Non-Gaussianity from self-coupling in inflation
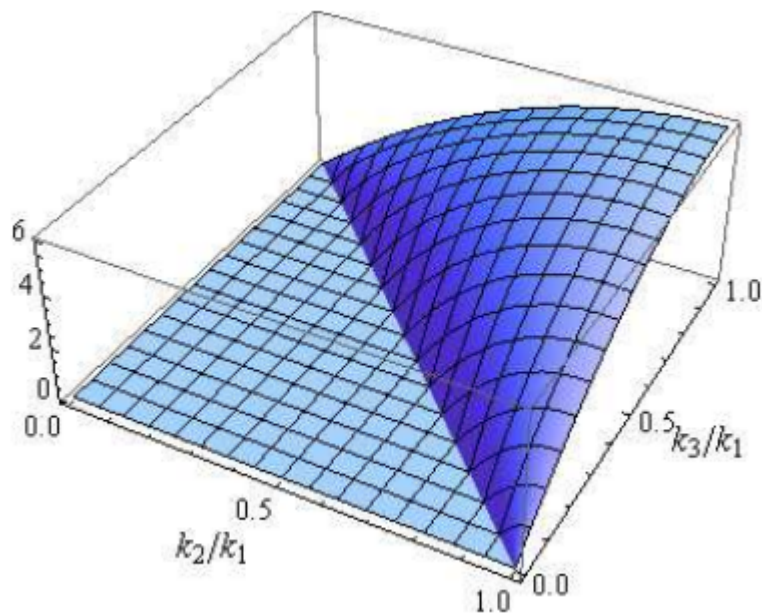
Utrecht University

Master's thesis - Theoretical Physics - Cosmology

Author:
Richard J. Dijk

Supervisor:
Dr. Enrico Pajer

15th December 2015

Abstract

Measurement of anisotropies in the cosmic microwave background (CMB) can give a wealth of information about the period of inflation that is assumed to have occurred in the early universe. The perturbations that cause these anisotropies are predominantly Gaussian, but there is the freedom for small non-Gaussianity, which is directly represented in the bispectrum (three-point function) of the perturbations. The non-Gaussianities can encode information on the underlying theory of inflation, but measuring these is hard, so theoretical work on the shape of the bispectrum is required. We will discuss how the bispectrum is typically calculated in single-field scalar models using slow-roll approximations. We calculate for the canonical scalar field model the contribution of an interaction term at next-to-leading order, not coming from coupling to gravity but from self-coupling. This term is proportional to $\epsilon\dot{\eta}$, using Hubble slow-roll parameters. We find that the contribution diverges and needs to be compensated by a boundary term in time that is commonly ignored in literature and we conclude that the slow-roll combination $\epsilon\dot{\eta}$ is not present in the bispectrum.

Contents

Chapter 1

Introduction

1.1 Introduction

The early universe is thought to have undergone a period of accelerated, exponential expansion, called inflation. There are many models of inflation, with the simplest model consisting of a single scalar field with canonical kinetic term and an arbitrary potential, subject to slow-roll restrictions, coupled to gravity. Classically, the scalar field reproduces a period of inflation for an isotropic universe, compatible with measurements of the cosmic microwave background (CMB). The CMB is almost isotropic but has small anisotropies.

When the classical field is quantised to become a quantum field theory (in curved space-time), the quantum fluctuations of the field can be a natural generator for these anisotropies. The field can be decomposed semi-classically, into an isotropic background that reproduces inflation and small perturbations on this background. The background can be kept classical and the perturbations quantised. The quantised perturbations can then be studied theoretically and the predictions of their primary observables, the two-field and three-field correlators of perturbations, can be compared against the measurement data of the CMB.

From the CMB, the correlators must be approximately Gaussian, but a small amount of non-Gaussianity is allowed. Gaussianity is characterised by vanishing of odd numbered correlators and so small non-Gaussianity is characterised by odd-numbered correlators deviating from zero. Considering now scalar perturbations, which give the predominant contributions to the CMB temperature anisotropies, the two-point function gives information about the size of the perturbations, whereas the three-point function gives information on the non-Gaussianity of the perturbations.

