3^3) + q_{1,2,b}(k_2 k_3^2 - k_3 k_2^2) \right], \quad (8.190)$$

where the overall factor is a necessary consequence of helicity scaling and the absence of divergences.

One of our main messages in this chapter is that parity-odd contact bispectra, arising from manifestly local cubic interactions, are small in number. In Table 8.2 we summarise the number of independent couplings associated with tree-level parity-odd bispectra of manifestly local scalars and gravitons, to all orders in derivatives, and with exact scale invariance. In inflationary models, we would expect

additional bispectra that mix scalars and gravitons and violate manifest locality. These will arise when we integrate out the non-dynamical modes. We still expect the shapes of such correlators to be heavily constrained given our discussion in Section 8.3. The primary difference is that in those cases the logs that appear in the wavefunction can in principle be multiplied by poles as one of the external energies is taken soft. In any case, in Section 8.5 we comment on when the above bispectra appear in the effective field theory of inflation and solid inflation.

| Parity-odd bispectra | SSS | SST | STT | TTT |
|----------------------|-----|-----|-----|-----|
| no. of couplings | 0 | 1 | 3 | 3 |

Table 8.2: For manifestly local and scale invariant theories, the table specifies the number of independent parity-odd tree-level bispectra for all possible combinations of scalars ("S") and gravitons ("T") to all orders in derivatives.

8.5 Graviton bispectra and symmetry breaking patterns

In this section, we want to study which of the graviton and graviton-scalar three-point functions we have discussed so far can arise during inflation depending on the particular way in which de Sitter boosts are broken. We consider the Effective Field Theory of Inflation (EFToI) [136] and the symmetry breaking pattern of a solid, i.e. Solid Inflation [130].

8.5.1 Effective field theory of inflation

Let us begin with the EFToI which is the most well-studied symmetry breaking pattern for inflation. Here the symmetry breaking is driven by a single scalar that acquires a time dependent vev. The background homogeneity and isotropy is then manifest and an approximate shift-symmetry for the resulting Goldstone mode ensures approximately scale invariant primordial correlators. In the decoupling limit and on subhorizon scales, where we can neglect gravity and the expansion of the universe, the Goldstone theory is that of a superfluid [268].

First, let us stress that at tree-level all cubic graviton interactions are manifestly local: one does not need to worry about non-manifestly local interactions coming from solving the Hamiltonian and momentum constraints in GR. The reason is that for the three-point function it is sufficient to solve the constraints at linear order⁶ [106] and at this order a two-tensor cannot mix with the scalars and transverse vector in $g^{0\mu}$. Hence, the Manifestly Local Test (MLT) we used throughout this chapter does indeed capture graviton bispectra in the EFToI. Now, what are the building blocks for the graviton operators? Initially consider operators that give rise to non-trivial cubic graviton self-interactions, but *do not* alter the graviton's quadratic action with respect to the GR contribution. This case corresponds to the setup in this chapter: standard dS mode functions for the massless graviton plus bispectra arising from manifestly-local cubic self-interactions. To find these building blocks we can stop the expansion of all geometric objects constructed from the foliation at leading order in perturbations. We can either use $\dot{\gamma}_{ij}$, which is $2 \, \delta K^i_{\ j}$ at leading order in perturbations ($\delta K_{\mu\nu}$ being the fluctuation in the extrinsic curvature of constant-time hypersurfaces), or $a^{-2}\partial_k\partial_l\gamma_{ij}$. The indices ijkl cannot be, however, chosen arbitrarily. We can either have the combination⁷

$$a^{-2} \left(\partial_k \partial_{[i} \gamma_{j]l} - \partial_l \partial_{[i} \gamma_{j]k} \right), \tag{8.191}$$

corresponding to the Riemann tensor ${}^{(3)}R^{ij}{}_{kl}$ on constant-time hypersurfaces, or

$$a^{-2}\partial^2\gamma_{ij}\,,\tag{8.192}$$

corresponding to the Ricci tensor ${}^{(3)}R^i_{\ j}$. We can then freely take further time derivatives or spatial derivatives of these building blocks, since we can project derivatives either parallel or orthogonal to n^{μ} , the normal four-vector to constant-time hypersurfaces.

Parity even Let us consider a few parity-even examples (*beyond* the bispectrum of GR) before moving to the parity-odd case. The constraints on the building blocks forbid us from having p < 3 for $\alpha = 0$ and p < 5 for $\alpha = 2$ and $\alpha = 4$. Two examples are the following: we have the dimension-6 and dimension-7 operators

$$\int d\eta d^3 x \, a(\eta) \, \gamma'_{ij} \gamma'_{jk} \gamma'_{ki} \quad \text{and} \quad \int d\eta d^3 x \, \gamma'_{ij} \gamma'_{jk} \partial^2 \gamma_{ki} \,. \tag{8.193}$$

⁶More generally, the solution of the constraints to order n is sufficient to write down the action to order (2n + 1) or less [90].

 $^{^{17}}$ The square brackets on a pair of indices denote anti-symmetrization with weight one, $A_{[ij]} \equiv (A_{ij} - A_{ji})/2$.

Since no spatial derivative is contracted with the indices of γ_{ij} , both have $\alpha = 0$ and both give the same trimmed wavefunction $\psi^{\text{trimmed}}(k_1, k_2, k_3)$ which may be assumed to be symmetric since the polarisation factors are. We find

$$\psi^{\text{trimmed}}(k_1, k_2, k_3) = \frac{e_3^2}{k_T^3},$$
(8.194)

which is our $\alpha = 0$, p = 3 solution from Section 8.4. In general we expect that the order of the total-energy pole is given by [231]

$$p = 1 + \sum_{A} (\Delta_A - 4),$$
 (8.195)

where the sum is over all vertices A with mass dimension Δ_A . For the tree-level bispectrum of gravitons and scalars this tells us that p is the total number of spatial and time derivatives. Indeed, for the first interaction in (8.193), we get a k_T^{-3} pole as expected. For the second interaction in (8.193), we naively expect a k_T^{-4} pole. However, the amplitude corresponding to this interaction vanishes, so the residue of the k_T^{-4} pole is zero⁸. A similar observation was made for the DBI limit of the EFToI in [167]. Another example is the dimension-7 operator

$$\int d\eta d^3x \,\gamma'_{jk} \gamma'_{il} \left(\partial_k \partial_{(i} \gamma_{j)l} - \partial_l \partial_{(i} \gamma_{j)k}\right), \tag{8.196}$$

which has $\alpha = 2$ because two spatial derivatives are contracted with the indices of γ_{ij} . The trimmed wavefunction coefficient is

$$\psi^{\text{trimmed}}(k_1, k_2, k_3) = \frac{k_1^2 k_2^2}{k_T^4} (k_1 + k_2 + 4k_3),$$
(8.197)

i.e. p = 4, as expected from (8.195). Note that this is the trimmed wavefunction for one of the permutations where the third leg in the diagram is not differentiated with respect to time. One would need to follow the rules outlined in Section 8.4 to find the final expressions with the correct symmetries. Despite appearances, this trimmed wavefunction does indeed satisfy the MLT for each leg.

⁸On-shell we can replace the ∂^2 with two time derivatives and then it is clear that this interaction is a total time derivative and does therefore not contribute to the energy conserving *S*-matrix.

Parity odd As we have seen already in the previous sections, the parity-odd case is much more constrained. While it is possible to write infinitely many parity-odd high-dimension operators, only one contributes to the three-point function. This is the three-dimensional Chern-Simons term, i.e.

$$\frac{M_{\rm pl}^2}{\Lambda} \int d\eta d^3 x \, a(\eta) \epsilon_{ijk} \left[\frac{{}^{(3)} \Gamma^l_{im} \partial_j {}^{(3)} \Gamma^m_{kl}}{2} + \frac{{}^{(3)} \Gamma^l_{im} {}^{(3)} \Gamma^m_{jn} {}^{(3)} \Gamma^n_{kl}}{3} \right], \tag{8.198}$$

where ${}^{(3)}\Gamma_{ij}^k$ are the Christoffel symbols for the covariant derivative on hypersurfaces of constant time and we have introduced a new scale Λ on dimensional grounds. It is possible to show (see e.g. [247], Appendix B), that this term is equal to the $W\tilde{W}$ combination (where W is shorthand for the Weyl tensor and \tilde{W} for its dual) multiplied by $f(\phi) \propto \phi$ where ϕ is the inflaton, up to a boundary term⁹ and an operator that has $\alpha = 1$ and p = 3 (which does not, then, contribute to correlators as we showed in Section 8.3). At cubic order in perturbations, this operator is equal to [247]

$$\frac{M_{\rm pl}^2}{\Lambda} \int d\eta d^3 x \, a(\eta) \left[\frac{1}{4} \epsilon_{ijk} \gamma_{kn} \partial_l \gamma_{nm} \partial_j \partial_l \gamma_{im} + \frac{1}{4} \epsilon_{ijk} \gamma_{ln} \partial_n \gamma_{im} \partial_j \partial_l \gamma_{km} - \frac{1}{4} \epsilon_{ijk} \partial_m \gamma_{nj} \partial_n \gamma_{lm} \partial_i \gamma_{lk} \right. \\
\left. - \frac{1}{12} \epsilon_{ijk} \partial_m \gamma_{lj} \partial_n \gamma_{mi} \partial_l \gamma_{nk} + \frac{1}{4} \epsilon_{ijk} \partial_n \gamma_{lm} \partial_k \gamma_{ln} \partial_j \gamma_{mi} + \frac{1}{4} \epsilon_{ijk} \partial_m \gamma_{nj} \partial_m \gamma_{lk} \partial_l \gamma_{ni} \right. \\
\left. + \frac{1}{4} \epsilon_{ijk} \partial_n \gamma_{mk} \partial_l \gamma_{mn} \partial_j \gamma_{li} \right].$$
(8.199)

We recognize the tensor structures summarized in Appendix 8.7.1. The first and sixth terms in the above equation are the first and second tensor structures in Eq. (8.289). Indeed they only have one spatial derivative contracted with the graviton's indices and have $\alpha = 1$. The other five terms are found in Eqs. (8.291)-(8.293) and have $\alpha = 3$. If we take the mode functions to be the usual massless ones in dS (see below for a discussion on this point) then the bispectrum coming from this sum of interactions is given by a linear sum of the parity-odd bispectra in Section 8.4.5. The relevant constraints are

$$g_{1,1} = -2g_{1,2} = -\frac{1}{6}g_{3,3}, \qquad (8.200)$$

and by fixing $g_{1,1}$ in terms of $M_{\rm pl}^2/(H\Lambda)$, and reinserting the factor of π , we find for the + + + configuration

⁹Given the high number of derivatives, such boundary term vanishes very fast at late times and does not contribute to the three-point function.



Figure 8.3: Three-point exchange diagram in the parity-odd sector of the EFT of Inflation

$$e_3^3 B_{\text{CS,contact}}^{+++} = \frac{\pi}{256} \frac{H}{\Lambda} \frac{H^4}{M_{\text{pl}}^4} \text{SH}_{+++} (99e_3 - 53k_T e_2 + 14k_T^3) , \qquad (8.201)$$

where the subscript CS stands for "Chern-Simons". If we rewrite this using the tunings of Eq. (8.200), we find

$$e_{3}^{3}B_{\text{CS,contact}}^{+++} = \frac{\pi}{256} \frac{H}{\Lambda} \frac{H^{4}}{M_{\text{pl}}^{4}} \text{SH}_{+++} \left(2k_{T}(k_{T}^{2} - 2e_{2}) - (-3e_{3} + k_{T}e_{2}) - 12I_{1}I_{2}I_{3} \right),$$
(8.202)

$$e_{3}^{3}B_{\text{CS,contact}}^{++-} = \frac{\pi}{256} \frac{H}{\Lambda} \frac{H^{4}}{M_{\text{pl}}^{4}} \text{SH}_{++-} \left(2I_{3}(k_{T}^{2} - 2e_{2}) - \left(k_{1}(k_{2}^{2} + k_{3}^{2}) + k_{2}(k_{1}^{2} + k_{3}^{2}) - k_{3}(k_{1}^{2} + k_{2}^{2}) \right) - 12I_{1}I_{2}k_{T} \right).$$

$$(8.203)$$

The non-linear realization of boosts not only forces the different operators to appear with tuned coefficients, it also forces a contribution to the quadratic graviton action. One can see how this is necessary from the fact that the bispectrum from Eq. (8.199), by itself, gives a contribution $\sim q^{-3}$ in the soft q limit¹⁰, spoiling the consistency relation of GR. The modification to the quadratic action is

$$-\frac{M_{\rm pl}^2}{4\Lambda}\int d\eta d^3x \,a(\eta)\,\epsilon_{ijk}\partial_i\partial_m\gamma_{jl}\partial_m\gamma_{lk}\,,\qquad(8.204)$$

and gives rise to the new three-point exchange diagram shown in Figure 8.3 where the cubic vertex is the one from GR. In this chapter we have not discussed such exchange diagrams. However, given

¹⁰This comes from the first two terms in Eq. (8.199).

that such an operator is the only source of parity violation in the EFToI, we find it interesting to consider the wavefunction coefficient and resulting bispectrum from this diagram which must appear in addition to the contact ones we just derived. Note that we are treating the correction to the two-point function perturbatively which, as shown in [245], is valid as long as the approximately constant Hubble scale during inflation is smaller than the scale Λ . Indeed, the correction to the late-time power spectrum $\langle \gamma_{\mathbf{k}}^{\pm} \gamma_{\mathbf{k}'}^{\pm} \rangle' = P^{\pm}(k)$ of the graviton is [245]

$$\delta P^{\pm}(k) = \mp \pi \, \frac{H}{2\Lambda} \frac{H^2}{M_{\rm pl}^2 k^3} \,, \tag{8.205}$$

where the factor of π is enforced by unitarity, as explained in Section 8.3, and the \pm indicates that the helicities have been split by this parity-odd correction. To compute this correction to the power spectrum the relation

$$i\epsilon_{ijk}k_j e^h_{km}(\mathbf{k}) = \lambda_h k \, e^h_{im}(\mathbf{k}) \tag{8.206}$$

proves useful, where $\lambda_{\pm} = \pm 1$. By considering this correction perturbatively, we can use the usual bulk-to-boundary and bulk-to-bulk propagators arising from the massless mode functions, as we have done to compute (8.202) and (8.203). It is worth noting that to get a parity-odd bispectrum one could use any parity-even vertex for the right-hand sub-diagram but this one is of the same order as the contact contributions arising from (8.199). The contribution to the wavefunction from the contact interactions (8.199) is $M_{\rm pl}^2/(H\Lambda)$. For the diagram in Figure 8.3, the GR vertex contributes a factor of $M_{\rm pl}^2/(H\Lambda)$, while the bulk-bulk propagator scales in the same way as the graviton power spectrum and so contributes a factor of $H^2/M_{\rm pl}^2$. Multiplying these factors together shows that the contact diagram and this exchange diagram contain the same dependence on $M_{\rm pl}$, H and Λ , as expected from the consistency relations of the EFToI [247]. If in Figure 8.3 we took the cubic vertex to be given by the sum in (8.199) then this (parity-even) contribution would scale as $M_{\rm pl}^2/(H\Lambda) \times H/\Lambda$ and for $H < \Lambda$ such a diagram is suppressed.

Now, up to cubic order the GR action is

$$S_{GR} = \frac{M_{\rm pl}^2}{8} \int d\eta d^3x \, a(\eta)^2 [\gamma_{ij}' \gamma_{ij}' - \partial_l \gamma_{ij} \partial_l \gamma_{ij} + (2\gamma_{ik}\gamma_{jl} - \gamma_{ij}\gamma_{kl})\partial_k \partial_l \gamma_{ij} + \mathcal{O}(\gamma^4)], \quad (8.207)$$

and given that both the GR vertex and the new quadratic term only have spatial derivatives, none of the propagators in the bulk time integrands are differentiated, so the relevant integrals are the same for all permutations. The only integral we need to compute for this exchange diagram is

$$-i \int d\eta \int d\eta' a(\eta) a(\eta')^2 K(k_1, \eta) G(k_1, \eta, \eta') K(k_2, \eta') K(k_3, \eta'), \qquad (8.208)$$

where we have used momentum conservation to write the internal energy as k_1 , and have included an overall -i as dictated by the Feynman rules. We will use (see e.g. [33])

$$K(k,\eta) = (1 - ik\eta)e^{ik\eta}$$
, (8.209)

$$G(k,\eta,\eta') = 2P(k)[\theta(\eta-\eta')K(k,\eta')\operatorname{Im}K(k,\eta) + (\eta\leftrightarrow\eta')], \qquad (8.210)$$

where P(k) is the GR power spectrum arising from the usual massless mode functions. We can compute this integral exactly and while the full expression is not very illuminating, we find that the result is *purely imaginary*. The result of this integral is then multiplied by a polarisation factor which is *purely real*, as we discussed in Section 8.3, so it follows that the contribution to the wavefunction from this digram is a pure phase, so it does not contribute to the bispectrum. It is interesting to note that this diagram contributes logarithms to the wavefunction which by unitarity have to come with $i\pi$ contributions too [12]. As we explained in Section 8.3, for contact diagrams the only logarithmic divergences are of the form $\log(-k_T\eta_0)$, so $i\pi$ contributions are inevitable. However, for this threepoint exchange diagram we find a number of logarithmic terms with different arguments and it turns out that the coefficients of these logs are such that all $i\pi$ contributions cancel out! This observation clearly deserves further attention and we plan to come back it in the near future.