Two 'sources' of non-Gaussianity can be distinguished. A free scalar field is Gaussian, but when it's coupled to gravity, because gravity is a non-linear theory, non-Gaussianity is automatically generated. Second, (self-)interaction terms in the action for the scalar field directly contribute non-Gaussianity. Literature results for the leading non-Gaussianity, coming from the coupling to gravity, of the canonical scalar field are known.

There are many different models of inflation and so far very few have been ruled out by observations. Generally, they can be divided into two categories: single-field and multi-field inflation. In this thesis, we focus our attention only on the single-field models, with a particular focus on the canonical scalar field, which is arguably the simplest. Definitions and methods of calculation reviewed here can be directly related to other single-field models.

It would help our understanding of inflation if more models could be ruled out. One can wait for measurements to improve and rule out models, of course, but it might also be possible to combine the bounds from data on different parameters in novel ways to strengthen the bounds. For example, the Lyth bound[19], which comes from the ratio (r) of the two-point functions of tensor perturbations to scalar perturbations, was strengthened using the spectral index[17] (n_{s}), which is the leading-order scale-dependence of the two-point function of scalar perturbations. There are also restrictions on how much more measurements can be improved, before effects other than sensitivity of measurement equipment become the limiting factor.

Self-interactions should play a role not only in the three-point function, but also in the scale-dependence of the two-point function (in the running, α, of the

spectral index), so relating the two might give better bounds on one or the other. This is one motivation for studying the non-Gaussianity. Another motivation for studying non-Gaussianity is because all interactions between perturbations are reflected only in the three-point function and not the two-point function. Said in another way, all the hints of 'interesting physics' are in the non-Gaussianity. For example, in [1] the authors view inflation as a 'cosmological particle collider' and showed that new particles created during inflation with masses close to the Hubble scale could leave characteristic imprints on the squeezed limit of the three-point function. It is therefore interesting, in general, to be able to calculate the three-point function for a theory of inflation.

In this thesis, we calculate the contribution to the three-point function of scalar perturbations of a term in the action, the $\epsilon\dot{\eta}$-term, related to self-interactions of the inflaton ($V'''\delta\phi^3 \subset S$). We find a divergence that should not be present and then from literature a boundary term that not only cancels this divergence, but cancels the entire term in the action. This boundary terms and others like it cannot, therefore, be neglected in deriving the action for perturbations. The calculation follows the methodology of [20] and has been described in detail and generality in this thesis, which allows the description to also be applicable to other terms or different models.

1.2 Overview

The thesis has been written to be self-contained where possible and assumes the reader to have basic knowledge of general relativity and quantum field theory. Understanding of basic cosmology definitely helps, but is not required. The structure of the thesis is as follows. First, the CMB is discussed and then a model that can reproduce the isotropy and inflation is introduced. We then consider perturbations to this model which can be quantised to give a quantum field theory model for perturbations, which can reproduce the anisotropy of the CMB. The primary observables, the two-point and three-point functions, of the quantum field theory are defined and discussed. After that, we show how these can be calculated and calculate a specific term related to self-interactions. We then conclude the thesis and offer some perspective on extensions of this simplest model.

1.2.1 Chapter contents

We start with preliminaries which introduce some of the notation and conventions, followed by an introduction of the CMB, highlighting its isotropy and the small anisotropies and what we learn from it, all in chapter 1.

The Einstein equations for an isotropic universe are given, which are the FLRW equations. Inflation is introduced and the motivation for a period of inflation is given. The slow-roll parameters are introduced, which are important when expanding quantities order by order. After this, we introduce the arguably simplest model that can produce inflation, the canonical scalar field with slow-roll conditions. This is the contents of chapter 2.

In chapter 3, small perturbations to the isotropic universe (and inflation model) are considered. First, perturbations are formally introduced, where it is mentioned how in linear perturbation theory the scalar degrees of freedom

decouple from the tensor degrees of freedom. The perturbations are conserved outside the horizon, as proved by Weinberg, and we discuss how this allows relating the perturbations to the CMB anisotropies. Gauge freedom of the perturbations is discussed. This is the contents of section 3.1. The next section applies perturbation theory to the canonical scalar field, but does not use the formalism discussed in the previous section. Instead, starting from the scalar field action, the ADM formalism is used to rewrite the action in Hamiltonian formalism, which will allow quantising the action. The elegant proof by Maldacena of conservation outside the horizon of perturbations for the canonical scalar field is explained. Finally, an action for the perturbations is given. This concludes section 3.2. In the last section of the chapter, the correlation functions and their derived quantities are defined. The two-point function defines the power spectrum and the three-point function defines the bispectrum. It is shown how the three-point function can be taken as the definition of non-Gaussianity. Literature results are quoted for these quantities and their values from CMB measurements are given. This is the contents of section 3.3.

In chapter 4, the perturbations derived in the previous chapter are quantised to give a quantum field theory for perturbations. Formula's are given for calculating the power spectrum and bispectrum in the in-in formalism. Solutions to the classical equations of motion and choice of vacuum are discussed. The power spectrum is calculated, giving the result from literature. The contribution to the bispectrum of the $\epsilon\dot{\eta}$-term, a term in the interactions for perturbations, is calculated. The calculation is described in a general way that allows it to be applied directly to other terms and extended to other models of inflation. Following the calculation of the $\epsilon\dot{\eta}$-contribution, a $\log|K\tau|$ divergence is found and a boundary term in time is found in literature to compensate this divergence.

Chapter 5 concludes the thesis. The conclusion is that the $\epsilon\dot{\eta}$ term does not contribute to the non-Gaussianity. A very general extension to the canonical scalar field model, the effective field theory of inflation, is mentioned, along with some interesting results from this effective field theory formalism.

1.3   Preliminaries

We will be working in units of $c = \hbar = 1$ and with two related sets of coordinates, corresponding to the flat FLRW metric (introduced in section 2.1), with the time t and coordinate distances $\vec{x}$ defined via

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + a(t)^2\,\mathrm{d}\vec{x}^2,$$

and the conformal time τ defined via

$$\mathrm{d}s^2 = a(\tau)^2(-\,\mathrm{d}\tau^2 + \mathrm{d}\vec{x}^2).$$

Time derivatives are indicates by a dot, e.g. $\dot{a} := \partial_t a$. We will also use primes to indicate derivative with respect to τ when a function explicitly depends on time, e.g. $a' := \partial_\tau a$. We assume the reader is familiar with the Hubble observations that space is always expanding, such that $\dot{a} > 0$. Coordinate distance is related to physical distance via the scale factor a and conformal time is related to time via

$$\mathrm{d}\tau = \frac{\mathrm{d}t}{a}. \tag{1.1}$$

Time runs from $-\infty$ to ∞ as conformal time runs from $-\infty$ to 0, so $\tau < 0$. For de Sitter space the scale factor is given by

$$a = e^{Ht}, \tag{1.2}$$

so for both de Sitter space-time and inflation up to slow-roll corrections, the relation between time and the scale factor is

$$\tau = \frac{-1}{aH} \quad \Leftrightarrow \quad a = \frac{-1}{H\tau}. \tag{1.3}$$

We will stick to the convention that Greek indices, μ, ν, etc., are space-time indices that go from 0 to 3 and Latin indices, i, j, etc., are space indices that go from 1 to 3.

The Fourier convention used here is

$$\hat{f}(\vec{k}) = \int_{\mathbb{R}^d} f(\vec{x}) e^{-i\vec{k}\cdot\vec{x}} \, \mathrm{d}^d x, \tag{1.4}$$

$$f(\vec{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(\vec{k}) e^{i\vec{k}\cdot\vec{x}} \, \mathrm{d}^d k. \tag{1.5}$$

There exist multiple conventions for the Fourier transform, so we mention the convention used here explicitly. Mixing conventions often results in incorrect factors of 2π. The Fourier transform of a product becomes a convolution, which when written explicitly in this convention becomes

$$\mathcal{F}(f \cdot g)(k) = (\hat{f} \star \hat{g})(k) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(\vec{q}) \hat{g}(\vec{k} - \vec{q}) \, \mathrm{d}^d q. \tag{1.6}$$

The mnemonic for this convention is to divide every momentum integration measure $\mathrm{d}^d k$ by $(2\pi)^d$.

When working with slow-roll parameters, ϵ, η, etc., the order notation $\mathcal{O}(\epsilon^n)$ denotes all terms that are nth power in slow-roll parameters, e.g. $\mathcal{O}(\epsilon) = \mathcal{O}(\epsilon, \eta, \xi, \ldots)$.