Although the contribution of Figure 8.3 to ψ_3 is pure phase, this modification of the quadratic action still leads to a parity-odd correction to the bispectrum, as we will now show¹¹. First, we write the full ψ_2 as $\psi_2^{(0)} + \delta \psi_2$, i.e. the leading part from GR plus a small correction due to the parity-odd quadratic term. To linear order in $1/\Lambda$, the relevant contributions to the wavefunction for computing the bispectra are

¹¹We thank Aaron Hillman for discussions about this point.

$$\Psi[\gamma, \eta_0] = \exp\left\{-\frac{1}{2}\int_{\mathbf{k}}\sum_{\lambda} \left(\psi_2^{(0)}(k) + \delta\psi_2^{\lambda}(k)\right)\gamma_{\mathbf{k}}^{\lambda}\gamma_{-\mathbf{k}}^{\lambda} - \frac{1}{3!}\int_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3} (2\pi)^3 \delta^{(3)}\left(\sum \mathbf{k}_i\right)\sum_{\{\mu_i\}} \psi_3^{\mu_1\mu_2\mu_3}(\{k_i\}, \{\mathbf{k}_i\})\gamma_{\mathbf{k}_1}^{\mu_1}\gamma_{\mathbf{k}_2}^{\mu_2}\gamma_{\mathbf{k}_3}^{\mu_3} + \dots\right\},$$
(8.211)

where ψ_3 contains all contributions from GR and our parity-breaking CS term, and we have used the fact that due to SO(3) invariance helicities do not mix at quadratic order. To linear order in $1/\Lambda$ we then have

$$B_{3} = \frac{1}{\prod_{i=1}^{3} \mathcal{P}_{2}^{(0)}(k_{i})} \left(-\mathcal{P}_{3}^{\{\lambda_{i}\}}(\{\mathbf{k}_{i}\}) + \mathcal{P}_{3}^{\{\lambda_{i}\}}(\{\mathbf{k}_{i}\}) \left(\frac{\delta \mathcal{P}_{2}^{\lambda_{1}}(k_{1})}{\mathcal{P}_{2}^{(0)}(k_{1})} + 2 \text{ permutations} \right) \right), \quad (8.212)$$

where the permutations are of both momenta and helicity labels, and we have defined

$$\mathcal{P}_{n}^{\{\mu_{i}\}}(\{k_{i}\},\{\mathbf{k}_{i}\}) = \psi_{n}^{\{\mu_{i}\}}(\{k_{i}\},\{\mathbf{k}_{i}\}) + \psi_{n}^{\{\mu_{i}\}}(\{k_{i}\},\{-\mathbf{k}_{i}\})^{*},$$
(8.213)

for $n \ge 3$, while for n = 2 we use SO(3) invariance to simplify the definition of $\mathcal{P}_2^{\lambda}(k)$ as $\psi_2^{\lambda}(k) + \psi_2^{\lambda}(k)^*$. As we explained above, the contribution to ψ_3 from Figure 8.3 drops out of \mathcal{P}_3 ; thus the only parity-odd contribution to \mathcal{P}_3 is fixed by the contact interactions in (8.199) and the contributions of these interactions are given by (8.202) and (8.203). However, $\delta \mathcal{P}_2$ is non-zero. The expressions that we now need to compute the full bispectra are

$$\mathcal{P}_2^{\pm}(k) = \frac{2M_{\rm pl}^2}{H^2} k^3 \left(1 \pm \frac{\pi H}{\Lambda}\right) \tag{8.214}$$

and

$$\mathcal{P}_{3,\text{GR}}^{+++}(\{\mathbf{k}_i\}) = \frac{M_{\text{pl}}^2}{32H^2} \text{SH}_{+++}(e_3 + k_T e_2 - k_T^3), \qquad (8.215)$$

$$\mathcal{P}_{3,\text{GR}}^{++-}(\{\mathbf{k}_i\}) = \frac{M_{\text{pl}}^2}{32H^2} \mathbf{SH}_{++-} \frac{I_3^2}{k_T^2} (e_3 + k_T e_2 - k_T^3), \qquad (8.216)$$

which we have computed from the cubic Einstein-Hilbert action using the bulk formalism. It follows that the *full* parity-odd contributions to the bispecta at $O(1/\Lambda)$, by summing all terms in (8.212) with those in Eqs. (8.202), (8.203), are given by

$$e_{3}^{3}B_{\text{CS,total}}^{+++} = \frac{\pi}{256} \frac{H}{\Lambda} \frac{H^{4}}{M_{\text{pl}}^{4}} \text{SH}_{+++} \left(2k_{T}(k_{T}^{2} - 2e_{2}) - (-3e_{3} + k_{T}e_{2}) - 12I_{1}I_{2}I_{3} - 3(k_{T}^{3} - e_{2}k_{T} - e_{3}) \right)$$

$$(8.217)$$

$$e_{3}^{3}B_{\text{CS,total}}^{++-} = \frac{\pi}{256} \frac{H}{\Lambda} \frac{H^{4}}{M_{\text{pl}}^{4}} \text{SH}_{++-} \left(2I_{3}(k_{T}^{2} - 2e_{2}) - \left(k_{1}(k_{2}^{2} + k_{3}^{2}) + k_{2}(k_{1}^{2} + k_{3}^{2}) - k_{3}(k_{1}^{2} + k_{2}^{2})\right) - 12I_{1}I_{2}k_{T} + \frac{I_{3}^{2}}{k_{T}^{2}}(e_{3} + k_{T}e_{2} - k_{T}^{3}) \right).$$

$$(8.218)$$

Again, we see that these parity-odd corrections are suppressed by H/Λ compared to the GR contributions. Using the relation $P^{\pm}(k) = 1/\mathcal{P}^{\pm}(k)$, it is straightforward to check that these bispectra satisfy the consistency condition for large wavelength gravitons, i.e. $\langle \gamma_{\mathbf{k}-\mathbf{q}/2}^{h_S} \gamma_{-\mathbf{k}-\mathbf{q}/2}^{h_S} \gamma_{\mathbf{q}}^{h_L} \rangle' \sim \frac{3}{2} P^{h_L}(q) P^{h_S}(k) e_{ij}^{h_L}(\mathbf{q}) \hat{k}^i \hat{k}^j$ for $q/k \to 0$.

To conclude, many of the bispectra we have computed in Section 8.4 do indeed arise in the EFToI, without corrections to the quadratic theory. The parity-odd contact bispectra, however, necessarily come with a correction to the two-point function that can be treated perturbatively and results in a *total* parity-odd contribution to the bispectrum in the EFToI given in (8.217) and (8.218). Although it is very interesting that all parity-odd corrections can be computed, they are suppressed relative to the GR contribution and will therefore be very difficult to detect observationally. This suppression was also noted in [250]. In this chapter we have restricted ourselves to exact scale invariance, and away from this limit other shapes are possible [249]. As we mentioned above, it would be very interesting to pick out this EFToI subset of our full catalogue directly at the level of the correlator rather than going back to the Lagrangian. We hope to return to this in the future.

In Section 8.4.6 we showed that (8.186) is the unique scale invariant, manifestly-local and parity-odd bispectrum of two scalars and a graviton. Here we will argue that this bispectrum *does not* appear in the EFToI. For the reader's convenience the structure of this bispectrum is

$$e_3^3 B_3^{00+} = h_{3,3} \frac{[13]^2 [23]^2}{k_3^2 [12]^2} I_3^2 k_3 , \qquad (8.219)$$

and the relevant polarisation factor is

$$\epsilon_{ijk} e_{im}^{h_3}(\mathbf{k}_3) k_1^j k_2^k k_1^m \,, \tag{8.220}$$

up to permutations. Let's first ask if such a tensor structure can appear in the EFToI without a correction to the graviton's two-point function. In this case the corresponding operator must be built out of the building blocks we listed above. Since such an interaction must have three spatial derivatives (simply from the form of the polarisation factor) and no time derivatives (to ensure $2n_{\partial_{\eta}} + n_{\partial_i} \leq 3$ and therefore a non-vanishing bispectrum) the only possible building blocks contain two derivatives acting on the graviton. However, it is easy to see from the structure of (8.220) that there is no way for two of the k's to correspond to the graviton. This implies that if such a bispectrum is to appear in the EFToI, it should come with a correction to the graviton's two-point function.

The leading correction to the graviton's two-point function is the one we discussed above, namely (8.204) which, as dictated by symmetries, appears in the EFToI in the form of the Chern-Simons term (8.198), see also [245]. We have checked that to leading order in slow-roll this Chern-Simons term does not contain a three-derivative scalar-scalar-graviton interaction and therefore cannot produce this bispectrum. We therefore conclude that (8.186) does not appear in the EFToI. We note that in [250] a scalar-scalar-graviton interaction with four-derivatives was derived from this Chern-Simons term. This interaction is slow-roll suppressed, but given our discussion in Section 8.3 it also has too many derivatives to produce a log in the corresponding wavefunction and so the corresponding bispectrum is zero (rather than simply small).

8.5.2 Solid inflation

Let us now switch to the symmetry breaking pattern of solid inflation [130]. Here the symmetry breaking is driven by a multiplet of scalar fields that pick up *spatial* vevs. Internal symmetries of the scalars then ensure that the background geometry is homogeneous and isotropic. In stark contrast to the EFToI, the fluctuations in solid inflation break spatial diffeomorphisms, and yield the following effective field theory description [248].

The building blocks in unitary gauge are constructed from the one-forms $\partial_{\mu}x^{i}$, with the requirement that latin indices are contracted in an SO(3)-invariant way. An important role is played by the trace

$$X = g^{ii}, (8.221)$$

which is a proxy for time, and by the four-vector

$$O^{\mu} = \frac{e^{\mu\nu\rho\sigma}\epsilon_{ijk}\partial_{\nu}x^{i}\partial_{\rho}x^{j}\partial_{\sigma}x^{k}}{6\sqrt{\det(g^{mn})}}, \qquad (8.222)$$

which we can use to take time derivatives of diffeomorphism scalars (via $O^{\mu}\nabla_{\mu}$). Then, we can take spatial derivatives of diffeomorphism scalars via

$$D_i^{\perp} = \frac{g^{i\mu} \nabla_{\mu}}{\sqrt{X/3}} \,. \tag{8.223}$$

This reduces to ∂_i/a at zeroth order in perturbations. The last ingredient is the SO(3) tensor

$$\Gamma^{ij} = \delta_{ij} - \frac{3g^{ij}}{X}, \qquad (8.224)$$

which is equal to γ_{ij} at leading order in perturbations.

With these building blocks it is then possible to write all possible manifestly-local cubic operators involving three gravitons. These are always built from an object of the form

$$(D_{i_1}^{\perp}\cdots D_{i_{\alpha_1}}^{\perp}\Gamma^{il})(D_{j_1}^{\perp}\cdots D_{j_{\alpha_2}}^{\perp}\Gamma^{jm})(D_{k_1}^{\perp}\cdots D_{k_{\alpha_3}}^{\perp}\Gamma^{kn}), \qquad (8.225)$$

where indices are contracted with δ_{ij} or ϵ_{ijk} . Contracting the indices $i_1, \ldots i_{\alpha_1}, j_1, \ldots j_{\alpha_1}, k_1, \ldots k_{\alpha_1}$ allows us to obtain all the tensor structures discussed in Section 8.4.1 and Appendix 8.7.1. Adding time derivatives to Γ^{il} , Γ^{jm} , Γ^{kn} only changes $\psi_3^{\text{trimmed}}(k_1, k_2, k_3)$ and is always allowed in this solid inflation EFT. So *all of the interactions we have considered in Section 8.4 can arise in solid inflation*. As it is clear from our bootstrap approach, there are a number of degeneracies at the level of the action that do not appear when working directly with observables. However, as an example, one possible set of interactions in solid inflation that can give rise to the parity-odd bispectra that we derived in Section 8.4.5 are

$$g_{1,1}: \quad \int d^4x \sqrt{-g} \,\epsilon_{ijk} \Gamma^{kn} \Gamma^{nm} D_j^{\perp} \Gamma^{im} \,, \tag{8.226}$$

$$g_{1,2}: \quad \int d^4x \sqrt{-g} \,\epsilon_{ijk} \Gamma^{kn} O^{\mu} \nabla_{\mu} \Gamma^{nm} D_j^{\perp} \Gamma^{im} \,, \tag{8.227}$$

$$g_{3,3}: \int d^4x \sqrt{-g} \,\epsilon_{ijk} D_m^{\perp} \Gamma^{lj} D_n^{\perp} \Gamma^{mi} D_l^{\perp} \Gamma^{nk} \,, \qquad (8.228)$$

where O^{μ} is defined in Eq. (8.222). In addition to parity-odd graviton bispectra, each of these operators

also generates mixed bispectra containing gravitons and scalars, such as for example scalar-scalargraviton and scalar-graviton-graviton bispectra. No purely scalar bispectra are generated because scalar bispectra cannot be parity-odd. To see why the scalars must enter the game, notice that to leading order in scalar and tensor perturbations we have

$$\Gamma^{ij} = \gamma_{ij} + 2\zeta \delta_{ij} - 6 \frac{\partial_i \partial_j}{\partial^2} \zeta , \qquad (8.229)$$

where ζ are curvature perturbations on constant-energy time slices. Furthermore, this expression makes it clear that the interactions involving the scalar ζ are *not* manifestly local due to the appearance of the inverse Laplacian in the last term in (8.229), a fact that we have verified with an explicit computation. Hence, the mixed scalar-scalar-graviton and scalar-graviton-graviton bispectra generated by these operators are not the ones we derived in Section 8.4.6, where we used the Manifestly Local Test (MLT) and therefore described only manifestly local interactions. In Section 8.5.4 we will see that, in solid inflation, the signal-to-noise ratio for the bispectra involving one or more scalar is always larger than that for the purely graviton bispectrum. Therefore, if the parity-odd graviton bispectra derived here were to be detected, then either one should also see the corresponding parity-odd mixed bispectra or one would conclude that the symmetry breaking pattern during inflation is different from that assumed in solid inflation (and the EFT of inflation).

Since in Section 8.4.6 we have bootstrapped *only* the manifestly local parity-odd scalar-scalar-graviton bispectrum, it is natural to ask whether that mixed bispectrum can be generated in solid inflation. It is straightforward to see that the answer is yes. Let us consider the operator

$$\int d^4x \sqrt{-g} X^{-2} \epsilon_{ijk} D_l^{\perp} X D_k^{\perp} X D_j^{\perp} \Gamma^{li} , \qquad (8.230)$$

which starts at cubic order in perturbations. Using $X = 3a^{-2}(1+2\zeta)$, together with Eq. (8.229), we see that to lowest order it is equal to

$$4\int d\eta d^3x \,a(\eta)\,\epsilon_{ijk}\partial_l\zeta\partial_k\zeta\partial_j\gamma_{li}\,.$$
(8.231)

Furthermore, the operator in (8.230) does not introduce any other mixed interaction. In particular, it does not generate terms with two gravitons and one ζ , with three gravitons (given that X does

not contain γ_{ij}), or with three scalars (they all vanish by integration by parts). The interaction in Eq. (8.231) gives the bispectrum of Eq. (8.186). Since the coupling constant of this operator can be large, we conclude that the parity-odd scalar-scalar-graviton bispectrum that we have bootstrapped can indeed arise and be large in solid inflation. We plot its shape in Figure 8.4.

Mode functions Let us now discuss our assumption that the mode functions for the graviton and scalar are the usual massless de Sitter ones. Let us start with the graviton. As one may have anticipated from the breaking of spatial diffeomorphisms, in solid inflation the graviton acquires a mass. Indeed, in unitary gauge solid inflation can be thought of as a theory of Lorentz-violating massive gravity [269]. On the surface this seems problematic for our assumptions, but it turns out that this mass is slow-roll suppressed. Let us quickly review how this happens (we refer the reader to [130] for more details.) We write the non-Einstein-Hilbert part of the action S as

$$S = \int d^4x \sqrt{-g} \left\{ \mathcal{L}_0(X) + M^4(X, \delta Y, \delta Z) \right\}, \qquad (8.232)$$

where we defined

$$\delta Y = Y - \frac{1}{3} = \frac{g^{ij}g^{ji}}{X^2} - \frac{1}{3}, \qquad (8.233)$$

$$\delta Z = Z - \frac{1}{9} = \frac{g^{ij}g^{jk}g^{ki}}{X^3} - \frac{1}{9}, \qquad (8.234)$$

and $X = g^{ii}$ was defined in Eq. (8.221). Without loss of generality, M^4 can be written as an expansion in powers of δY and δZ , each multiplied by a function of X. This mimics the expansion in powers of $g^{00} + 1$ (with time-dependent coefficients) of the EFToI action at zeroth order in derivatives, thanks to the fact that δY and δZ start at second order in perturbations around a FLRW background. Consequently, it is only \mathcal{L}_0 whose dependence is fixed by the background evolution: we have

$$3M_{\rm pl}^2 H^2 = -\mathcal{L}_0 \,, \tag{8.235}$$

$$\varepsilon \equiv -\frac{\dot{H}}{H^2} = \frac{d\log \mathcal{L}_0}{d\log X}, \qquad (8.236)$$

where we recognize the slow-roll parameter ε on the left-hand side of (8.236). What are the contributions of Eq. (8.232) to the graviton action (which are added to those from Einstein gravity)? It is straightforward to see that

$$S_{\gamma\gamma} = \int d^4x \, a^3 \left(\frac{1}{6} \frac{d\mathcal{L}_0}{d\log X} + \frac{1}{9} \frac{\partial M^4}{\partial Y} + \frac{1}{9} \frac{\partial M^4}{\partial Z} \right) \gamma_{ij} \gamma_{ji} \,. \tag{8.237}$$

Using Eq. (8.236), together with the fact that the propagation speed c_T^2 of the transverse part of the Goldstone modes π^i is [130]

$$c_{\rm T}^2 = 1 + \frac{2}{3} \frac{M_{,Y}^4 + M_{,Z}^4}{X \mathcal{L}_{0,X}}, \qquad (8.238)$$

we see that the graviton has a small mass given by

$$m_{\gamma}^2 = -2\dot{H}c_{\rm T}^2 = 2H^2\varepsilon c_{\rm T}^2$$
. (8.239)

One might ask what happens if other (higher-derivative) operators are turned on. No other operator can contribute to the mass term aside from Eq. (8.237). They could, however, modify Eq. (8.236) and therefore modify how one converts from Eq. (8.237) to Eq. (8.239) via the definition of the speed of sound c_T^2 . As long as these operators are only small perturbative corrections to the solid inflation action of Eq. (8.232), the graviton mass will still be given by Eq. (8.239) at leading order, and thus it will be slow-roll-suppressed. Therefore, to leading order in slow-roll one can use the massless dS mode functions to compute the bispectra, as we have been doing and as was done in [248, 270]. In the scale invariant limit, it then follows that *all* of the parity-even and parity-odd bispectra we have constructed in this chapter using the MLT are the leading contributions to the graviton bispectra in solid inflation.

Let us now turn to the scalar's mode functions. We have shown that the unique parity-odd B_3^{00+} can indeed arise from an interaction in solid inflation under the assumptions we have made in this chapter which translate into constraints from the COT, MLT and boostless bootstrap rules that enable us to write down an ansatz for the wavefunction. It turns out that in solid inflation the mode functions for ζ are not the usual ones for a massless scalar in dS [130], yet the COT, MLT and our ansatz still apply and so this unique bispectrum can indeed be generated when using the corrected mode functions for ζ .

In more detail, it was shown in [130] that in the slow-roll limit the ζ mode functions contain an extra term relative to the usual ones for a massless scalar in dS which corrects the ζ bulk-boundary propagator to

$$K_{\zeta}(k,\eta) \sim (1 - ic_L k\eta + c_L^2 k^2 \eta^2 / 3) e^{ic_L k\eta},$$
 (8.240)

where we have omitted overall constant factors¹². Now, one can easily verify that this new bulkboundary propagator satisfies (8.54) and any contact diagrams that we derive using this propagator satisfy the contact COT (8.55). One can also see that this propagator will lead to wavefunction coefficients that satisfy the MLT. Indeed, the first derivative of (8.240) vanishes at k = 0. Finally, wavefunction coefficients due to manifestly-local interactions that are derived in the bulk formalism using this bulk-boundary propagator still take the form we assume in this chapter the energy dependence corresponds to rational functions, with only total-energy poles, with the additional possibility of $log(-k_T\eta_0)$ terms multiplied by polynomials. This can be easily seen from the fact that any integrand in the bulk formalism that is a function of (8.240) and its derivatives, can also be written in terms of the usual expression for $K(k, \eta)$ and its derivatives. To capture the effects of the $k^2\eta^2$ correction, we would need to include terms with extra time derivatives which will in turn result in wavefunction coefficients where the degree of the leading total-energy pole will not equal the number of derivatives in the corresponding bulk interaction. Indeed, we have

$$K_{\zeta}(k,\eta) = K(k,\eta) - \frac{\eta}{3} \frac{\partial K(k,\eta)}{\partial \eta} \,. \tag{8.241}$$

For our interests in this chapter, the important point is that the interaction in (8.231) still generates a logarithm in the wavefunction. Indeed, given that the relevant interaction vertex has $\alpha = 3$, terms coming from the second term on the RHS of (8.241) will violate $2n_{\partial_{\eta}} + n_{\partial_i} \leq 3$ and so will not alter the coefficient of the log divergence coming from three copies of $K(k, \eta)$. It follows that (8.186) does indeed arise as a ζ -correlator in solid inflation.

8.5.3 Phenomenology of parity-odd interactions in solid inflation

In the EFToI, we showed above that the parity-odd correction to the *graviton* bispectrum is small relative to the GR contribution since in our analysis we took the correction to the quadratic action to be perturbative. Furthermore, the mixed scalar-scalar-graviton parity-odd bispectrum that we found in Section 8.4.6 does not arise in the EFToI. In solid inflation, however, we can choose operators that do not affect the quadratic action and can in principle give rise to large parity-odd bispectra: see (8.226) - (8.228) for gravitons (which do also introduce three-point functions involving the curvature pertur-

¹²We note that ζ and γ are not conserved in solid inflation which induces time dependence in both the power spectra and bispectra of these modes. However, in the slow-roll limit this time dependence is small and to capture it one needs to keep slow-roll corrections in the mode functions. Importantly for us, the shapes of solid inflation bispectra are unaffected by this mild time dependence. We refer the reader to [130] for details.

bation ζ) and (8.230) for scalar-scalar-graviton (which is the only three-point function arising from that operator).

The reason we focus on these three-point functions is two-fold. On the one hand, if parity-odd non-Gaussianities involving scalars are suppressed in some model of inflation beyond the ones we discuss here, then graviton-graviton-graviton bispectra will be the leading signal. On the other hand, in the generic case where non-Gaussianities involving scalars are not suppressed (as is the case in EFToI and solid inflation), then the scalar-scalar-graviton signal will be the leading one, as we show in Sec-



Figure 8.4: The shape of the unique parity odd, manifestly local B_3^{00+} .

tion 8.5.4. In this short section, we therefore study the phenomenology of the parity-odd, *manifestly local* graviton-graviton-graviton and scalar-scalar-graviton bispectra in more detail by plotting and commenting on the shapes of each possibility that we presented in Sections 8.4.5 and 8.4.6.

Let us start with the unique scalar-scalar-graviton bispectrum, which is

$$(\alpha = 3, p = 3): \quad B_3^{00+} = h_{3,3} \frac{1}{e_3^3} \frac{[23]^2 [31]^2}{[12]^2 k_3} I_3^2.$$
 (8.242)

Taking $\mathbf{k}_3 = k_3 \hat{\mathbf{z}}$ without loss of generality, the bispectrum can be rewritten in terms of the graviton polarisation tensor

$$e^{\pm}(\mathbf{k}_3) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & \pm i \\ 0 & \pm i & -1 \end{pmatrix}, \qquad (8.243)$$

as

$$B_3^{00h} = -4\lambda_h h_{3,3} \frac{1}{e_3^3} e_{ij}^h(\mathbf{k}_3) k_i^1 k_j^2 k_3 = \lambda_h \frac{h_{3,3}}{e_3^3 k_3} k_T I_1 I_2 I_3 , \qquad (8.244)$$

where $\lambda_{\pm} = \pm 1$. To see the shape of this bispectrum, in Figure 8.4 we plot $B_3^{00+}e_3^2$, which has a

vanishing scaling dimension. The correlator vanishes in the folded limit and does not peak in the squeezed limit.

Let us now move to the three parity-odd graviton bispectra. For the convenience of the reader we recall that these are

$$\alpha = 1, \ p = 1: \quad B_3^{+++} = g_{1,1} \mathrm{SH}_{+++} \frac{k_T \left(k_T^2 - 2e_2\right)}{e_3^3},$$
$$B_3^{++-} = g_{1,1} \mathrm{SH}_{++-} \frac{I_3 \left(k_T^2 - 2e_2\right)}{e_3^3},$$
(8.245)

$$\alpha = 1, \ p = 2: \quad B_3^{+++} = g_{1,2} \mathbf{SH}_{+++} \frac{-3e_3 + k_T e_2}{e_3^3},$$
$$B_3^{++-} = g_{1,2} \mathbf{SH}_{++-} \frac{k_1 (k_2^2 + k_3^2) + k_2 (k_1^2 + k_3^2) - k_3 (k_1^2 + k_2^2)}{e_3^3}, \qquad (8.246)$$

$$\alpha = 3, \ p = 3: \quad B_3^{+++} = g_{3,3} \mathrm{SH}_{+++} \frac{I_1 I_2 I_3}{e_3^3},$$
$$B_3^{++-} = g_{3,3} \mathrm{SH}_{++-} \frac{I_1 I_2 k_T}{e_3^3}.$$
(8.247)

Now, for each of these bispectra the polarisation factor is unique and is fixed by the helicity transformations of the external spinors. In terms of polarisation tensors we have

$$SH_{++\pm} = -e_{ij}^{+}(\mathbf{k}_{1})e_{jk}^{+}(\mathbf{k}_{2})e_{ik}^{\pm}(\mathbf{k}_{3}), \qquad (8.248)$$

which we can express solely in terms of the energies k_1, k_2, k_3 . Using momentum conservation and SO(3) invariance, we can make each of the three external vectors lie in the (x, y) plane with

$$\mathbf{k}_1 = k_1(1,0,0), \qquad \mathbf{k}_2 = k_2(\cos\theta,\sin\theta,0), \qquad \mathbf{k}_3 = k_3(\cos\varphi,\sin\varphi,0), \qquad (8.249)$$

where

$$\cos\theta = \frac{k_3^2 - k_1^2 - k_2^2}{2k_1k_2}, \qquad \cos\varphi = \frac{k_2^2 - k_3^2 - k_1^2}{2k_1k_3}.$$
(8.250)

The angles θ and φ are simply those formed by \mathbf{k}_1 with \mathbf{k}_2 and \mathbf{k}_3 respectively. With this representation for \mathbf{k}_a we can write the polarisation tensors as

$$e^{\pm}(\mathbf{k}_{1}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & \pm i \\ 0 & \pm i & -1 \end{pmatrix}, \qquad (8.251)$$

$$e^{\pm}(\mathbf{k}_{2}) = \begin{pmatrix} \sin^{2}\theta & -\sin\theta\cos\theta & \mp i\sin\theta \\ -\sin\theta\cos\theta & \cos^{2}\theta & \pm i\cos\theta \\ \mp i\sin\theta & \pm i\cos\theta & -1 \end{pmatrix}, \quad (8.252)$$
$$\begin{pmatrix} \sin^{2}\varphi & -\sin\varphi\cos\varphi & \mp i\sin\varphi \end{pmatrix}$$

$$e^{\pm}(\mathbf{k}_3) = \begin{pmatrix} -\sin\varphi\cos\varphi & \cos^2\varphi & \pm i\cos\varphi \\ \mp i\sin\varphi & \pm i\cos\varphi & -1 \end{pmatrix}.$$
 (8.253)

It is then straightforward to see that

$$SH_{+++} = -\frac{k_T^3 \left(8e_3 - 4k_T e_2 + k_T^3\right)}{e_3^2}, \qquad (8.254)$$

$$SH_{++-} = -\frac{I_3^3 \left(-8e_3 - 4I_3e_2' + I_3^3\right)}{e_3^2}.$$
(8.255)

Note that, perhaps surprisingly, these expressions are purely rational. Here we have defined e'_2 which is simply e_2 with the sign of k_3 flipped i.e. $e'_2 = k_1k_2 - (k_1 + k_2)k_3$.

To see the behaviour of these different shapes we plot $B_3^{++\pm}e_3^2$ for each of the three couplings. These combinations have vanishing scaling weight and can be written as functions of the dimensionless parameters

$$x_2 \equiv \frac{k_2}{k_1}, \qquad x_3 \equiv \frac{k_3}{k_1}.$$
 (8.256)

The shapes can be found in Figure 8.5. We see that both $\alpha = 1$ parity-odd bispectra peak in the squeezed limit for all helicities, but have an angular dependence which causes them to vanish when all spatial momenta are parallel. More specifically, in the squeezed limit $k_3 \ll k_1, k_2$, all $\alpha = 1$ bispectra are proportional to $\sin^2(\measuredangle(\mathbf{k}_1, \mathbf{k}_3))$. By contrast, the $\alpha = 3$ parity-odd bispectrum vanishes in the squeezed limit for all helicities, and is large in the equilateral configuration.



Figure 8.5: Shapes of each of the three tree-level, contact parity-odd bispectra consistent with the MLT.

8.5.4 On the detectability of graviton and scalar bispectra

In this subsection we discuss the signal-to-noise ratio S/N for general bispectra. We use this analysis to argue that, since $(S/N)^2$ scales with the power spectrum of the fields involved, it is larger for bispectra that contain more scalar fields, all other things being equal.

So far we have seen that the three parity-odd graviton bispectra that we have bootstrapped to all orders in derivatives can indeed arise in solid inflation. Since there cannot be any parity-odd scalar bispectra, the graviton bispectra do not have any counterpart in the purely scalar sector and are therefore unconstrained by current data. In solid inflation they can appear with a large coefficient and should therefore be considered an important observational target for observations of the polarization of the Cosmic Microwave Background (CMB). It would be interesting search for these parity-odd graviton bispectra also with gravitational wave interferometers. Since both ground and space based interferometers probe scales that are very much shorter than cosmological scales, the possibility to detect a primordial stochastic background of gravitational waves in the conceivable future hinges on having a blue tilt in the tensor power spectrum. It is worth keeping in mind that such a blue tilt is at odds with the assumption of scale invariance that we have used extensively in this work.

The operators in (8.226)-(8.228) generate a parity-odd graviton bispectrum, but also scalar-scalargraviton and scalar-graviton-graviton bispectra. It is therefore interesting to ask which of these signals can be seen first. To assess the theoretical detectability of a bispectrum we look at the signal-to-noise ratio. For concreteness and convenience, we assume that we can access the full three-dimensional distribution of the fields within a volume V and up to a resolution of order k_{max}^{-1} . Let us consider the following action for three massless fields φ_a with a = 1, 2, 3, which can be scalars or gravitons,

$$S = \int d^3x d\eta \, a^4 \left[\sum_{a=1}^3 \frac{\Delta_a^2}{2} \left(\partial_\mu \phi_a \right)^2 + g a^{-p} \partial^p \phi_1 \phi_2 \phi_3 \right] \,, \tag{8.257}$$

where g is a coupling constant, Δ_a is an arbitrary normalization, and we have schematically indicated that the interaction has p derivatives and therefore comes with the appropriate power of the scale factor required by scale invariance. The indices of the spatial derivatives can be contracted with the indices of the gravitons, with δ_{ij} or with the anti-symmetric Levi-Civita symbol ϵ_{ijk} , so that this discussion captures parity-odd interactions as well. For example, for the graviton γ_{ij} we would have $\Delta_{\gamma} = M_{\rm P}/2$ and for curvature perturbations $\Delta_{\zeta} = M_{\rm P}\sqrt{2\epsilon} \ll \Delta_{\gamma}$. The power spectra are found to be

$$\langle \phi_a \phi_a \rangle' = \frac{H^2}{2\Delta_a^2} \frac{1}{k^3} \equiv \frac{A_a}{k^3} \,, \tag{8.258}$$

where the prime indicates that we are dropping the factor $(2\pi)^3 \delta^3(\mathbf{k})$. The bispectrum induced by the interactions in (8.257) in the in-in formalism takes the following schematic form

$$\langle \phi_1 \phi_2 \phi_3 \rangle' = B_{123} = \int d\eta \langle [H_{\text{int}}, \phi_1 \phi_2 \phi_3] \rangle$$
(8.259)

$$\sim g H^{p-4} \left(\prod_{a=1}^{3} \frac{A_a}{k_a^3}\right) R_3(k_1, k_2, k_3),$$
 (8.260)

where R_3 is a rational function of the momenta that scales as k^3 , up to possible logarithmic terms. For parity-even interactions we expect $R_3 \sim \text{Poly}_{p+3}/k_T^p$, while for parity-odd interactions we have proven that no k_T pole can arise and so $R_3 \sim \text{Poly}_3$. Notice that the bispectrum therefore scales as the power spectrum of each of the fields. Then, we define the dimensionless signal-to-noise ratio S/Nas (see e.g. [271])

$$\left(\frac{S}{N}\right)^2 = V^3 \int_{\mathbf{k}_1 \mathbf{k}_2 \mathbf{k}_3} \frac{\langle \phi_1(\mathbf{k}_1)\phi_2(\mathbf{k}_2)\phi_3(\mathbf{k}_3) \rangle \langle \phi_1(\mathbf{k}_1)\phi_2(\mathbf{k}_2)\phi_3(\mathbf{k}_3) \rangle}{\langle \phi_1(\mathbf{k}_1)\phi_2(\mathbf{k}_2)\phi_3(\mathbf{k}_3)\phi_1(\mathbf{k}_1)\phi_2(\mathbf{k}_2)\phi_3(\mathbf{k}_3) \rangle}$$
(8.261)

$$= V^3 \int_{\mathbf{k}_1 \mathbf{k}_2 \mathbf{k}_3} (2\pi)^3 \delta\left(\sum_{a=1}^3 \mathbf{k}_a\right) \frac{B_{123}(k_1, k_2, k_3)^2 \times (2\pi)^3 \delta(\mathbf{0})}{\prod_{a=1}^3 (2\pi)^3 \delta(\mathbf{0}) P_a},$$
(8.262)

where we estimated the denominator, i.e. the noise, in the Gaussian theory since we have in mind non-Gaussianities that are perturbatively close to the Gaussian theory. For a finite-volume survey we substitute $(2\pi)^3 \delta(\mathbf{0}) = V$ and use (8.259) to find

$$\left(\frac{S}{N}\right)^2 = V \int_{\mathbf{k}_1 \mathbf{k}_2} \frac{\left(g H^{p-4} R_3 \prod_{a=1}^3 A_a\right)^2}{e_3^3 \prod_{a=1}^3 A_a}$$
(8.263)

$$= g^2 H^{2p-8} \left(\prod_{a=1}^3 A_a\right) \left(Vk_{\max}^3\right) \,, \tag{8.264}$$

where we estimated the momentum integrals with dimensional analysis¹³. Since we can write $V \sim k_{\min}^{-3}$ and the number of independent data points is $N_{\text{data}} \sim (k_{\max}/k_{\min})^3$, the last factor confirms

¹³Here we focus our attention on the parametric scaling of S/N. The reader should be mindful that this discussion neglects the fact that different bispectra might have very different shapes and so momentum integrals might give rise to large numerical factors that are not captured by dimensional analysis. This is not the case for the parity-odd bispectra we are considering in this work.

the intuition that our ability to detect a signal scales as $S/N \sim N_{data}^{1/2}$. From the above expression we deduce that if two interactions have the same coupling constant g, then the interaction involving fields with the largest power spectrum has the most signal-to-noise ratio and therefore should be the main observational target.

If we apply this result to the parity-odd bispectra generated in solid inflation by the operators (8.226)-(8.228) we conclude that *the scalar-scalar-graviton bispectrum is expected to have an S/N larger than the graviton bispectrum by a factor of* ϵ^{-1} , which is the inverse of the small tensor-to-scalar ratio. (See [272] for a detailed analysis of efficient CMB estimators of this signal.) To summarize, we briefly discuss some possible scenarios in which the manifestly local parity-odd bispectra that we computed in this work can be the leading observational signals:

- The manifestly local, parity-odd scalar-scalar-graviton bispectrum that we computed in (8.185) and which is generated in solid inflation by the interaction in (8.230) does not have a purely scalar counterpart because of symmetry, and therefore can be the leading observational signal in solid inflation or in other non-minimal symmetry breaking patterns.
- If one has access only to the gravitational sector, as it is the case for example if one considers only interferometric and pulsar observations of gravitational waves, then the parity-odd graviton bispectra in (8.175)-(8.182) can be the leading observational signals in solid inflation. A detection of these signals would rule out the effective field theory of inflation.
- A detection of the parity odd graviton bispectra in (8.175)-(8.182) that is not accompanied by correlated parity-odd scalar-scalar-graviton and scalar-graviton-graviton bispectra with a much higher signal-to-noise ratio would rule out both the effective field theory of inflation and solid inflation. It would be interesting to investigate what symmetry breaking pattern could be consistent with this possibility.

8.5.5 Perturbativity, naturalness and strong coupling

Since we have claimed at the beginning of this chapter that the bispectra we study, in particular the parity-odd bispectra of Section 8.5.3 can be *large*, we need to verify how large they can be without compromising the validity of our analysis. One might worry that loop corrections could spoil our conclusions. Such corrections can come in a number of forms. Loops could introduce brand

new shapes coming from performing new bulk time integrals coming from loop diagrams. These will introduce more complicated shapes that we have not considered here, but these will always be suppressed relative to the ones we have computed as long as we work below the strong coupling scale which we estimate below. Loops could also alter the quadratic action which we have assumed takes on the GR form. Such corrections could be in the form of operators with three or more derivatives that introduce new diagrams that contribute to the bispectrum. We will show below that these corrections are always small if we work below the strong coupling scale. Corrections to the quadratic action could also arise in the form of a large mass correction to the graviton. In Solid Inflation, where our large parity-odd bispectra can arise, the graviton is massive but the mass is very small and in this section we pay special attention to the issue of large mass corrections within the context of naturalness. A reader not interested in the details of the calculation may skip to the end of this section, where we summarize our findings.