1.4   Cosmic Microwave Background

The cosmic microwave background, abbreviated as the CMB, is thermal radiation coming to us from the early universe. It's the oldest light we can observe, created during the epoch of recombination which occurred $375\,000 \pm 1800$ years after the big bang[8]. Before this epoch, the light was in thermal equilibrium with the baryonic matter contents of the universe, forming an opaque plasma, with light being effectively trapped inside small regions. Because of this thermalisation process, the spectrum of this radiation became a blackbody spectrum; see fig. 1.1. A blackbody spectrum is characterised by its temperature, T, so this associates a temperature to the incoming photons of the CMB. When the CMB temperature measurement is repeated for each angle in the sky a temperature map is created; see fig. 1.2. The CMB has a temperature of $T = 2.725\,48 \pm 0.000\,57$ K [14]. It's one of the most precisely measured blackbody spectrums in nature, which gives confidence that what is measured is primordial (from the thermal history of the universe). During the epoch of recombination, as the universe cools due to its expansion, the electrons in the plasma become bound into hydrogen and helium atoms. This causes the universe to become transparent to the photons and they fall out of thermal equilibrium, after which the photons are allowed to travel freely. This point in time is referred to as last scattering and the origin of the CMB photons as observed from earth is called the surface of last scattering. The photons that were 'released' during last scattering cool due to the expansion of the universe, whilst preserving the blackbody distribution of the spectrum during this free streaming. The blackbody spectrum temperature drops down from around 3000 K [1] to its current value of 2.725 K.

What is striking about the CMB is its isotropy: the differences between the spectrum at different angles in the sky are of order $400\,\mu$K on a background of 2.725 K.[22] This means that different patches of the universe that were naïvely never in causal contact with one another somehow thermalised to the same temperature to striking precision. This is a strong motivator for a history of the universe that allows these patches to have been in causal contact with each other at some point in the past; this history is believed to be a period of inflation, as we'll explain and further motivate in section 2.2.

The fluctuations in temperature, $\Delta T/T$, between different points in the sky are called the anisotropy. They can be traced back to fluctuations present during inflation. These primordial fluctuations seeded the CMB anisotropy and the matter density contrast, $\delta = \delta\rho/\rho$, in large-scale structure observations.

1.4.1   Relating CMB to primordial fluctuations

To relate the fluctuations produced during inflation to the CMB anisotropies, the evolution of the fluctuations needs to be traced from the time the fluctuations were created to the time the fluctuations were observed. We will primarily focus on a particular type of fluctuations, namely scalar perturbations, which are the predominant source of the temperature anisotropies in the CMB. First we discuss the observation of the CMB anisotropy and then we show the relation between this and the primordial perturbations.

---

[1]Using[25] $T = T_0(1 + z)$ as a function of redshift z, with $T_0 = 2.725$ K and $z = 1090$ [23] gives the temperature of one particular time during recombination.
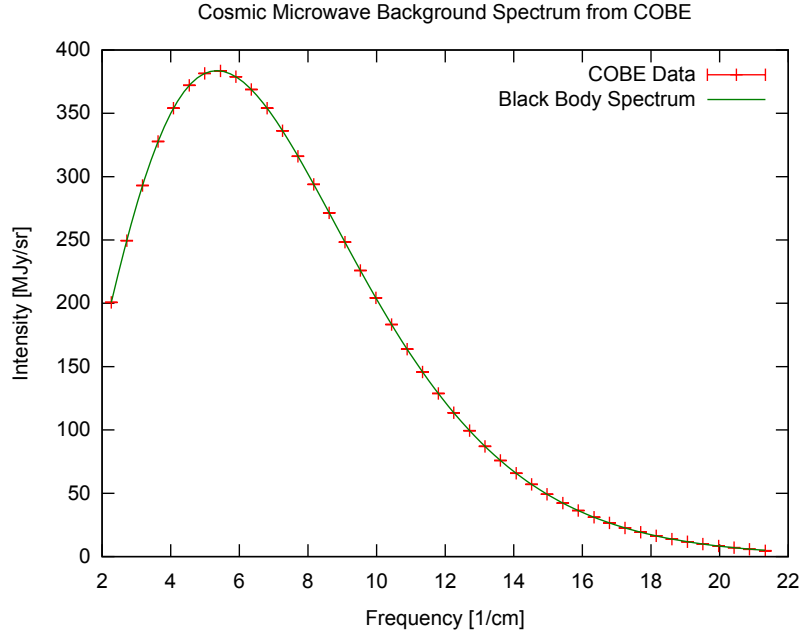
Figure 1.1: Blackbody spectrum of CMB photons.[26]