We can write a general Lagrangian up to cubic order as

$$\mathcal{L}[\gamma_c] = \mathcal{L}_{GR}[\gamma_c] + \sum_i f_m^{(i)} \frac{(H\eta)^{m-4}}{\Lambda^{m-2}} \partial^m \gamma_c^2 + \sum_i g_n^{(i)} \frac{(H\eta)^{n-4}}{M_{\rm pl}\Lambda^{n-2}} \partial^n \gamma_c^3 + \mathcal{O}\left(\gamma_c^4\right) \,, \quad (8.265)$$

where $\gamma_c := M_{\rm pl}\gamma$ is the canonically normalized field, $\partial^n \gamma_c^m$ is a shorthand notation for an n-derivative operator and $f_n^{(i)}$ and $g_n^{(i)}$ are the dimensionless coupling constants. We have included all powers of η that are required by scale invariance. The parity-odd interactions that contribute to bispectra have n = 1, 2 and 3, and their dimensionless coupling constants are denoted by g_1, g_2, g_3 , respectively. At tree-level these operators do not correct the quadratic action which allows us to conclude that they can yield a large contribution to the graviton non-Gaussianity relative to the GR contribution if

$$(\mathcal{L}_3)_{\text{new}} \gg (\mathcal{L}_3)_{\text{GR}} \sim \frac{H}{M_{\text{P}}} \mathcal{L}_2$$
 (8.266)

In the above, \mathcal{L}_2 is simply the GR quadratic Lagrangian, as we have assumed throughout the chapter. Ideally, we want a stronger notion of a large non-Gaussianity, namely that the signal-to-noise (S-to-N) in the 3-point function is close to that of the power spectrum. This would mean

$$f_{\rm NL} \gamma \lesssim \mathcal{O}(1).$$
 (8.267)

Crucially the non-Gaussian contributions need to be smaller than the vacuum one to remain within a perturbative analysis. These two conditions entail, respectively, (at horizon crossing)

$$g_n \frac{H^{n-2}}{\Lambda^{n-2}} \gg 1, \tag{8.268}$$

$$g_n \frac{H^{n-2}}{\Lambda^{n-2}} \lesssim \frac{M_{\rm pl}}{H} \,, \tag{8.269}$$

and these would need to be satisfied for $H \ll M_{\rm pl}$ again so that the vacuum contribution dominates over the GR cubic contribution. We see that it is possible to have large non-Gaussianities relative to GR, while remaining perturbative. However, we must remember that the tree-level bispectra derived in this work are good approximations only if the loop contributions to the bispectra are suppressed, while the GR quadratic Lagrangian assumed throughout is only natural if loop corrections to it are insignificant. Let us therefore estimate (i) the size of quantum corrections to the quadratic Lagrangian and (ii) the size of loop contributions to the bispectra.

We start by estimating the UV cutoff scale Λ_c . This can be done by deriving the scale Λ_* at which the theory becomes strongly coupled, since at that energy new physics is expected to be important [273]. This corresponds to a limiting scenario where loop corrections are the largest, although it is still possible that the cutoff lies much below the strong coupling scale, which would correspond to a weakly coupled UV completion which we will comment on later. Now consider a general cubic operator with *n* derivatives. A rough estimate of the strong coupling scale can be derived by examining the breakdown of perturbative unitarity in flat-space i.e. by asking when the $\gamma\gamma \rightarrow \gamma\gamma$ scattering amplitude is of order 1. We work in flat space as this is a good approximation for energy scales well above the Hubble scale and indeed we want the theory to be valid in such a regime. A back-of-the-envelope estimate in the flat space limit yields

$$\mathcal{A}_{4} \sim g_{n}^{2} \frac{E^{n}}{M_{\rm pl}\Lambda^{n-2}} \frac{1}{E^{2}} \frac{E^{n}}{M_{\rm pl}\Lambda^{n-2}} = \left(g_{n} \frac{E^{n-1}}{M_{\rm pl}\Lambda^{n-2}}\right)^{2}, \qquad (8.270)$$

implying that the strong coupling scale Λ_* associated to an *n*-derivative operator is

$$\Lambda_* \sim \left(\frac{1}{g_n} M_{\rm pl} \Lambda^{n-2}\right)^{\frac{1}{n-1}}.$$
(8.271)

(For n = 1, we take $\Lambda_* \sim M_{\rm pl}$, since we expect new physics to be relevant at $M_{\rm pl}$, if not earlier.) For

the EFT to be useful, we should require that it is valid at least up to Hubble scale, so that $\Lambda_* > H$. Thanks to the presence of the $M_{\rm pl}$ factor, this is consistent with the above estimate of Λ_* , as well as with (8.268).

Let's now estimate the size of the loop corrections to the quadratic Lagrangian. First we focus our attention on a particular *n*-derivative operator and cut off the loop momentum at the relevant Λ_* given above. In the absence of a symmetry that would protect a small value of a given coupling, the radiative correction to the coefficient of a $(\partial^a \gamma_c)^2$ operator due to a loop with two *n*-derivative vertices is of the order

$$\delta \mathcal{L}_{(\partial^a \gamma_c)^2} \sim \frac{1}{M_{\rm pl} \Lambda^{n-2}} \frac{1}{M_{\rm pl} \Lambda^{n-2}} \int^{\Lambda_*} d^4 p g_n^2 \frac{p^{n-a} p^{n-a}}{p^4} \sim g_n^2 \frac{\Lambda_*^{2n-2a}}{M_{\rm pl}^2 \Lambda^{2n-4}} \sim \Lambda_*^{2-2a}.$$
 (8.272)

The ratio of the loop contribution $(\mathcal{L}_2)_{new}$ to the GR contribution at $E \sim H$ is of the order

$$\frac{(\mathcal{L}_2)_{\text{new}}}{(\mathcal{L}_2)_{\text{GR}}} \sim \left(\frac{\Lambda_*}{H}\right)^{2-2a}.$$
(8.273)

Now for a > 1 this is a small contribution since we take $\Lambda_* > H$. For a = 1 we would have a correction to the two derivative GR action but such corrections are harmless since we can always do field redefinitions that bring the quadratic action into the canonical form [245]. However, we see that the mass term (a = 0) could receive a large quantum correction. An important exception is for n = 3 where we have a shift symmetry. In this case a small graviton mass is protected by the shift symmetry of the interaction (8.183). For the other two operators (8.179)-(8.180) it looks like a large mass could be generated, but before jumping to such a conclusion one would need to perform a fully fledged computation to check if once all polarisation sums are included such a correction is still non-zero and large. We leave such an analysis for future work.

In the above we have assumed that there is only one cubic operator which not only generates radiative corrections, but also defines the cutoff scale. Suppose, however, that we have multiple cubic operators $\mathcal{O}_1, \ldots, \mathcal{O}_k$. If the g_n couplings do not differ by too many orders of magnitude, then the cutoff scale Λ_c is the one associated to the highest-dimension operator and this can alter our conclusions about large corrections to the mass. In the case of our three parity-odd interactions (8.179)-(8.180), (8.183) the lowest cutoff is $\Lambda_* = \sqrt{\frac{M_P\Lambda}{g_3}}$. The radiative correction to the coefficient at $(\partial^a \gamma)^2$ due to a loop

with two n-derivative vertices is then of the order

$$\delta \mathcal{L}_{(\partial^a \gamma_c)^2} \sim \frac{1}{M_{\rm pl} \Lambda^{n-2}} \frac{1}{M_{\rm pl} \Lambda^{n-2}} \int^{\Lambda_*} d^4 p g_n^2 \frac{p^{n-a} p^{n-a}}{p^4} \sim g_n^2 \frac{\Lambda_*^{2n-2a}}{M_{\rm pl}^2 \Lambda^{2n-4}} \sim \frac{g_n^2}{g_3^{n-a}} M_{\rm pl}^{n-a-2} \Lambda^{4-n}$$
(8.274)

Comparing this with the GR contribution at $E \sim H$, we have

$$\frac{(\mathcal{L}_2)_{\text{new}}}{(\mathcal{L}_2)_{\text{GR}}} \sim \frac{g_n^2}{g_3^{n-a}} \left(\frac{H^2}{M_{\text{P}}\Lambda}\right)^{a-1} \left(\frac{\Lambda}{M_{\text{P}}}\right)^{3-n}.$$
(8.275)

For $a \ge 1$ the corrections are small. For a = 0, only n = 1, 2 contribute due to the shift symmetry for n = 3. We have

$$n = 1: \frac{(\mathcal{L}_2)_{m^2}}{(\mathcal{L}_2)_{\rm GR}} \sim \frac{g_1^2}{g_3} \frac{\Lambda^3}{M_{\rm P} H^2}, \quad n = 2: \frac{(\mathcal{L}_2)_{m^2}}{(\mathcal{L}_2)_{\rm GR}} \sim \frac{g_2^2}{g_3^2} \frac{\Lambda^2}{H^2}.$$
(8.276)

We see that the g_1 could dominate the GR contribution $(f_{NL} \gg 1, \text{ but } f_{NL} \ll M_P/H)$ while keeping δm^2 small. This applies as long as the cutoff scale is dictated by the n = 3 operator which, as we have shown before, could be very large $(f_{NL} \sim M_P/H)$ since its loops do not correct the mass. On the other hand, we need a hierarchy $g_3 \gg g_2 \frac{\Lambda}{H}$ if the radiative corrections to m^2 from the n = 2 parity-odd operator are supposed to be small. In this case it is difficult to keep non-Gaussianity from g_2 larger than the GR non-Gaussianity, while keeping loop corrections under control.

Let us now study loop contributions to the parity-odd tree-level shapes we have computed in this chapter. We are interested mostly in the regime where the energy in the loop is large (close to Λ_*), so the loop is effectively deep inside the horizon. We can therefore again work in flat-space and estimate the size of the loop corrections to the three-particle *amplitude* and compare it with the tree-level result. We should be careful, however, to only put derivatives on the external legs in such a way that we reproduce the structure of one of our parity-odd operators, since otherwise the loop diagram will not contribute to the parity-odd bispectrum as we have shown in Section 8.4.5. We can therefore put m = 1, 2 or 3 derivatives on the external legs. An estimate of such a loop diagram proceeds similarly as before. For loop diagrams with three instances of the same *n*-derivative parity-odd operator, assuming the cutoff $\Lambda_* = \left(\frac{1}{g_n}M_{\rm P}\Lambda^{n-2}\right)^{1/(n-1)}$ is dictated by that operator, we find

$$\frac{\mathcal{A}_3^{1-\text{loop}}}{\mathcal{A}_3^{\text{tree}}} \sim g_n^{\frac{m-1}{n-1}} g_m^{-1} \left(\frac{M_{\text{P}}}{\Lambda}\right)^{\frac{n-m}{n-1}}.$$
(8.277)

- If n = 3, then shift symmetry must be preserved at any loop order, which means that loop diagrams cannot generate the shapes (8.175)-(8.178), but only the three-derivative shape. Thus, it suffices to consider m = 3. It seems that the ratio (8.277) is equal to 1, although in our estimate we neglected combinatorial factors as well as factors of (2π). In any case, identifying the cutoff with the perturbative unitarity breakdown scale is supposed to only give us an order-of-magnitude estimate, and it is not unnatural to have a slightly lower cutoff which would further suppress loops which scale as Λ⁴_c.
- If n = 2, then loop contributions are small for m = 3, again O(1) for m = 2, while they are large for m = 1. However, if the n = 3 interaction is also present and dominates the signal, then the cutoff is lowered from M_{pl}/g₂ to √M_{pl}Λ/g₃, and all the loop contributions from n = 2 are small.
- For n = 1, high energies are suppressed and we don't observe any UV divergences in the loop constructed out of three copies of the n = 1 operator. Instead, we ought to consider the loop constructed out of two GR vertices and one n = 1 operator. Regardless of the structure of derivatives on the external legs, this loop diagram is suppressed, relative to tree level, by M_{pl}⁻², but has at most two factors of Λ_{*} since at most four derivatives can be put on the internal legs. Therefore, loop contributions due to n = 1 are small.

Let us conclude this section by summarising our findings:

- If we consider the parity-odd operators *individually*, then g_3 (and only g_3) can be so large that the (8.182) bispectrum has a S-to-N ratio comparable to that of the power spectrum, without the need for fine-tuning. This is because of shift symmetry of (8.183), which protects a small mass from receiving quantum corrections. Meanwhile, the other two parity-odd operators do contribute to the mass via radiative corrections, and natural values of g_1 and g_2 must be very small, meaning that the associated signals are weaker than the GR bispectra.
- By identifying the cutoff scale of the theory with the scale at which the three-derivative parity-odd interactions given by (8.183) become strongly coupled, $\Lambda_c = \sqrt{M_{\rm P}\Lambda/g_3}$, we can consider a more general case in which we have multiple parity-odd operators. In this case, non-Gaussianities generated by (8.179) may be larger than GR non-Gaussianities (but with S-to-N smaller than in the power spectrum) while g_1 remains natural. However, the coefficient

of (8.180) is bounded by $g_2 \ll g_3 \frac{H}{\Lambda}$, implying that the region of parameter space where (8.177)-(8.178) are larger than GR is very limited.

• Tree level calculations are a good approximation for the three-derivative parity-odd interaction: loop contributions to the bispectrum can be suppressed without the need for fine-tuning. Overall, the g_3 operator is best placed to give large non-Gaussianities, both in the sense of being large compared to GR but also with a sizeable S-to-N, while keeping loop corrections under control.

8.6 Summary and future directions

In this work we have, for the first time, bootstrapped tree-level inflationary graviton bispectra to all orders in derivatives. Under a minimal set of assumptions, we have detailed how one can write down these bispectra without working with a concrete inflationary model. We used spatial translations, spatial rotations and scale invariance to write down a general ansatz for the corresponding wavefunction of the universe. Assuming that the mode functions are the usual ones of a massless graviton with Bunch-Davies initial conditions, we used locality and unitarity to constrain the wavefunction coefficients. We considered all possible tree-level contributions, including IR-divergences at future infinity, $\eta_0 \rightarrow 0$. We imposed locality by demanding that the wavefunction coefficients satisfy the Manifestly Local Test (MLT) introduced in [33] which is a simple differential constraint that all *n*-point functions of massless gravitons should satisfy. Solutions to the MLT replace solutions to the time integrals that one is required to calculate in the bulk formalism. The beauty of the MLT is that it allows us to compute non-Gaussian shapes without having to consider the unobservable bulk time evolution. We imposed bulk unitarity using the Cosmological Optical Theorem (COT) [12]. We presented our results succinctly in Section 8.4 using the cosmological spinor helicity formalism of [171], and we computed all bispectra for both parity-even and parity-odd interactions.

In Section 8.3, we showed which part of the wavefunction contributes to the correlator, for contact diagrams. We concentrated on contact diagrams since our focus in this chapter is on tree-level bispectra but many of our results in that section hold for any tree-level *n*-point function. We showed that only the part of the wavefuction that breaks the $\{k\} \rightarrow \{-k\}$ symmetry, where $\{k\}$ are the external energies, can contribute to the correlator. This is a direct consequence of bulk unitarity and can be easily derived from the COT for contact diagrams. For graviton bispectra, this tells us that for parity-even interactions both the rational part and the log part of the wavefunction can appear in

the correlator, whereas for parity-odd interactions the only allowed contributions are regular at both $\eta_0 \rightarrow 0$ and $k_T \rightarrow 0$. Indeed, unitarity in the form of the COT tells us that the log must always appear in the combination $\log(-k_T\eta_0) + \frac{i\pi}{2}$ and for parity-odd interactions it is the $\frac{i\pi}{2}$ piece that contributes to the correlator. This allowed us to show that, to all orders in derivatives, for parity-odd graviton self-interactions there are only three independent couplings that contribute to the bispectrum. This is not evident when using concrete Lagrangians and the in-in formalism and therefore offers a neat example of where the bootstrap approach can be very advantageous.

In Section 8.5, we showed that our parity-breaking graviton bispectra appear in both the Effective Field Theory of Inflation (EFToI), and in solid inflation. For the former, a correction to the two-point function is forced by the non-linearly realized symmetries. By accounting for this correction, we computed the full parity-odd contribution to the graviton bispectrum. The associated non-Gaussianity is too small to be detected observationally in any conceivable future. Conversely, for solid inflation there is no symmetry that forces a correction to the two-point function, so the three parity-odd bispectra we have computed can indeed arise with arbitrary coefficients. Given that such operators do not contribute to the bispectrum of curvature perturbations, which cannot violate parity, there are no strong observational bounds on the size of these non-Gaussianities. We plotted the associated shapes in Figure 8.5.

With this catalogue of graviton non-Gaussianities at hand, we outline here a few directions for future work

• To derive our catalogue of graviton bispectra we did not assume any particular symmetry breaking pattern for the inflationary dynamics. Indeed, we have captured all scale invariant contributions, assuming the usual massless mode functions. It would be very interesting to develop further criteria to identify those non-Gaussianites that are consistent only in the presence of additional degrees of freedom. For example, we expect that only some couplings can appear in the EFToI and in future work we plan to use soft theorems/consistency relations to extract this subset. It would also be very interesting to take these three-point building blocks and to glue them together to form four-point functions. By demanding that the full four-point function satisfies some consistency constraints, we will also be able to pick out interesting subsets of our full catalogue. This approach would be very similar to that used to constrain cubic interactions in flat space with *S*-matrix consistency conditions [29, 2], and in [181] assuming invariance

under de Sitter boosts. Deriving this full catalogue is the first step towards distinguishing between different symmetry breaking patterns for inflation directly at the level of the observable. This will complement the recent Lagrangian analysis of [248] and ultimately lead to a more efficient way of "simplifying" inflationary predictions [246]. For example, we expect there to be only a single three-derivative correction to the graviton bispectrum in the EFToI [247], and we plan to develop bootstrap techniques that enables us to efficiently extract this result without having to use the Lagrangian or bulk time evolution.

- Given the small number of possible parity-odd graviton bispectra in solid inflation, it would be interesting to study the associated bulk operators. In particular one would like to know when those same operators give also rise to interactions between the graviton and curvature perturbations. It is also very important to study the quantum stability of these operators and possible perturbative unitarity bounds on their size.
- Finally, we notice that Ref. [274] showed that for manifestly-local interactions, all parity-odd scalar correlation functions vanish at tree level. It would be interesting to see if their result can be generalized to spinning particles.

Our understanding of physical observables in nature becomes increasingly more opaque as we approach the real world. In anti-de Space (AdS) we have the gauge-gravity duality that provides us with a good understanding of the structure of boundary observables. In flat-space, the object of interest is the *S*-matrix. The *S*-matrix bootstrap has lead to a good understanding of the tree-level properties of amplitudes, with progress now being made on the analytic structure at loop level. Finally, we have de Sitter space, which appears to describe the early and late phases of our universe very well. We are only now starting to understand the general structure of cosmological correlators in de Sitter, both at tree and loop level. We hope that our results will contribute to broadening this understanding and to provide theoretical guidance on the physical modeling of inflation.