The observation of the CMB is performed by looking at every angle in the sky from a 'fixed' point; it is a probe of a sphere (of last scattering). The temperature as a function of angle over the sphere, $T(\varphi, \theta)$, can be decomposed using *spherical harmonics*, which can be thought of as the Fourier basis on a sphere. This automatically gives a form of averaging the measurement over the entire sky, just like Fourier for function on real space. The spherical harmonics are a set of functions on the spherical coordinates $(\varphi, \theta) \in [0, 2\pi) \times [0, \pi)$,

$$Y_l^m(\varphi, \theta), \quad l \geq 0, \ -l \leq m \leq l. \tag{1.7}$$

With the inner product for functions on the sphere given by

$$\langle f, g \rangle = \frac{1}{4\pi} \int_0^\pi \mathrm{d}\theta \int_0^{2\pi} \sin\theta \, \mathrm{d}\varphi \ f(\varphi, \theta)g(\varphi, \theta)^*, \tag{1.8}$$

the spherical harmonics form an orthonormal basis of functions on the sphere:

$$\left\langle Y_l^m, Y_{l'}^{m'} \right\rangle = \frac{1}{4\pi} \int_0^\pi \mathrm{d}\theta \int_0^{2\pi} \sin\theta \, \mathrm{d}\varphi \ Y_l^m(\varphi, \theta) Y_{l'}^{*m'}(\varphi, \theta) = \delta_{l\,l'}\delta_{mm'}, \tag{1.9}$$

$$\|Y_l^m\|^2 = \langle Y_l^m, Y_l^m \rangle = 1. \tag{1.10}$$

Furthermore, they are defined such that they obey the relation

$$Y_l^{*m} = Y_l^{-m}. \tag{1.11}$$

The temperature variations $\Delta T(\hat{n})$, with $\hat{n}$ the normal vector indicating the
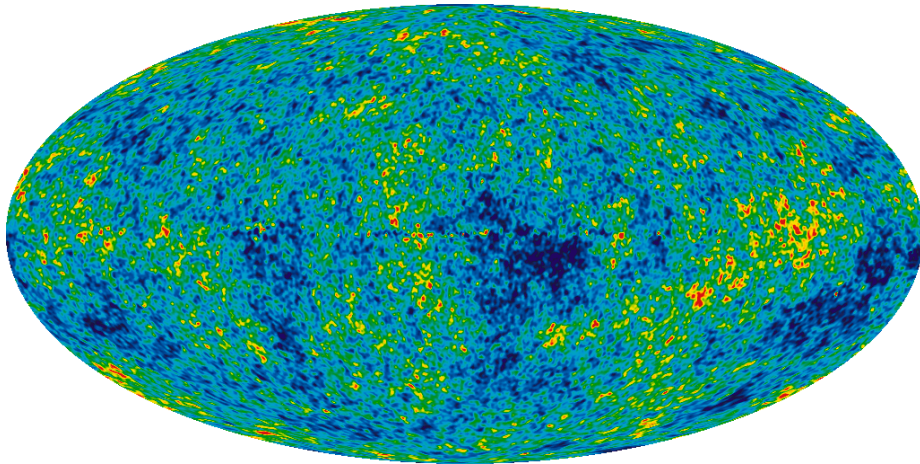
Figure 1.2: Map of the temperature fluctuations in the CMB sky. The average temperature is $2.725\,\mathrm{K}$, with fluctuations shown as colour differences with a range of $\pm 200\,\mu\mathrm{K}$.[22]

angle in the sky being short-hand for (φ, θ), are defined and decomposed as

$$\Delta T(\hat{n}) := T(\hat{n}) - T_0 = \sum_{lm} a_{lm} Y_l^m(\hat{n}), \tag{1.12}$$

$$T_0 := \frac{1}{4\pi} \int \mathrm{d}^2\hat{n}\, T(\hat{n}). \tag{1.13}$$

Here, a_{lm} are the spherical harmonics components of ΔT (like Fourier components); T_0 is defined by averaging the temperature over the sky. Reality of ΔT imposes that