8.7 Appendix

8.7.1 From polarisations to spinors

In this appendix we construct all possible polarisation factors for three gravitons and explain how one can convert these into spinor expressions using the spinor helicity formalism. We consider parity-even
and parity-odd structures separately. Throughout we suppress the momentum dependence of the polarisation tensors. Note that throughout we only contract momenta with polarisation tensors as any pair of contracted momenta can be written in terms of the energies (norms) which we include in the trimmed part of the wavefunction c.f. (8.37). Indeed, we have

$$\mathbf{k}_{a} \cdot \mathbf{k}_{b} = \frac{1}{2} (k_{c}^{2} - k_{a}^{2} - k_{b}^{2}), \qquad a \neq b \neq c.$$
(8.278)

When we convert the following expressions into spinors, their symmetry properties will become manifest.

Parity-even tensor structures

For parity-even structures we need to contract spatial momenta with

$$e_{i_1i_2}^{h_1} e_{i_3i_4}^{h_2} e_{i_5i_6}^{h_3}, agenum{(8.279)}$$

using δ_{ij} . We work order by order in the total number of derivatives α .

lpha=0 In this case there is clearly only a single structure which is given by

$$e_{ij}^{h_1} e_{jk}^{h_2} e_{ki}^{h_3} \,. \tag{8.280}$$

This structure is fully symmetric and when converted to spinors this contraction simply yields

$$SH_{+++}$$
, (8.281)

for the all-plus configuration.

 $\alpha = 2$ In this case we have two possibilities. For the first we contract the two momenta with the same polarisation tensor and for the second we contract each momentum with different polarisations. Using momentum conservation and the transversality of the polarisation tensors, there is then a single option for the labels of the momenta, up to permutations. We have

$$e_{lm}^{h_1} e_{lm}^{h_2} e_{ij}^{h_3} k_1^i k_2^j$$
 and $e_{lm}^{h_1} e_{il}^{h_2} e_{jm}^{h_3} k_1^i k_1^j$. (8.282)

These two structures appear in GR with tuned coefficients. The first structure is symmetric in labels 1 and 2 while the second is symmetric in 2 and 3. If we sum over permutations and convert to spinors then we have

$$SH_{+++} \times Poly_2 = SH_{+++} \left(a_0 e_2 + a_2 k_T^2 \right)$$
 (8.283)

 $\alpha = 4$ In this case we have a single option. All momenta need to be contracted with polarisation tensors and then using the fact that the polarisations are traceless yields a single possibility. Again, momentum conservation and transversality yields a single possibility for the labels, up to permutations. We have

$$e_{lk}^{h_1} e_{mk}^{h_2} e_{ij}^{h_3} k_1^i k_2^j k_3^l k_3^m . aga{8.284}$$

This structure is symmetric in 2 and 3 and when we sum over permutations and convert to spinors we have

$$SH_{+++} \times Poly_4 = SH_{+++} \left(k_T^4 - k_T^2 e_2 + 8k_T e_3 \right) .$$
(8.285)

 $\alpha = 6$ Finally, in this case there is a single option with all polarisation tensor indices contracted with momenta. We have

$$e_{il}^{h_1} e_{jm}^{h_2} e_{kn}^{h_3} k_2^i k_3^l k_3^j k_1^m k_1^k k_2^n . aga{8.286}$$

This structure is fully symmetric and yields

$$\mathbf{SH}_{+++} \times \mathbf{Poly}_6 = \mathbf{SH}_{+++} \left(k_T^6 - 8k_T^4 e_2 + 16k_T^3 e_3 + 16k_T^2 e_2^2 - 64k_T e_2 e_3 + 64e_3^2 \right) .$$
(8.287)

when we convert to spinors.

Parity-odd tensor structures

We now turn to parity-odd structures where we need to contract momenta with

$$\epsilon_{i_1i_2i_3}e^{h_1}_{i_4i_5}e^{h_2}_{i_6i_7}e^{h_3}_{i_8i_9}. (8.288)$$

As above, in all cases there is a single option for the labels, up to permutations.

 $\alpha = 1$ In this case there are two possible structures with the single momentum either contracted with a polarisation tensor or with the epsilon tensor. We have

$$\epsilon_{ijk}e^{h_1}_{il}e^{h_2}_{lm}e^{h_3}_{km}k^j_3$$
 and $\epsilon_{ijk}e^{h_1}_{il}e^{h_2}_{jm}e^{h_3}_{kl}k^m_3$. (8.289)

The first of these is symmetric under the exchange $1 \leftrightarrow 2$, while the second is symmetric under $1 \leftrightarrow 3$. When symmetrized over all possible permutations of the three energies, these two contractions coincide up to a minus sign. This fact can be checked using explicit expressions for the polarization tensors, but it is not at all obvious. Conversely, it is easy to see in the spinor helicity formalism where both contractions must take the form

$$SH_{+++} \times Poly_1$$
, (8.290)

where the only permutation-invariant linear symmetric polynomial is $Poly_1 = k_T$.

 $\alpha = 3$ In this case we have six possibilities and we classify them according to how many momenta are contracted with the epsilon tensor. First consider the case where none of the momenta are contracted with the epsilon tensor. Given the properties of the polarisation tensors, we then have a single possibility given by

$$\epsilon_{ijk} e_{il}^{h_1} e_{jm}^{h_2} e_{kn}^{h_3} k_3^l k_1^m k_2^n .$$
(8.291)

Now when one of the momenta is contracted with the epsilon tensor we have two possibilities since the remaining two momenta can be contracted with the same polarisation tensor or with two different ones. We have

$$\epsilon_{ijk}e^{h_1}_{nl}e^{h_2}_{jm}e^{h_3}_{km}k^n_2k^i_3k^l_3 \quad , \quad \epsilon_{ijk}e^{h_1}_{jl}e^{h_2}_{nm}e^{h_3}_{kn}k^m_1k^l_2k^i_1 \quad \text{and} \quad \epsilon_{ijk}e^{h_1}_{jl}e^{h_2}_{nm}e^{h_3}_{kn}k^m_1k^l_2k^i_3 \ . \tag{8.292}$$

Finally, we can contract two momenta with the epsilon tensor. There are then two possibilities: the third momentum must be contracted with a polarisation tensor, and the other index of this polarisation can be contracted with the epsilon tensor or another polarisation. We have

$$\epsilon_{ijk}e_{nl}^{h_1}e_{nm}^{h_2}e_{il}^{h_3}k_1^mk_2^jk_3^k \quad \text{and} \quad \epsilon_{ijk}e_{im}^{h_1}e_{ln}^{h_2}e_{ln}^{h_3}k_1^jk_2^kk_3^m.$$
(8.293)

Upon symmetrization over all possible permutations of the three energies, only three of the above six contractions are linearly independent (for example (8.291) and the first two in (8.292)). To see this with explicit polarization tensors requires a laborious calculation. Conversely, this can be easily seen using spinor helicity variables, where the most generic $\alpha = 3$ (symmetrized) contraction must take the form

$$SH_{+++} \times Poly_3 = SH_{+++} \left(a_0 e_3 + a_1 k_T e_2 + a_3 k_T^3 \right),$$
 (8.294)

which has indeed three free coefficients $a_{0,1,3}$.

 $\alpha = 5$ In this case we have a total of three possibilities. One of them corresponds to having only one momentum contracted with the epsilon tensor, while for the others two of the momenta are contracted with the epsilon tensor. We have

$$\epsilon_{ijk}e^{h_1}_{mq}e^{h_2}_{nq}e^{h_3}_{lk}k^i_1k^j_2k^m_2k^n_1k^l_1, \quad \epsilon_{ijk}e^{h_1}_{mn}e^{h_2}_{lq}e^{h_3}_{qk}k^i_1k^j_2k^m_2k^n_1k^l_1, \text{ and } \epsilon_{ijk}e^{h_1}_{mn}e^{h_2}_{qj}e^{h_3}_{lk}k^i_2k^m_3k^q_1k^l_1k^n_3.$$
(8.295)

When we sum over permutations and convert to spinors we have only two structures:

$$SH_{+++}Poly_5 = a SH_{+++}(-3k_T^5 + 20k_T^3e_2 - 24k_T^2e_3 - 32k_Te_2^2 + 64e_2e_3)$$
(8.296)

$$+ b \operatorname{SH}_{+++}(k_T^5 - 8k_T^3 e_2 + 8k_T^2 e_3 + 16k_T e_2^2 - 32e_2 e_3).$$
(8.297)

lpha=7 Finally, in this last case we have a single possibility given by

$$\epsilon_{ijk}e^{h_1}_{mn}e^{h_2}_{qp}e^{h_3}_{lk}k^i_1k^j_2k^m_2k^n_2k^q_1k^p_1k^p_1k^l_1, \qquad (8.298)$$

and once we sum over permutations and convert to spinors we have

$$SH_{+++}Poly_7 = SH_{+++}(k_T^7 - 8k_T^5e_2 + 16k_T^4e_3 + 16k_T^3e_2^2 - 64k_T^2e_2e_3 + 16k_Te_3^2).$$
(8.299)

Note that in the above we have used the fact that three momenta cannot be contracted with the epsilon tensor due to momentum conservation.

Converting to spinors

Now that we have all of the possible polarisation factors, we can convert them into spinor expressions using the spinor helicity formalism. As we explained in detail in Section 8.4, given the form of the + + + polarisation factor, one can easily construct the ones for the other helicity configurations. The following expressions hold for three-point kinematics only. In the parity-even case the only expressions we need are

$$e^{a+} \cdot e^{b+} = -\frac{[ab]^2}{2k_a k_b},$$
 (8.300a)

$$p^a \cdot e^{b+} = \frac{(ab)[ab]}{2k_b},$$
 (8.300b)

where we have used the relations presented in Section 8.2.5. For parity-odd structures we use the general expression

$$\epsilon_{ijk}V_i^a V_j^b V_k^c = \frac{i}{4} (\langle ab \rangle [ab](cc) + \langle ab \rangle [ca](cb) + \langle bc \rangle [ab](ac)), \qquad (8.301)$$

where each SO(3) vector contains the spatial parts of a null four-vector V_{μ} which is converted to spinors using the standard expressions

$$V^{\mu} = -\frac{1}{2} (\bar{\sigma}^{\mu})^{\dot{\alpha}\alpha} V_{\alpha\dot{\alpha}}, \qquad V_{\alpha\dot{\alpha}} = V^{\mu} (\sigma_{\mu})_{\alpha\dot{\alpha}}, \qquad V_{\alpha\dot{\alpha}} = \lambda_{\alpha} \tilde{\lambda}_{\dot{\alpha}} . \tag{8.302}$$

The expressions that we need are then

$$\epsilon_{ijk} e_i^{a+} e_j^{b+} e_k^{c+} = i \frac{[ab][bc][ca]}{2k_a k_b k_c} , \qquad (8.303a)$$

$$\epsilon_{ijk} p_i^a e_j^{a+} e_k^{b+} = i \frac{[ab]^2}{2k_b} , \qquad (8.303b)$$

$$\epsilon_{ijk} p_i^a p_j^b e_k^{a+} = \frac{i}{2} [ab](ba) .$$
 (8.303c)

Note that by momentum conservation, we only need to consider cases where one of the momenta has the same label as one of the polarisation tensors. We used these relations to derive the list of possible $h_{\alpha}(k_1, k_2, k_3)$ in (8.85) to (8.92). Notice that for some α there are fewer choices for h_{α} compared to the polarisation structures above.

Chapter 9

Discussion

In this thesis I have emphasized that cosmic anisotropies and inhomogeneities can be traced back to primordial fluctuations generated during inflation. These fluctuations probe energies of order of the inflationary Hubble scale and therefore may serve as valuable data about high energy physics. I have therefore explored the problem of constraining inflationary correlators using the framework of the cosmological bootstrap. The importance of model-insensitive predictions that follow from reparametrization invariance of General Relativity [1], unitarity of time evolution and Bunch-Davies initial conditions [12, 14, 4], locality [33, 4] and little group scaling [2] follows from the fact that they can be used to guide the observational effort towards detection of particular signals. By relying on first principles, these techniques are an alternative to the EFT Lagrangian description, and can be employed to efficiently organize the correlators that originate from unknown UV corrections to General Relativity [4]. We have derived a number of consistency relations that are valid under wide assumptions which can be falsified if the consistency relation was observed to be violated. On the other hand, an observational confirmation would support the models in which the relations are valid.

In this final chapter, I summarize the main results presented in this thesis and sketch an outlook for future research in the area.

9.1 Summary of the main results

Cosmological correlators in a curved universe. Observations constrain the mean curvature of the universe to be small, consistent with zero [39], $\Omega_K = 0.0007 \pm 0.0019$ (68% CI). Still, the effects of curvature on primordial correlators could be of interest. As was first observed by [112], in the

EFT of Inflation these effects scale as K/c_s^2 , where c_s is the effective speed of sound. In Chapter 5, we derived $\mathcal{O}(K/c_s^2)$ corrections to the scalar power spectrum and the bispectrum and showed how the ratio Ω_K/c_s^2 can in principle be constrained by the low CMB multipoles, although at present the constraint is not meaningful in its regime of applicability. The discussed effect, linear in K, is model dependent and therefore cannot be captured by soft theorems, implying a violation of the Maldacena's soft theorem already at order K/q_l^2 (where q_l is the momentum of the longest mode) in a curved universe.

Constraints on boost-violating flat space theories. The on-shell S-matrix bootstrap is a rich program that has been successful in constructing higher-point amplitudes from simple building blocks, bypassing complex Feynman diagram computations. It encodes symmetry, locality and unitarity directly on the level of scattering amplitudes, using these general principles to constrain their form, often allowing only for one unique amplitude. In Chapters 6 and 7 ([2, 3]), we showed how these methods can be extended to flat space theories that do not respect boost invariance. We made the assumption that all particles are massless, propagate at the same speed and that Lorentz violations enter the action only through time derivatives. We found that such theories are tightly constrained: spin 1 particles cannot couple to themselves through a (+1, +1, -1) amplitude, while spin 2 particles may interact with themselves only in a Lorentz-invariant way. In fact, in the presence of a graviton all interactions that dominate in the IR must be Lorentz invariant, with a coefficient that is identical to the pure spin 2 coupling constant. We determined that (+S, +S, -S) amplitudes are only consistent for $S \leq 2$. In Chapter 7 we derived the conclusions for the interactions of identical particles under the same assumptions using the BCFW formalism, demonstrating its validity in a boost-violating setting. While the assumptions of Chapters 6 and 7 are quite restrictive, there is a possibility of giving them a physical interpretation related to the manner in which the theories considered therein are UV-completed by a Lorentz-invariant theory. It would be interesting to explore this issue further.

Constraints on graviton non-Gaussianities. In [4] (Chapter 8), we used new techniques developed in [171, 2, 33, 12] - the spinor-helicity formalism for boost-violating settings, the Manifestly Local Test and the Cosmological Optical Theorem - to constrain and catalogue all tree-level graviton three-point non-Gaussianities consistent with scale invariance. We did not restrict the analysis to any particular symmetry breaking pattern, and our results capture the physics of EFT of single field inflation as well as the Solid Inflation [130], the latter of which can in fact produce the most general graviton non-Gaussianities. Remarkably, we found only three parity-odd graviton bispectra in addition to infinitely many parity-even ones, for which we wrote down a general formula.

9.2 Outlook

Within the last several decades, there has been steady progress in the measurements of cosmological correlators [refs] and the inference of primordial fluctuations. On the large scales, the CMB power spectrum has been measured with a precision approaching the fundamental variance of the limited sample we have access to in a finite universe (the cosmic variance limit). The main focus of some upcoming experiments will be therefore the Large Scale Structure [275] and primordial gravitational waves [276, 277, 55], which should contain more information than the CMB alone. The current upper limit on the tensor-to-scalar ratio is $r_{0.05} < 0.036$ (95 % CL) [278], which is consistent with some single-field slow-roll models (see Section 2.3.1). The uncertainty is expected to be improved to $\sigma(r) \sim 0.003$ within the next few years which could confirm or exclude many models. The next generation experiments include the LiteBIRD mission (expected $\sigma(r) \sim 0.001$), planned to be launched in the late 2020s [279], and gravitational wave inteferometers such as LISA [55], sensitive to metric perturbations in the 10^{-2} Hz frequency range.

Given the cautious optimism about new data on scalar and tensor fluctuations, it is important to understand how scalar and graviton non-Gaussianities depend on inflation dynamics, field content and symmetry breaking patterns. The work outlined in this thesis makes several contributions to this project. In particular, in [4] (Chapter 8), we obtained shapes of tree-level tensor non-Gaussianities that can be large, including in the squeezed limit. If an expriment detects a tensor power spectrum, both the corresponding $\langle \gamma \gamma \gamma \rangle$ bispectrum and $\langle \gamma \gamma \zeta \rangle$ could be "just around the corner". Since such three-point correlators contain additional information, measuring them would provide even more valuable data about the early universe [280].

It is important to continue making progress on the theoretical side of the problem as well. A problem not discussed in this thesis are non-Gaussianities in the presence of massive degrees of freedom, including scalar or spinning particles. Recent developments can be found in [177, 178, 281], which predict a characteristic pattern of oscillations in the scalar bispectrum (as a function of momentum ratios) as a signature of massive particle exchange.

Virtually all the results for (quasi) de Sitter spacetime discussed in this thesis are perturbative in nature. Even those that are valid to any loop order, such as the Cosmological Optical Theorem or

the Manifestly Local Test assume a perturbative expansion of wavefunction coefficients. It would be worthwhile to explore the possibility of generalizing these results beyond perturbation theory, only using fundamental physical postulates such as unitarity and causality. Hopefully, the results of [1] might also be generalized to any order in background curvature K; [282] indeed derived the power spectrum in the special case where inflation is split into a phase dominated by the kinetic term and the slow-roll phase.

One of the remaining challenges is to explore the problem of consistent UV completions of inflationary EFTs. In a non-gravitational theory in flat space, assuming the existence of a UV completion leads to a set of inequalities that must be satisfied by the forward limit ($t \rightarrow 0$) of a four-particle scattering amplitude which can be expressed as a set of *positivity bounds* on the low energy EFT coefficients. These positivity bounds [283–287] have been studied in depth in the context of flat space but still do not have a full generalization to de Sitter space. Since amplitudes cannot be defined in the latter case, one could instead look for analogous constraints on wavefunction coefficients or correlators, but it is not known how to formulate such a procedure (but see [288]).

Effort is also being made to develop (quasi) de Sitter holography that aims to construct a boundary theory corresponding to the gravitational physics in the bulk spacetime. Since the IR limit in the bulk corresponds to the UV limit on the boundary and vice versa, such a holographic method could automate the computations of non-Gaussianities, including in the strongly coupled regime, and generate constraints that cannot be derived by studying the IR physics alone. We should certainly look forward to seeing new developments in this and related areas.