$$a_{lm}^* = a_{l(-m)}. \tag{1.14}$$

Cosmic variance

To get information out of the CMB that allows determining the underlying statistics, an average over the position from which the CMB is measured should be taken; ideally, the CMB would be measured not just from earth, but from numerous places in the universe. For the same reason, ideally the measurement would be performed over an ensemble of 'universes', to be able to average over all possible quantum fluctuations. Because we can only measure the CMB from one particular spot and we can only see the events of one particular sequence of quantum fluctuations, what we end up measuring is only one instance of a random variable. We need to know how this one measurement relates to the statistics of what is being measured; in other words, we need to make an estimate of the difference between the proper statistical measurement we would like to perform and the 'single' measurement we actually can perform. We will now describe how these can be related, closely following Weinberg [25]. First a remark: the ergodic theorem says that these two kinds averages are the same, under reasonable assumptions about the separability of the measurements. For the rest of this piece on relating CMB observations to models, the average will be taken to mean both quantum and positional averaging.

Using isotropy, all averages $\langle \Delta T(\hat{n}_1) \Delta T(\hat{n}_2) \cdots \rangle$ are rotationally invariant; in particular, this means that $\langle \Delta T(\hat{n}) \rangle$ is independent of $\hat{n}$. Combining this with the fact that the sky-average of ΔT is 0 by definition, it follows that:

$$\frac{1}{4\pi} \int \mathrm{d}^2 \hat{n} \, \langle \Delta T(\hat{n}) \rangle = 0 \;\Rightarrow\; \langle \Delta T(\hat{n}) \rangle = 0. \tag{1.15}$$

The addition theorem for spherical harmonics states that, with P_l the Legendre polynomials,

$$P_l(\hat{n} \cdot \hat{n}') = \frac{4\pi}{2l+1} \sum_m Y_l^m(\hat{n}) Y_l^{*(-m)}(\hat{n}'). \tag{1.16}$$

Therefore, if the correlator between two temperature variations is to be rotationally invariant,

$$\langle \Delta T(\hat{n}) \Delta T(\hat{n}') \rangle = \sum_{lm} \sum_{l'm'} \langle a_{lm} a_{l'm'} \rangle Y_l^m(\hat{n}) Y_{l'}^{m'}(\hat{n}'), \tag{1.17}$$

it follows that

$$\langle a_{lm} a_{l'm'} \rangle = \delta_{ll'} \delta_{m(-m')} C_l, \tag{1.18}$$

with C_l an l-labeled set of numbers, after which the correlator can be written as

$$\langle \Delta T(\hat{n}) \Delta T(\hat{n}') \rangle = \sum_l C_l \left(\frac{2l+1}{4\pi} \right) P_l(\hat{n} \cdot \hat{n}'). \tag{1.19}$$

By inverting the Legendre transform, C_l can be expressed in terms of the correlator,

$$C_l = \frac{1}{4\pi} \int \mathrm{d}^2 \hat{n} \, \mathrm{d}^2 \hat{n}' \, P_l(\hat{n} \cdot \hat{n}') \, \langle \Delta T(\hat{n}) \Delta T(\hat{n}') \rangle. \tag{1.20}$$

What is actually observed is not the true average that gives C_l, but rather an average over m:

$$C_l^{\mathrm{obs}} = \frac{1}{2l+1} \sum_m a_{lm} a_{l(-m)}. \tag{1.21}$$

The difference between the theoretical C_l and the observed C_l is called *cosmic variance*. Fortunately, the difference between the observed C_l^{obs} and the underlying theoretical C_l for (near) Gaussian random variables decreases with l:

$$\left\langle \left(\frac{C_l - C_l^{\mathrm{obs}}}{C_l} \right)^2 \right\rangle = \frac{2}{2l+1}. \tag{1.22}$$

That means that cosmic variance is an issue with the low-l modes, but for higher l what is observed from the CMB can be taken as actually being the average over position and quantum fluctuations. Higher-l modes can give information on the statistics of the primordial universe. The measured multipole coëfficients, C_l, of the CMB are shown in figure 1.3.

Transfer function

The relation between the perturbations created during inflation, characterised by the power spectrum $P_{\mathcal{R}}$, which we will define in section 3.3, and the measured multipole coëfficients of the CMB, C_l, is given by:[13, 5]

$$C_l = \frac{2}{\pi} \int \mathrm{d}k \, k^2 P_{\mathcal{R}}(k) \mathcal{T}_l(k), \tag{1.23}$$
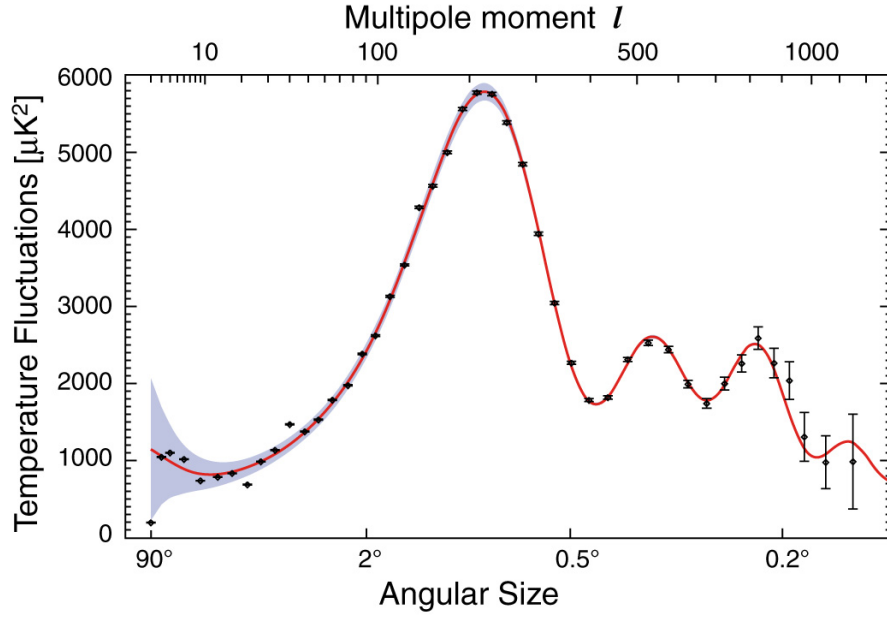
Figure 1.3: Measured multipole coëfficients (anisotropy) in the CMB; the quantity $l(l+1)C_l/2\pi$ is plotted against the multipole moment l.[21]

where $\mathcal{T}_l(k)$ is a known transfer function that accounts for evolving the perturbations from the moment they 're-enter the horizon' until the time they are measured. We will discuss what it means for a mode to 're-enter the horizon' in section 3.1. For now, the message is that scalar perturbations during inflation can be directly related to multipole coëfficients measured in the CMB. Looking again at the CMB anistropy measurements, fig. 1.3, the characteristic peaks arise from the transfer function; the scalar perturbations must have a nearly scale invariant power spectrum ($P_\mathcal{R}(k)$ is approximately constant). This approximate scale-dependence is one of the key predictions of models of inflation that is observed in the data.

Chapter 2

Isotropic universe and inflation

2.1   Friedmann-Lemaître-Robertson-Walker

To construct models of the (early) universe and perform calculations, the starting point is general relativity and the Einstein-Hilbert action:

$$S_{\text{EH}} = \frac{M_{\text{Pl}}^2}{2} \int \mathrm{d}^4 x \sqrt{-g}\, R, \tag{2.1}$$

where we have set $\hbar = c = 1$. Gravity is (minimally) coupled to other fields by adding a relativistic action for the Lagrangian, $\mathcal{L}_{\text{m}}$, describing the fields,

$$S_{\text{m}} = \int \mathrm{d}^4 x \sqrt{-g}\, \mathcal{L}_{\text{m}}, \tag{2.2}$$

to the total action

$$S = S_{\text{EH}} + S_{\text{m}}. \tag{2.3}$$

Variation of this action with respect to the metric, $g^{\mu\nu}$, gives the Einstein equations[1]

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \frac{1}{M_{\text{Pl}}^2} T_{\mu\nu}\,, \tag{2.4}$$

where $T_{\mu\nu}$ is the stress-energy tensor of S_{m}, given by

$$T_{\mu\nu} = \frac{-2}{\sqrt{-g}} \frac{\delta S_{\text{m}}}{\delta g^{\mu\nu}}. \tag{2.5}$$

From the CMB observation, we know that our universe is homogeneous and isotropic to a high degree.[2] We therefore look for isotropic solutions to the Einstein equations, by assuming an isotropic form of the metric and demanding that the matter content of the universe, which sources the metric, has the same symmetry.