Bibliography

- G. Avis, S. Jazayeri, E. Pajer, and J. Supeł, "Spatial Curvature at the Sound Horizon," JCAP 02 (2020) 034, arXiv:1911.04454 [astro-ph.CO].
- [2] E. Pajer, D. Stefanyszyn, and J. Supeł, "The Boostless Bootstrap: Amplitudes without Lorentz boosts," JHEP 12 (2020) 198, arXiv: 2007.00027 [hep-th].
- [3] D. Stefanyszyn and J. Supeł, "The Boostless Bootstrap and BCFW Momentum Shifts," JHEP
 03 (2021) 091, arXiv:2009.14289 [hep-th].
- [4] G. Cabass, E. Pajer, D. Stefanyszyn, and J. Supeł, "Bootstrapping large graviton non-Gaussianities," JHEP 05 (2022) 077, arXiv:2109.10189 [hep-th].
- [5] G. Lemaître, "Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques," *Annales de la Société Scientifique de Bruxelles* 47 (Jan., 1927) 49–59.
- [6] E. Hubble, "A relation between distance and radial velocity among extra-galactic nebulae," *Proceedings of the National Academy of Sciences* 15 no. 3, (1929) 168–173, https://www.pnas.org/doi/pdf/10.1073/pnas.15.3.168.
- [7] R. H. Cyburt, B. D. Fields, K. A. Olive, and T.-H. Yeh, "Big bang nucleosynthesis: Present status," *Rev. Mod. Phys.* 88 (Feb, 2016) 015004. https://link.aps.org/doi/10.1103/RevModPhys.88.015004.
- [8] A. Bazavov, H.-T. Ding, P. Hegde, O. Kaczmarek, F. Karsch, N. Karthik, E. Laermann,
 A. Lahiri, R. Larsen, S.-T. Li, S. Mukherjee, H. Ohno, P. Petreczky, H. Sandmeyer, C. Schmidt,
 S. Sharma, and P. Steinbrecher, "Chiral crossover in qcd at zero and non-zero chemical

potentials," Physics Letters B 795 (2019) 15-21.

https://www.sciencedirect.com/science/article/pii/S0370269319303223.

- [9] ESA and the Planck Collaboration, "Planck CMB," March, 2013. https://www.esa.int/ESA_Multimedia/Images/2013/03/Planck_CMB. Accessed 22 March 2022. Licence: CC BY-SA IGO 3.0.
- [10] Y.-F. Cai, X. Chen, M. H. Namjoo, M. Sasaki, D.-G. Wang, and Z. Wang, "Revisiting non-Gaussianity from non-attractor inflation models," *JCAP* 05 (2018) 012, arXiv:1712.09998 [astro-ph.CO].
- [11] Planck Collaboration, Y. Akrami *et al.*, "Planck 2018 results. X. Constraints on inflation," arXiv:1807.06211 [astro-ph.CO].
- [12] H. Goodhew, S. Jazayeri, and E. Pajer, "The Cosmological Optical Theorem," JCAP 04 (2021) 021, arXiv:2009.02898 [hep-th].
- [13] H. Goodhew, S. Jazayeri, M. H. Gordon Lee, and E. Pajer, "Cutting cosmological correlators," JCAP 08 (2021) 003, arXiv: 2104.06587 [hep-th].
- [14] S. Melville and E. Pajer, "Cosmological Cutting Rules," JHEP 05 (2021) 249, arXiv:2103.09832 [hep-th].
- [15] J. M. Maldacena, "The Large N limit of superconformal field theories and supergravity," Adv. Theor. Math. Phys. 2 (1998) 231–252, arXiv:hep-th/9711200.
- [16] V. E. Hubeny, "The AdS/CFT Correspondence," Class. Quant. Grav. 32 no. 12, (2015) 124010, arXiv:1501.00007 [gr-qc].
- [17] I. R. Klebanov and J. M. Maldacena, "Solving quantum field theories via curved spacetimes," *Phys. Today* 62 (2009) 28–33.
- [18] E. Witten, "Anti-de Sitter space and holography," Adv. Theor. Math. Phys. 2 (1998) 253-291, arXiv:hep-th/9802150.
- [19] S. Ryu and T. Takayanagi, "Holographic derivation of entanglement entropy from AdS/CFT," *Phys. Rev. Lett.* 96 (2006) 181602, arXiv:hep-th/0603001.

- [20] J. Casalderrey-Solana, H. Liu, D. Mateos, K. Rajagopal, and U. A. Wiedemann, *Gauge/String Duality, Hot QCD and Heavy Ion Collisions*. Cambridge University Press, 2014. arXiv:1101.0618 [hep-th].
- [21] A. Christodoulou and K. Skenderis, "Holographic Construction of Excited CFT States," JHEP 04 (2016) 096, arXiv:1602.02039 [hep-th].
- [22] J. Maldacena, "Non-gaussian features of primordial fluctuations in single field inflationary models," arXiv:astro-ph/0210603.
- [23] A. Strominger, "The dS / CFT correspondence," JHEP 10 (2001) 034, arXiv:hep-th/0106113.
- [24] A. Strominger, "Inflation and the dS / CFT correspondence," JHEP 11 (2001) 049, arXiv:hep-th/0110087.
- [25] E. Witten, "Quantum gravity in de Sitter space," in *Strings 2001: International Conference*. 6, 2001. arXiv:hep-th/0106109.
- [26] P. McFadden and K. Skenderis, "Cosmological 3-point correlators from holography," JCAP 06 (2011) 030, arXiv:1104.3894 [hep-th].
- [27] A. Bzowski, P. McFadden, and K. Skenderis, "Holography for inflation using conformal perturbation theory," JHEP 04 (2013) 047, arXiv:1211.4550 [hep-th].
- [28] C. Cheung, "TASI Lectures on Scattering Amplitudes," in Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics : Anticipating the Next Discoveries in Particle Physics (TASI 2016): Boulder, CO, USA, June 6-July 1, 2016, pp. 571–623. 2018. arXiv:1708.03872 [hep-ph].
- [29] P. Benincasa and F. Cachazo, "Consistency Conditions on the S-Matrix of Massless Particles," arXiv:0705.4305 [hep-th].
- [30] P. Benincasa, "New structures in scattering amplitudes: a review," Int. J. Mod. Phys. A29 no. 5, (2014) 1430005, arXiv:1312.5583 [hep-th].
- [31] H. Elvang and Y.-t. Huang, "Scattering Amplitudes," arXiv:1308.1697 [hep-th].

- [32] P. Benincasa, "New structures in scattering amplitudes: a review," arXiv:1312.5583.
- [33] S. Jazayeri, E. Pajer, and D. Stefanyszyn, "From locality and unitarity to cosmological correlators," JHEP 10 (2021) 065, arXiv:2103.08649 [hep-th].
- [34] S. Nadathur, "Seeing patterns in noise: Gigaparsec-scale 'structures' that do not violate homogeneity," *Mon. Not. Roy. Astron. Soc.* 434 (2013) 398–406, arXiv:1306.1700 [astro-ph.CO].
- [35] R. Wald, *General Relativity*. University of Chicago Press, 1984. https://books.google.co.uk/books?id=FQgAmQEACAAJ.
- [36] T. Clifton, P. G. Ferreira, A. Padilla, and C. Skordis, "Modified Gravity and Cosmology," *Phys. Rept.* 513 (2012) 1–189, arXiv:1106.2476 [astro-ph.CO].
- [37] E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess, and J. Silk, "In the realm of the Hubble tension—a review of solutions," *Class. Quant. Grav.* 38 no. 15, (2021) 153001, arXiv:2103.01183 [astro-ph.CO].
- [38] E. Di Valentino and A. Melchiorri, "Neutrino Mass Bounds in the era of Tension Cosmology," arXiv:2112.02993 [astro-ph.CO].
- [39] Planck Collaboration, N. Aghanim et al., "Planck 2018 results. VI. Cosmological parameters," arXiv:1807.06209 [astro-ph.CO].
- [40] D. J. Fixsen, "The Temperature of the Cosmic Microwave Background," Astrophys. J. 707 (2009) 916–920, arXiv:0911.1955 [astro-ph.CO].
- [41] K. Lake, "The Flatness problem and lambda," *Phys. Rev. Lett.* 94 (2005) 201102, arXiv:astro-ph/0404319.
- [42] G. Efstathiou, J. R. Bond, and S. D. M. White, "COBE Background radiation anisotropies and large scale structure in the universe," *Mon. Not. Roy. Astron. Soc.* 258 (1992) 1–6.
- [43] WMAP Collaboration, C. L. Bennett *et al.*, "First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Preliminary maps and basic results," *Astrophys. J. Suppl.* 148 (2003) 1–27, arXiv:astro-ph/0302207.

- [44] Planck Collaboration, N. Aghanim *et al.*, "Planck 2018 results I. Overview and the cosmological legacy of Planck," A&A 641 (2020) A1. https://doi.org/10.1051/0004-6361/201833880.
- [45] Planck Collaboration, N. Aghanim *et al.*, "Planck 2018 results V. CMB power spectra and likelihoods," *A&A* 641 (2020) A5. https://doi.org/10.1051/0004-6361/201936386.
- [46] Planck Collaboration, Y. Akrami *et al.*, "Planck 2018 results VII. Isotropy and statistics of the CMB," A&A 641 (2020) A7. https://doi.org/10.1051/0004-6361/201935201.
- [47] Planck Collaboration, N. Aghanim *et al.*, "Planck 2018 results VIII. Gravitational lensing," A&A 641 (2020) A8. https://doi.org/10.1051/0004-6361/201833886.
- [48] Planck Collaboration, N. Aghanim *et al.*, "Planck 2018 results IX. Constraints on primordial non-Gaussianity," A&A 641 (2020) A9. https://doi.org/10.1051/0004-6361/201935891.
- [49] B. Pandey and S. Sarkar, "Probing large scale homogeneity and periodicity in the LRG distribution using Shannon entropy," *Mon. Not. Roy. Astron. Soc.* 460 no. 2, (2016) 1519–1528, arXiv:1512.06350 [astro-ph.CO].
- [50] G. Zhi-yuan, "Testing the homogeneity of large-scale structure with the sdss data," *Chinese Astronomy and Astrophysics* 34 no. 1, (2010) 23–29.
 https://www.sciencedirect.com/science/article/pii/S0275106209001349.
- [51] A. H. Guth and S. H. H. Tye, "Phase Transitions and Magnetic Monopole Production in the Very Early Universe," *Phys. Rev. Lett.* 44 (1980) 631.
- [52] A. H. Guth, "The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems," *Phys. Rev. D* 23 (1981) 347–356.
- [53] K. Dimopoulos, "Ultra slow-roll inflation demystified," Phys. Lett. B 775 (2017) 262–265, arXiv:1707.05644 [hep-ph].
- [54] J. Martin, "Inflationary cosmological perturbations of quantum-mechanical origin," Lect. Notes Phys. 669 (2005) 199–244, arXiv:hep-th/0406011.

- [55] P. Campeti, E. Komatsu, D. Poletti, and C. Baccigalupi, "Measuring the spectrum of primordial gravitational waves with CMB, PTA and Laser Interferometers," *JCAP* 01 (2021) 012, arXiv:2007.04241 [astro-ph.CO].
- [56] M. Kamionkowski and E. D. Kovetz, "The quest for b modes from inflationary gravitational waves," Annual Review of Astronomy and Astrophysics 54 no. 1, (2016) 227–269, https://doi.org/10.1146/annurev-astro-081915-023433.
- [57] BICEP2, Keck Array Collaboration, P. A. R. Ade *et al.*, "BICEP2 / Keck Array x: Constraints on Primordial Gravitational Waves using Planck, WMAP, and New BICEP2/Keck Observations through the 2015 Season," *Phys. Rev. Lett.* 121 (2018) 221301, arXiv:1810.05216 [astro-ph.CO].
- [58] P. Adshead, R. Easther, and E. A. Lim, "The "in-in" formalism and cosmological perturbations," arXiv:0904.4207.
- [59] S. Weinberg, "Quantum contributions to cosmological correlations," *Phys. Rev. D* 72 (2005) 043514, arXiv:hep-th/0506236.
- [60] D. Baumann, Cosmology. Cambridge University Press, 7, 2022.
- [61] J. Fergusson, "Part III Cosmology." http://www.damtp.cam.ac.uk//user/examples/3R2La.pdf. Unpublished lecture notes, accessed 08/08/2022.
- [62] D. Tong, "Cosmology." http://www.damtp.cam.ac.uk/user/tong/cosmo/cosmo.pdf. Unpublished lecture notes, accessed 08/08/2022.
- [63] WMAP Collaboration, E. Komatsu *et al.*, "Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation," *Astrophys. J. Suppl.* 180 (2009) 330–376, arXiv:0803.0547 [astro-ph].
- [64] R. G. Crittenden and N. Turok, "Looking for Lambda with the Rees-Sciama effect," *Phys. Rev. Lett.* 76 (1996) 575, arXiv:astro-ph/9510072.
- [65] A. Lewis and A. Challinor, "Weak gravitational lensing of the CMB," *Phys. Rept.* 429 (2006)
 1-65, arXiv:astro-ph/0601594.

- [66] M. Birkinshaw, "The Sunyaev-Zel'dovich effect," Phys. Rept. 310 (1999) 97–195, arXiv:astro-ph/9808050.
- [67] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, "Large scale structure of the universe and cosmological perturbation theory," *Phys. Rept.* 367 (2002) 1–248, arXiv:astro-ph/0112551.
- [68] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, "The Effective Field Theory of Cosmological Large Scale Structures," JHEP 09 (2012) 082, arXiv:1206.2926 [astro-ph.CO].
- [69] R. A. Porto, L. Senatore, and M. Zaldarriaga, "The Lagrangian-space Effective Field Theory of Large Scale Structures," JCAP 05 (2014) 022, arXiv:1311.2168 [astro-ph.CO].
- [70] R. A. Porto, "The effective field theorist's approach to gravitational dynamics," *Phys. Rept.*633 (2016) 1–104, arXiv:1601.04914 [hep-th].
- [71] L. Senatore and M. Zaldarriaga, "The IR-resummed Effective Field Theory of Large Scale Structures," JCAP 02 (2015) 013, arXiv:1404.5954 [astro-ph.CO].
- [72] J. J. M. Carrasco, S. Foreman, D. Green, and L. Senatore, "The Effective Field Theory of Large Scale Structures at Two Loops," JCAP 07 (2014) 057, arXiv:1310.0464 [astro-ph.CO].
- [73] J. Adamek, D. Daverio, R. Durrer, and M. Kunz, "gevolution: a cosmological N-body code based on General Relativity," JCAP 07 (2016) 053, arXiv:1604.06065 [astro-ph.CO].
- [74] L. H. Garrison, D. J. Eisenstein, and P. A. Pinto, "A High-Fidelity Realization of the Euclid Code Comparison N-body Simulation with Abacus," *Mon. Not. Roy. Astron. Soc.* 485 no. 3, (2019) 3370–3377, arXiv:1810.02916 [astro-ph.CO].
- [75] H. K. Dreiner, H. E. Haber, and S. P. Martin, "Two-component spinor techniques and Feynman rules for quantum field theory and supersymmetry," *Phys. Rept.* 494 (2010) 1–196, arXiv:0812.1594 [hep-ph].
- [76] "Mandelstam variables." https://en.wikipedia.org/wiki/Mandelstam_variables, 2022.

- [77] D. Griffiths, Introduction to elementary particles. 2008.
- [78] S. Weinberg, *The Quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, 6, 2005.
- [79] M. Dorato and E. Rossanese, "The nature of representation in Feynman diagrams," *Perspect. Sci.* 26 (2018) 443–458, arXiv:1711.03790 [physics.hist-ph].
- [80] S. J. Parke and T. R. Taylor, "Gluonic Two Goes to Four," Nucl. Phys. B 269 (1986) 410-420.
- [81] S. J. Parke and T. R. Taylor, "Amplitude for *n*-gluon scattering," *Phys. Rev. Lett.* 56 (Jun, 1986) 2459–2460. https://link.aps.org/doi/10.1103/PhysRevLett.56.2459.
- [82] F. A. Berends and W. T. Giele, "Recursive Calculations for Processes with n Gluons," *Nucl. Phys. B* 306 (1988) 759–808.
- [83] N. Arkani-Hamed, J. L. Bourjaily, F. Cachazo, A. B. Goncharov, A. Postnikov, and J. Trnka, *Grassmannian Geometry of Scattering Amplitudes*. Cambridge University Press, 4, 2016. arXiv:1212.5605 [hep-th].
- [84] M. E. Peskin and D. V. Schroeder, An Introduction to quantum field theory. Addison-Wesley, Reading, USA, 1995.
- [85] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3, 2014.
- [86] A. Peres, "Causality in S matrix theory," Annals of Physics 37 (1966) 179–208.
- [87] J. R. Hance, J. Rarity, and J. Ladyman, "Could wavefunctions simultaneously represent knowledge and reality?," *Quant. Stud. Math. Found.* 9 no. 3, (2022) 333–341, arXiv:2101.06436 [quant-ph].
- [88] D. Skinner, "Quantum Field Theory II." http://www.damtp.cam.ac.uk/user/dbs26/AQFT.html. Unpublished lecture notes, accessed 08/08/2022.
- [89] G. V. Dunne, "Functional determinants in quantum field theory," J. Phys. A 41 (2008) 304006, arXiv:0711.1178 [hep-th].

- [90] E. Pajer, G. L. Pimentel, and J. V. S. Van Wijck, "The Conformal Limit of Inflation in the Era of CMB Polarimetry," JCAP 1706 no. 06, (2017) 009, arXiv:1609.06993 [hep-th].
- [91] X. Tong, Y. Wang, and Y. Zhu, "Cutting rule for cosmological collider signals: a bulk evolution perspective," JHEP 03 (2022) 181, arXiv:2112.03448 [hep-th].
- [92] D. Baumann, W.-M. Chen, C. Duaso Pueyo, A. Joyce, H. Lee, and G. L. Pimentel, "Linking the Singularities of Cosmological Correlators," arXiv:2106.05294 [hep-th].
- [93] 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].
- [94] T. S. Floss, "Inflationary consistency conditions and shift-symmetric cosmologies," Master's thesis, 2018.
- [95] L. Hui, A. Joyce, and S. S. C. Wong, "Inflationary soft theorems revisited: A generalized consistency relation," JCAP 1902 (2019) 060, arXiv:1811.05951 [hep-th].
- [96] E. Pajer and S. Jazayeri, "Systematics of Adiabatic Modes: Flat Universes," JCAP 1803 no. 03, (2018) 013, arXiv:1710.02177 [astro-ph.CO].
- [97] E. Pajer, "Lecture Notes on Cosmological Soft Theorems." http://www.damtp.cam.ac.uk/user/ep551/notes_cosmo_soft_theorems.pdf, 2019. Accessed: 2022-07-29.
- [98] B. Finelli, G. Goon, E. Pajer, and L. Santoni, "Soft Theorems For Shift-Symmetric Cosmologies," *Phys. Rev.* D97 no. 6, (2018) 063531, arXiv:1711.03737 [hep-th].
- [99] G. L. Pimentel, "Inflationary Consistency Conditions from a Wavefunctional Perspective," JHEP 02 (2014) 124, arXiv:1309.1793 [hep-th].
- [100] S. Weinberg, "Adiabatic modes in cosmology," Phys. Rev. D 67 (2003) 123504, arXiv:astro-ph/0302326.
- [101] R. G. Barrera, G. A. Estevez, and J. Giraldo, "Vector spherical harmonics and their application to magnetostatics," *European Journal of Physics* 6 no. 4, (Oct, 1985) 287–294. https://doi.org/10.1088/0143-0807/6/4/014.

- [102] F. K. Manasse and C. W. Misner, "Fermi Normal Coordinates and Some Basic Concepts in Differential Geometry," J. Math. Phys. 4 (1963) 735–745.
- [103] T. Baldauf, U. Seljak, L. Senatore, and M. Zaldarriaga, "Galaxy Bias and non-Linear Structure Formation in General Relativity," JCAP 1110 (2011) 031, arXiv:1106.5507 [astro-ph.CO].
- [104] 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].
- [105] L. Dai, E. Pajer, and F. Schmidt, "Conformal Fermi Coordinates," *JCAP* 1511 no. 11, (2015)
 043, arXiv:1502.02011 [gr-qc].
- [106] J. M. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," JHEP 05 (2003) 013, arXiv:astro-ph/0210603 [astro-ph].
- [107] P. Creminelli and M. Zaldarriaga, "Single field consistency relation for the 3-point function," JCAP 0410 (2004) 006, arXiv:astro-ph/0407059 [astro-ph].
- [108] K. Hinterbichler, L. Hui, and J. Khoury, "Conformal Symmetries of Adiabatic Modes in Cosmology," JCAP 1208 (2012) 017, arXiv:1203.6351 [hep-th].
- [109] V. Assassi, D. Baumann, and D. Green, "On Soft Limits of Inflationary Correlation Functions," JCAP 1211 (2012) 047, arXiv:1204.4207 [hep-th].
- [110] P. Creminelli, G. D'Amico, M. Musso, and J. Noreña, "The (not so) squeezed limit of the primordial 3-point function," *JCAP* 1111 (2011) 038, arXiv:1106.1462 [astro-ph.CO].
- [111] P. Creminelli, J. Noreña, and M. Simonović, "Conformal consistency relations for single-field inflation," JCAP 1207 (2012) 052, arXiv:1203.4595 [hep-th].
- [112] P. Creminelli, A. Perko, L. Senatore, M. Simonović, and G. Trevisan, "The Physical Squeezed Limit: Consistency Relations at Order q²," JCAP 1311 (2013) 015, arXiv:1307.0503 [astro-ph.CO].

- [113] L. Berezhiani and J. Khoury, "Slavnov-Taylor Identities for Primordial Perturbations," JCAP
 1402 (2014) 003, arXiv:1309.4461 [hep-th].
- [114] L. Berezhiani, J. Khoury, and J. Wang, "Non-Trivial Checks of Novel Consistency Relations," JCAP 1406 (2014) 056, arXiv:1401.7991 [hep-th].
- [115] N. Kundu, A. Shukla, and S. P. Trivedi, "Constraints from Conformal Symmetry on the Three Point Scalar Correlator in Inflation," JHEP 04 (2015) 061, arXiv:1410.2606 [hep-th].
- [116] N. Kundu, A. Shukla, and S. P. Trivedi, "Ward Identities for Scale and Special Conformal Transformations in Inflation," JHEP 01 (2016) 046, arXiv:1507.06017 [hep-th].
- [117] L. Bordin, P. Creminelli, M. Mirbabayi, and J. Noreña, "Tensor Squeezed Limits and the Higuchi Bound," JCAP 1609 no. 09, (2016) 041, arXiv:1605.08424 [astro-ph.CO].
- [118] S. Weinberg, "Adiabatic modes in cosmology," Phys. Rev. D67 (2003) 123504, arXiv:astro-ph/0302326 [astro-ph].
- [119] M. Mirbabayi and M. Zaldarriaga, "CMB Anisotropies from a Gradient Mode," JCAP 1503 no. 03, (2015) 056, arXiv:1409.4777 [astro-ph.CO].
- [120] W. H. Kinney, "Horizon crossing and inflation with large eta," Phys. Rev. D72 (2005) 023515, arXiv:gr-qc/0503017 [gr-qc].
- [121] B. Finelli, G. Goon, E. Pajer, and L. Santoni, "The Effective Theory of Shift-Symmetric Cosmologies," arXiv:1802.01580 [hep-th].
- [122] M. H. Namjoo, H. Firouzjahi, and M. Sasaki, "Violation of non-Gaussianity consistency relation in a single field inflationary model," *EPL* 101 no. 3, (2013) 39001, arXiv:1210.3692 [astro-ph.CO].
- [123] X. Chen, H. Firouzjahi, M. H. Namjoo, and M. Sasaki, "A Single Field Inflation Model with Large Local Non-Gaussianity," *EPL* 102 no. 5, (2013) 59001, arXiv:1301.5699 [hep-th].

- [124] M. Akhshik, H. Firouzjahi, and S. Jazayeri, "Cosmological Perturbations and the Weinberg Theorem," JCAP 1512 no. 12, (2015) 027, arXiv:1508.03293 [hep-th].
- [125] S. Mooij and G. A. Palma, "Consistently violating the non-Gaussian consistency relation," JCAP 1511 no. 11, (2015) 025, arXiv:1502.03458 [astro-ph.CO].
- [126] R. Bravo, S. Mooij, G. A. Palma, and B. Pradenas, "A generalized non-Gaussian consistency relation for single field inflation," arXiv:1711.02680 [astro-ph.CO].
- [127] L. Bordin, P. Creminelli, M. Mirbabayi, and J. Noreña, "Solid Consistency," JCAP 1703 no. 03, (2017) 004, arXiv:1701.04382 [astro-ph.CO].
- [128] E. Pajer, S. Jazayeri, and D. van der Woude, "Solid Soft Theorems," arXiv:1902.09020 [hep-th].
- [129] A. Gruzinov, "Elastic inflation," Phys. Rev. D70 (2004) 063518, arXiv:astro-ph/0404548 [astro-ph].
- [130] S. Endlich, A. Nicolis, and J. Wang, "Solid Inflation," JCAP 10 (2013) 011, arXiv:1210.0569 [hep-th].
- [131] N. Bartolo, D. Cannone, A. Ricciardone, and G. Tasinato, "Distinctive signatures of space-time diffeomorphism breaking in EFT of inflation," *JCAP* 1603 no. 03, (2016) 044, arXiv:1511.07414 [astro-ph.CO].
- [132] A. Ricciardone and G. Tasinato, "Primordial gravitational waves in supersolid inflation," *Phys. Rev.* D96 no. 2, (2017) 023508, arXiv:1611.04516 [astro-ph.CO].
- [133] F. Piazza, D. Pirtskhalava, R. Rattazzi, and O. Simon, "Gaugid inflation," JCAP 1711 no. 11, (2017) 041, arXiv:1706.03402 [hep-th].
- [134] T. Fujita, S. Kuroyanagi, S. Mizuno, and S. Mukohyama, "Blue-tilted Primordial Gravitational Waves from Massive Gravity," *Phys. Lett.* B789 (2019) 215–219, arXiv:1808.02381
 [gr-qc].
- [135] X. Xiao, "Note on Adiabatic Modes and Ward Identities In A Closed Universe," arXiv:1407.8486 [hep-th].

- [136] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, "The Effective Field Theory of Inflation," *JHEP* 03 (2008) 014, arXiv:0709.0293 [hep-th].
- [137] D. Seery, J. E. Lidsey, and M. S. Sloth, "The inflationary trispectrum," arXiv:astro-ph/0610210.
- [138] D. Seery, M. S. Sloth, and F. Vernizzi, "Inflationary trispectrum from graviton exchange," JCAP 0903 (2009) 018, arXiv:0811.3934 [astro-ph].
- [139] X. Chen, B. Hu, M. xin Huang, G. Shiu, and Y. Wang, "Large primordial trispectra in general single field inflation," arXiv:0905.3494.
- [140] W. Handley, "Primordial power spectra for curved inflating universes," arXiv:1907.08524 [astro-ph.CO].
- [141] T. Tanaka and M. Sasaki, "The Spectrum of gravitational wave perturbations in the one bubble open inflationary universe," *Prog. Theor. Phys.* 97 (1997) 243–262, arXiv:astro-ph/9701053 [astro-ph].
- [142] D. H. Lyth and A. Woszczyna, "Large scale perturbations in the open universe," *Phys. Rev.* D52 (1995) 3338–3357, arXiv:astro-ph/9501044 [astro-ph].
- [143] J. Garriga, X. Montes, M. Sasaki, and T. Tanaka, "Spectrum of cosmological perturbations in the one bubble open universe," *Nucl. Phys.* B551 (1999) 317–373, arXiv:astro-ph/9811257 [astro-ph].
- [144] M. Sasaki, T. Tanaka, and K. Yamamoto, "Euclidean vacuum mode functions for a scalar field on open de Sitter space," *Phys. Rev.* D51 (1995) 2979–2995, arXiv:gr-qc/9412025 [gr-qc].
- [145] K. Yamamoto, M. Sasaki, and T. Tanaka, "Large angle CMB anisotropy in an open universe in the one bubble inflationary scenario," Astrophys. J. 455 (1995) 412–418, arXiv:astro-ph/9501109 [astro-ph].
- [146] K. Yamamoto, M. Sasaki, and T. Tanaka, "Quantum fluctuations and CMB anisotropies in one bubble open inflation models," *Phys. Rev.* D54 (1996) 5031–5048, arXiv:astro-ph/9605103 [astro-ph].

- [147] 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 12 (2006) 080, arXiv:hep-th/0606090 [hep-th].
- [148] K. M. Smith, L. Senatore, and M. Zaldarriaga, "Optimal analysis of the CMB trispectrum," arXiv:1502.00635 [astro-ph.CO].
- [149] 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].
- [150] "Planck legacy archive." https://pla.esac.esa.int.
- [151] J. M. Stewart, "Perturbations of Friedmann-Robertson-Walker cosmological models," *Class. Quant. Grav.* 7 (1990) 1169–1180.
- [152] B. Ratra, "Inflation in a closed universe," Phys. Rev. D96 no. 10, (2017) 103534, arXiv:1707.03439 [astro-ph.CO].
- [153] R. Bluhm, "Overview of the SME: Implications and phenomenology of Lorentz violation," Lect. Notes Phys. 702 (2006) 191–226, arXiv:hep-ph/0506054.
- [154] V. A. Kostelecky and N. Russell, "Data Tables for Lorentz and CPT Violation," *Rev. Mod. Phys.* 83 (2011) 11–31, arXiv:0801.0287 [hep-ph].
- [155] J. D. Tasson, "What Do We Know About Lorentz Invariance?," *Rept. Prog. Phys.* 77 (2014) 062901, arXiv:1403.7785 [hep-ph].
- [156] C. M. Will, "The Confrontation between General Relativity and Experiment," *Living Rev. Rel.* **17** (2014) 4, arXiv:1403.7377 [gr-qc].
- [157] S. Weinberg, "What is quantum field theory, and what did we think it is?," in Conference on Historical Examination and Philosophical Reflections on the Foundations of Quantum Field Theory, pp. 241–251. 3, 1996. arXiv:hep-th/9702027.
- [158] R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The analytic S-matrix*. Cambridge Univ. Press, Cambridge, 1966.

- [159] R. Britto, F. Cachazo, and B. Feng, "New recursion relations for tree amplitudes of gluons," *Nucl. Phys. B* 715 (2005) 499–522, arXiv:hep-th/0412308.
- [160] R. Britto, F. Cachazo, B. Feng, and E. Witten, "Direct proof of tree-level recursion relation in Yang-Mills theory," *Phys. Rev. Lett.* 94 (2005) 181602, arXiv:hep-th/0501052.
- [161] P. Benincasa, C. Boucher-Veronneau, and F. Cachazo, "Taming Tree Amplitudes In General Relativity," JHEP 11 (2007) 057, arXiv:hep-th/0702032.
- [162] A. Nicolis, R. Penco, F. Piazza, and R. Rattazzi, "Zoology of condensed matter: Framids, ordinary stuff, extra-ordinary stuff," *JHEP* 06 (2015) 155, arXiv:1501.03845 [hep-th].
- [163] LIGO Scientific, Virgo, Fermi-GBM, INTEGRAL Collaboration, B. P. Abbott *et al.*, "Gravitational Waves and Gamma-rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A," *Astrophys. J. Lett.* 848 no. 2, (2017) L13, arXiv:1710.05834 [astro-ph.HE].
- [164] L. Alberte and A. Nicolis, "Spontaneously broken boosts and the Goldstone continuum," arXiv:2001.06024 [hep-th].
- [165] Planck Collaboration, N. Aghanim *et al.*, "Planck 2013 results. XXVII. Doppler boosting of the CMB: Eppur si muove," *Astron. Astrophys.* 571 (2014) A27, arXiv:1303.5087 [astro-ph.CO].
- [166] A. Nicolis and F. Piazza, "Spontaneous Symmetry Probing," JHEP 06 (2012) 025, arXiv:1112.5174 [hep-th].
- [167] T. Grall, S. Jazayeri, and D. Stefanyszyn, "The cosmological phonon: symmetries and amplitudes on sub-horizon scales," JHEP 11 (2020) 097, arXiv:2005.12937 [hep-th].
- [168] D. Green and E. Pajer, "On the Symmetries of Cosmological Perturbations," arXiv:2004.09587 [hep-th].
- [169] D. Baumann, D. Green, and T. Hartman, "Dynamical Constraints on RG Flows and Cosmology," JHEP 12 (2019) 134, arXiv:1906.10226 [hep-th].

- [170] S. R. Coleman and J. Mandula, "All Possible Symmetries of the S Matrix," *Phys. Rev.* 159 (1967) 1251–1256.
- [171] J. M. Maldacena and G. L. Pimentel, "On graviton non-Gaussianities during inflation," JHEP
 09 (2011) 045, arXiv:1104.2846 [hep-th].
- [172] S. Raju, "New Recursion Relations and a Flat Space Limit for AdS/CFT Correlators," Phys. Rev. D 85 (2012) 126009, arXiv:1201.6449 [hep-th].
- [173] E. Pajer, "The Symmetries of Cosmological Correlators,". https://drive.google.com/file/d/15BG9LiUziqxUWB4-CTW7OIFoRL3YTOSC/view.
- [174] P. Creminelli, "Conformal invariance of scalar perturbations in inflation," Phys. Rev. D 85 (2012) 041302, arXiv:1108.0874 [hep-th].
- [175] I. Mata, S. Raju, and S. Trivedi, "CMB from CFT," JHEP 07 (2013) 015, arXiv:1211.5482 [hep-th].
- [176] A. Ghosh, N. Kundu, S. Raju, and S. P. Trivedi, "Conformal Invariance and the Four Point Scalar Correlator in Slow-Roll Inflation," JHEP 07 (2014) 011, arXiv:1401.1426 [hep-th].
- [177] N. Arkani-Hamed, D. Baumann, H. Lee, and G. L. Pimentel, "The Cosmological Bootstrap: Inflationary Correlators from Symmetries and Singularities," JHEP 04 (2020) 105, arXiv:1811.00024 [hep-th].
- [178] D. Baumann, C. Duaso Pueyo, A. Joyce, H. Lee, and G. L. Pimentel, "The cosmological bootstrap: weight-shifting operators and scalar seeds," *JHEP* 12 (2020) 204, arXiv:1910.14051 [hep-th].
- [179] C. Sleight, "A Mellin Space Approach to Cosmological Correlators," JHEP 01 (2020) 090, arXiv:1906.12302 [hep-th].
- [180] C. Sleight and M. Taronna, "Bootstrapping Inflationary Correlators in Mellin Space," JHEP 02 (2020) 098, arXiv:1907.01143 [hep-th].

- [181] D. Baumann, C. Duaso Pueyo, A. Joyce, H. Lee, and G. L. Pimentel, "The Cosmological Bootstrap: Spinning Correlators from Symmetries and Factorization," arXiv:2005.04234 [hep-th].
- [182] N. Arkani-Hamed and J. Maldacena, "Cosmological Collider Physics," arXiv:1503.08043 [hep-th].
- [183] N. Arkani-Hamed, P. Benincasa, and A. Postnikov, "Cosmological Polytopes and the Wavefunction of the Universe," arXiv:1709.02813 [hep-th].
- [184] N. Arkani-Hamed and P. Benincasa, "On the Emergence of Lorentz Invariance and Unitarity from the Scattering Facet of Cosmological Polytopes," arXiv:1811.01125 [hep-th].
- [185] P. Benincasa, "From the flat-space S-matrix to the Wavefunction of the Universe," arXiv:1811.02515 [hep-th].
- [186] P. Benincasa, "Cosmological Polytopes and the Wavefuncton of the Universe for Light States," arXiv:1909.02517 [hep-th].
- [187] A. Hillman, "Symbol Recursion for the dS Wave Function," arXiv:1912.09450 [hep-th].
- [188] P. Creminelli and F. Vernizzi, "Dark Energy after GW170817 and GRB170817A," Phys. Rev. Lett. 119 no. 25, (2017) 251302, arXiv:1710.05877 [astro-ph.CO].
- [189] J. Sakstein and B. Jain, "Implications of the Neutron Star Merger GW170817 for Cosmological Scalar-Tensor Theories," *Phys. Rev. Lett.* **119** no. 25, (2017) 251303, arXiv:1710.05893 [astro-ph.CO].
- [190] E. J. Copeland, M. Kopp, A. Padilla, P. M. Saffin, and C. Skordis, "Dark energy after GW170817 revisited," *Phys. Rev. Lett.* **122** no. 6, (2019) 061301, arXiv:1810.08239 [gr-qc].
- [191] L. Bordin, E. J. Copeland, and A. Padilla, "Dark energy loopholes some time after GW170817," arXiv:2006.06652 [astro-ph.CO].

- [192] LIGO Scientific, Virgo Collaboration, B. Abbott *et al.*, "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral," *Phys. Rev. Lett.* 119 no. 16, (2017) 161101, arXiv:1710.05832 [gr-qc].
- [193] A. Goldstein *et al.*, "An Ordinary Short Gamma-Ray Burst with Extraordinary Implications: Fermi-GBM Detection of GRB 170817A," *Astrophys. J. Lett.* 848 no. 2, (2017) L14, arXiv:1710.05446 [astro-ph.HE].
- [194] V. A. Kostelecky and J. D. Tasson, "Constraints on Lorentz violation from gravitational Cerenkov radiation," *Phys. Lett.* B749 (2015) 551–559, arXiv:1508.07007 [gr-qc].
- [195] C. Talmadge, J. P. Berthias, R. W. Hellings, and E. M. Standish, "Model Independent Constraints on Possible Modifications of Newtonian Gravity," *Phys. Rev. Lett.* 61 (1988) 1159–1162.
- [196] LIGO Scientific, Virgo Collaboration, B. Abbott *et al.*, "Tests of general relativity with GW150914," *Phys. Rev. Lett.* 116 no. 22, (2016) 221101, arXiv:1602.03841 [gr-qc].
 [Erratum: Phys.Rev.Lett. 121, 129902 (2018)].
- [197] C. de Rham, J. T. Deskins, A. J. Tolley, and S.-Y. Zhou, "Graviton Mass Bounds," *Rev. Mod. Phys.* 89 no. 2, (2017) 025004, arXiv:1606.08462 [astro-ph.CO].
- [198] S. R. Choudhury, G. C. Joshi, S. Mahajan, and B. H. J. McKellar, "Probing large distance higher dimensional gravity from lensing data," *Astropart. Phys.* 21 (2004) 559–563, arXiv:hep-ph/0204161 [hep-ph].
- [199] G. Dvali, A. Gruzinov, and M. Zaldarriaga, "The Accelerated universe and the moon," *Phys. Rev.* D68 (2003) 024012, arXiv:hep-ph/0212069 [hep-ph].
- [200] M. P. Hertzberg and M. Sandora, "Special Relativity from Soft Gravitons," Phys. Rev. D 96 no. 8, (2017) 084048, arXiv:1704.05071 [hep-th].
- [201] S. Weinberg, "Infrared photons and gravitons," Phys. Rev. 140 (1965) B516–B524.
- [202] S. Weinberg, "Photons and gravitons in perturbation theory: Derivation of Maxwell's and Einstein's equations," *Phys. Rev.* 138 (1965) B988–B1002.

- [203] V. A. Kostelecky, "Gravity, Lorentz violation, and the standard model," Phys. Rev. D69 (2004) 105009, arXiv:hep-th/0312310 [hep-th].
- [204] J. Khoury, G. E. J. Miller, and A. J. Tolley, "On the Origin of Gravitational Lorentz Covariance," *Class. Quant. Grav.* **31** (2014) 135011, arXiv:1305.0822 [hep-th].
- [205] H. K. Dreiner, H. E. Haber, and S. P. Martin, "Two-component spinor techniques and Feynman rules for quantum field theory and supersymmetry," *Phys. Rept.* 494 (2010) 1–196, arXiv:0812.1594 [hep-ph].
- [206] M. P. Hertzberg, J. A. Litterer, and M. Sandora, "Symmetries from Locality Part 2: Gravitation and Lorentz Boosts," arXiv:2005.01744 [hep-th].
- [207] M. P. Hertzberg and J. A. Litterer, "Symmetries from Locality Part 1: Electromagnetism and Charge Conservation," *Phys. Rev. D* 102 no. 2, (2020) 025022, arXiv:2005.01731 [hep-th].
- [208] D. A. McGady and L. Rodina, "Higher-spin massless S-matrices in four-dimensions," Phys. Rev. D90 no. 8, (2014) 084048, arXiv:1311.2938 [hep-th].
- [209] P. C. Schuster and N. Toro, "Constructing the Tree-Level Yang-Mills S-Matrix Using Complex Factorization," JHEP 06 (2009) 079, arXiv:0811.3207 [hep-th].
- [210] N. Arkani-Hamed, T.-C. Huang, and Y.-t. Huang, "Scattering Amplitudes For All Masses and Spins," arXiv:1709.04891 [hep-th].
- [211] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3, 2014.
- [212] S. Weinberg, *The Quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, 2005.
- [213] S. J. Parke and T. R. Taylor, "An Amplitude for n Gluon Scattering," Phys. Rev. Lett. 56 (1986) 2459.
- [214] A. R. White, "The Past and future of S matrix theory," arXiv:hep-ph/0002303 [hep-ph].

- [215] M. E. Peskin and D. V. Schroeder, An Introduction to quantum field theory. Addison-Wesley, Reading, USA, 1995. http://www.slac.stanford.edu/~mpeskin/QFT.html.
- [216] C. Chandler, "Causality in S matrix theory," *Phys. Rev.* 174 (1968) 1749–1758. [Erratum: Phys. Rev.179,1632(1969)].
- [217] D. Olive, "Exploration of s-matrix theory," Physical Review 135 no. 3B, (1964) 745–760.
- [218] P. Benincasa and E. Conde, "Exploring the S-Matrix of Massless Particles," Phys. Rev. D 86 (2012) 025007, arXiv:1108.3078 [hep-th].
- [219] Z. Bern, J. Carrasco, and H. Johansson, "New Relations for Gauge-Theory Amplitudes," *Phys. Rev. D* 78 (2008) 085011, arXiv:0805.3993 [hep-ph].
- [220] D. Ponomarev and E. Skvortsov, "Light-Front Higher-Spin Theories in Flat Space," J. Phys. A 50 no. 9, (2017) 095401, arXiv:1609.04655 [hep-th].
- [221] D. Baumann, D. Green, H. Lee, and R. A. Porto, "Signs of Analyticity in Single-Field Inflation," *Phys. Rev.* D93 no. 2, (2016) 023523, arXiv:1502.07304 [hep-th].
- [222] X. Bekaert, N. Boulanger, and P. Sundell, "How higher-spin gravity surpasses the spin two barrier: no-go theorems versus yes-go examples," *Rev. Mod. Phys.* 84 (2012) 987–1009, arXiv:1007.0435 [hep-th].
- [223] T. Grall and S. Melville, "Inflation in motion: unitarity constraints in effective field theories with (spontaneously) broken Lorentz symmetry," JCAP 09 (2020) 017, arXiv:2005.02366 [gr-qc].
- [224] N. Afshordi, D. J. Chung, and G. Geshnizjani, "Cuscuton: A Causal Field Theory with an Infinite Speed of Sound," *Phys. Rev. D* 75 (2007) 083513, arXiv:hep-th/0609150.
- [225] C. Cheung, K. Kampf, J. Novotny, C.-H. Shen, and J. Trnka, "On-Shell Recursion Relations for Effective Field Theories," *Phys. Rev. Lett.* **116** no. 4, (2016) 041601, arXiv:1509.03309 [hep-th].
- [226] A. Padilla, D. Stefanyszyn, and T. Wilson, "Probing Scalar Effective Field Theories with the Soft Limits of Scattering Amplitudes," JHEP 04 (2017) 015, arXiv:1612.04283 [hep-th].

- [227] B. Feng, K. Zhou, C. Qiao, and J. Rao, "Determination of Boundary Contributions in Recursion Relation," JHEP 03 (2015) 023, arXiv:1411.0452 [hep-th].
- [228] S. Raju, "Four Point Functions of the Stress Tensor and Conserved Currents in AdS₄/CFT₃," *Phys. Rev. D* 85 (2012) 126008, arXiv:1201.6452 [hep-th].
- [229] P. Benincasa and E. Conde, "On the Tree-Level Structure of Scattering Amplitudes of Massless Particles," JHEP 11 (2011) 074, arXiv:1106.0166 [hep-th].
- [230] C. Sleight and M. Taronna, "From AdS to dS Exchanges: Spectral Representation, Mellin Amplitudes and Crossing," arXiv:2007.09993 [hep-th].
- [231] E. Pajer, "Building a Boostless Bootstrap for the Bispectrum," JCAP 01 (2021) 023, arXiv:2010.12818 [hep-th].
- [232] J. Bonifacio, E. Pajer, and D. G. Wang, "A Boostless Bootstrap for the Trispectrum," arXiv:2105.xxxx.
- [233] L. Di Pietro, V. Gorbenko, and S. Komatsu, "Analyticity and Unitarity for Cosmological Correlators," arXiv:2108.01695 [hep-th].
- [234] C. Sleight and M. Taronna, "From dS to AdS and back," arXiv:2109.02725 [hep-th].
- [235] M. Hogervorst, J. a. Penedones, and K. S. Vaziri, "Towards the non-perturbative cosmological bootstrap," arXiv:2107.13871 [hep-th].
- [236] D. Meltzer, "The Inflationary Wavefunction from Analyticity and Factorization," arXiv:2107.10266 [hep-th].
- [237] C. Sleight and M. Taronna, "On the consistency of (partially-)massless matter couplings in de Sitter space," arXiv:2106.00366 [hep-th].
- [238] S. Céspedes, A.-C. Davis, and S. Melville, "On the time evolution of cosmological correlators," arXiv:2009.07874 [hep-th].
- [239] H. Gomez, R. L. Jusinskas, and A. Lipstein, "Cosmological Scattering Equations," arXiv:2106.11903 [hep-th].

- [240] J. Soda, H. Kodama, and M. Nozawa, "Parity Violation in Graviton Non-gaussianity," JHEP 08 (2011) 067, arXiv:1106.3228 [hep-th].
- [241] A. Bzowski, P. McFadden, and K. Skenderis, "Implications of conformal invariance in momentum space," JHEP 03 (2014) 111, arXiv:1304.7760 [hep-th].
- [242] S. Jain, R. R. John, A. Mehta, A. A. Nizami, and A. Suresh, "Momentum space parity-odd CFT 3-point functions," arXiv:2101.11635 [hep-th].
- [243] S. Jain, R. R. John, A. Mehta, A. A. Nizami, and A. Suresh, "Higher spin 3-point functions in 3d CFT using spinor-helicity variables," arXiv:2106.00016 [hep-th].
- [244] C. Armendariz-Picon, T. Damour, and V. F. Mukhanov, "k inflation," Phys. Lett. B 458 (1999) 209-218, arXiv:hep-th/9904075.
- [245] 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].
- [246] L. Bordin, G. Cabass, P. Creminelli, and F. Vernizzi, "Simplifying the EFT of Inflation: generalized disformal transformations and redundant couplings," JCAP 09 (2017) 043, arXiv:1706.03758 [astro-ph.CO].
- [247] L. Bordin and G. Cabass, "Graviton non-Gaussianities and Parity Violation in the EFT of Inflation," JCAP 07 (2020) 014, arXiv: 2004.00619 [astro-ph.CO].
- [248] G. Cabass, "Zoology of Graviton non-Gaussianities," arXiv:2103.09816 [hep-th].
- [249] N. Bartolo, L. Caloni, G. Orlando, and A. Ricciardone, "Tensor non-Gaussianity in chiral scalar-tensor theories of gravity," JCAP 03 (2021) 073, arXiv:2008.01715 [astro-ph.CO].
- [250] N. Bartolo and G. Orlando, "Parity breaking signatures from a Chern-Simons coupling during inflation: the case of non-Gaussian gravitational waves," JCAP 07 (2017) 034, arXiv:1706.04627 [astro-ph.CO].

- [251] M. Shiraishi, D. Nitta, and S. Yokoyama, "Parity Violation of Gravitons in the CMB Bispectrum," Prog. Theor. Phys. 126 (2011) 937–959, arXiv:1108.0175 [astro-ph.CO].
- [252] J. Guven, B. Lieberman, and C. T. Hill, "Schrodinger Picture Field Theory in Robertson-walker Flat Space-times," *Phys. Rev. D* 39 (1989) 438.
- [253] D. Anninos, T. Anous, D. Z. Freedman, and G. Konstantinidis, "Late-time Structure of the Bunch-Davies De Sitter Wavefunction," JCAP 11 (2015) 048, arXiv:1406.5490 [hep-th].
- [254] G. Goon, K. Hinterbichler, A. Joyce, and M. Trodden, "Shapes of gravity: Tensor non-Gaussianity and massive spin-2 fields," JHEP 10 (2019) 182, arXiv:1812.07571 [hep-th].
- [255] J. B. Hartle and S. W. Hawking, "Wave Function of the Universe," *Phys. Rev. D* 28 (1983) 2960–2975.
- [256] R. Holman and A. J. Tolley, "Enhanced Non-Gaussianity from Excited Initial States," JCAP 05 (2008) 001, arXiv:0710.1302 [hep-th].
- [257] D. Green and R. A. Porto, "Signals of a Quantum Universe," *Phys. Rev. Lett.* 124 no. 25, (2020) 251302, arXiv:2001.09149 [hep-th].
- [258] O. Aharony, L. F. Alday, A. Bissi, and E. Perlmutter, "Loops in AdS from Conformal Field Theory," JHEP 07 (2017) 036, arXiv:1612.03891 [hep-th].
- [259] D. Meltzer, E. Perlmutter, and A. Sivaramakrishnan, "Unitarity Methods in AdS/CFT," JHEP 03 (2020) 061, arXiv:1912.09521 [hep-th].
- [260] D. Meltzer and A. Sivaramakrishnan, "CFT unitarity and the AdS Cutkosky rules," JHEP 11 (2020) 073, arXiv:2008.11730 [hep-th].
- [261] V. Gorbenko and L. Senatore, " $\lambda \phi^4$ in dS," arXiv:1911.00022 [hep-th].
- [262] T. Cohen and D. Green, "Soft de Sitter Effective Theory," JHEP 12 (2020) 041, arXiv:2007.03693 [hep-th].

- [263] H. Goodhew and E. Pajer, "In preparation,".
- [264] K. Skenderis, "Lecture notes on holographic renormalization," Class. Quant. Grav. 19 (2002) 5849–5876, arXiv:hep-th/0209067.
- [265] A. Hillman and E. Pajer, "A differential representation of cosmological wavefunctions," JHEP04 (2022) 012, arXiv:2112.01619 [hep-th].
- [266] J. Bonifacio, E. Pajer, and D.-G. Wang, "From Amplitudes to Contact Cosmological Correlators," arXiv:2106.15468 [hep-th].
- [267] E. Komatsu and D. N. Spergel, "Acoustic signatures in the primary microwave background bispectrum," Phys. Rev. D 63 (2001) 063002, arXiv:astro-ph/0005036.
- [268] D. T. Son, "Low-energy quantum effective action for relativistic superfluids," arXiv:hep-ph/0204199.
- [269] S. L. Dubovsky, "Phases of massive gravity," JHEP 10 (2004) 076, arXiv:hep-th/0409124.
- [270] S. Endlich, B. Horn, A. Nicolis, and J. Wang, "Squeezed limit of the solid inflation three-point function," *Phys. Rev.* D90 no. 6, (2014) 063506, arXiv:1307.8114 [hep-th].
- [271] S. R. Behbahani, A. Dymarsky, M. Mirbabayi, and L. Senatore, "(Small) Resonant non-Gaussianities: Signatures of a Discrete Shift Symmetry in the Effective Field Theory of Inflation," JCAP 1212 (2012) 036, arXiv:1111.3373 [hep-th].
- [272] A. J. Duivenvoorden, P. D. Meerburg, and K. Freese, "CMB B-mode non-Gaussianity: optimal bispectrum estimator and Fisher forecasts," *Phys. Rev. D* 102 no. 2, (2020) 023521, arXiv:1911.11349 [astro-ph.CO].
- [273] D. Baumann and D. Green, "Equilateral Non-Gaussianity and New Physics on the Horizon," JCAP 09 (2011) 014, arXiv:1102.5343 [hep-th].
- [274] T. Liu, X. Tong, Y. Wang, and Z.-Z. Xianyu, "Probing P and CP Violations on the Cosmological Collider," JHEP 04 (2020) 189, arXiv:1909.01819 [hep-ph].

- [275] N. Sailer, E. Castorina, S. Ferraro, and M. White, "Cosmology at high redshift a probe of fundamental physics," *JCAP* 12 no. 12, (2021) 049, arXiv:2106.09713 [astro-ph.CO].
- [276] C. Caprini and D. G. Figueroa, "Cosmological Backgrounds of Gravitational Waves," *Class. Quant. Grav.* 35 no. 16, (2018) 163001, arXiv:1801.04268 [astro-ph.CO].
- [277] M. C. Guzzetti, N. Bartolo, M. Liguori, and S. Matarrese, "Gravitational waves from inflation," *Riv. Nuovo Cim.* **39** no. 9, (2016) 399–495, arXiv:1605.01615 [astro-ph.CO].
- [278] BICEP, Keck Collaboration, P. A. R. Ade *et al.*, "Improved Constraints on Primordial Gravitational Waves using Planck, WMAP, and BICEP/Keck Observations through the 2018 Observing Season," *Phys. Rev. Lett.* 127 no. 15, (2021) 151301, arXiv:2110.00483
 [astro-ph.CO].
- [279] LiteBIRD Collaboration, M. Hazumi *et al.*, "LiteBIRD: JAXA's new strategic L-class mission for all-sky surveys of cosmic microwave background polarization," *Proc. SPIE Int. Soc. Opt. Eng.* 11443 (2020) 114432F, arXiv:2101.12449 [astro-ph.IM].
- [280] G. Orlando, "Probing parity-odd bispectra with anisotropies of GW V modes," arXiv:2206.14173 [astro-ph.CO].
- [281] G. L. Pimentel and D.-G. Wang, "Boostless Cosmological Collider Bootstrap," arXiv:2205.00013 [hep-th].
- [282] Z. Shumaylov and W. Handley, "Primordial power spectra from k-inflation with curvature," *Phys. Rev. D* 105 no. 12, (2022) 123532, arXiv:2112.07547 [astro-ph.CO].
- [283] C. de Rham, S. Melville, A. J. Tolley, and S.-Y. Zhou, "UV complete me: Positivity Bounds for Particles with Spin," JHEP 03 (2018) 011, arXiv:1706.02712 [hep-th].
- [284] A. J. Tolley, Z.-Y. Wang, and S.-Y. Zhou, "New positivity bounds from full crossing symmetry," arXiv:2011.02400 [hep-th].
- [285] B. Bellazzini, "Softness and amplitudes' positivity for spinning particles," JHEP 02 (2017) 034, arXiv:1605.06111 [hep-th].

- [286] S. Melville and J. Noller, "Positivity in the Sky: Constraining dark energy and modified gravity from the UV," *Phys. Rev. D* 101 no. 2, (2020) 021502, arXiv:1904.05874 [astro-ph.CO]. [Erratum: Phys.Rev.D 102, 049902 (2020)].
- [287] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, "Causality, analyticity and an IR obstruction to UV completion," *JHEP* 10 (2006) 014, arXiv:hep-th/0602178.
- [288] T. Grall and S. Melville, "Positivity Bounds without Boosts," arXiv:2102.05683 [hep-th].