The assumed metric is the Friedmann-Lemaître-Robertson-Walker metric, which for zero spatial curvature is

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + a(t)^2\, \mathrm{d}\vec{x}^2. \tag{2.6}$$

Here, $a(t)$ is the scale factor, which gives how much the universe has expanded. For instance, if $a(0) = 1$ and $a(t) = 2$, then in the time from 0 to t space has expanded by a factor of 2. This means that the distance between two points in space with (fixed) coordinates $\vec{x}_1$ and $\vec{x}_2$ has been increased by a factor of 2 and the volume of space has increased by a factor of 8 during this time.

The stress energy tensor, under the assumption of isotropy and zero curvature, takes on the form of a perfect fluid[25]

$$T^{00} = \rho(t), \quad T^{0i} = 0, \quad T^{ij} = a(t)^{-2} p(t) \delta^{ij}. \tag{2.7}$$

---

[1] To be completely accurate, the Gibbons-Hawking-York boundary term needs to be added to the Einstein-Hilbert action to reproduce the Einstein equations.

[2] Homogeneous corresponds to translational symmetry in all spatial directions, while isotropic corresponds to rotational symmetry in all angles and at each point. Isotropy is a superset of homogeneity; by saying the universe is isotropic we are automatically saying it is homogeneous. However, homogeneity is possible without isotropy, for instance an electric field equal in all of space aligned in one particular direction.

Taking both together gives the Friedmann equations for zero curvature

$$-\frac{2\dot{a}^2}{a^2} - \frac{\ddot{a}}{a} = -\frac{1}{2M_{\text{Pl}}^2}(\rho - p), \tag{2.8}$$

$$\frac{3\ddot{a}}{a} = -\frac{1}{2M_{\text{Pl}}^2}(3p + \rho). \tag{2.9}$$

These express the evolution of the scale factor in terms of the matter contents in the universe. The stress-energy tensor is a conserved quantity in general relativity, characterised by the conservation law

$$T^{\mu\nu}{}_{;\nu} := \nabla_\nu T^{\mu\nu} = 0. \tag{2.10}$$

Here, ∇ denotes the covariant derivative. For the perfect fluid form, these equations reduce to the conservation equation

$$\dot{\rho} + \frac{3\dot{a}}{a}\left(p + \rho\right) = 0. \tag{2.11}$$

These three equations together, the Friedmann equations plus the energy-momentum conservation, characterise the general solution.

2.2 Inflation

It is believed that at the beginning of the universe, before the period of baryonic matter creation, there was a period of approximately exponential expansion of the universe, dubbed *inflation*. During inflation, the scale factor a, so also physical distances, grew by a factor of at least 10^{26}. The driving 'force' behind inflation is thought to be one, or more, scalar fields that slowly roll down a potential, causing a to grow approximately exponentially. There is no direct evidence of inflation and the details are not known, but the idea of inflation has become very plausible: inflation solves three classical problems of cosmology, and theories of inflation make predictions for the properties of CMB anisotropies that have been observed.

2.2.1 Friedmann equations

Inflation is the approximately exponential expansion of the universe,

$$a(t) \sim e^{Ht}, \tag{2.12}$$

where H is approximately constant. The choice of calling the factor H is not arbitrary, because H is actually the Hubble parameter, as we shall see shortly. More generally, we can look for solutions of the FLRW equations (2.8, 2.9) with a positive scale factor a, so a can be written as

$$a(t) = a_0 e^{\int_{t_0}^t \mathrm{d}t H(t)}, \tag{2.13}$$

for some function $H(t)$. Another way to write the dependence of a on H is

$$\frac{\dot{a}}{a} = H, \tag{2.14}$$

and with this the Friedmann equations can be recombined to give

$$H^2 = \frac{1}{3M_{\mathrm{Pl}}^2}\rho, \tag{2.15}$$

$$2\dot{H} + 3H^2 = -\frac{1}{M_{\mathrm{Pl}}^2}p, \tag{2.16}$$

where we used that

$$\dot{H} = \frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2}. \tag{2.17}$$

If $\dot{a} > 0$ for all times, which we will henceforth assume, then a and t are two different ways of denoting (cosmological) time. A convenient quantity is the number of *e-folds* from a time t to some fixed (later) time t_0, defined as

$$\Delta\mathcal{N} = \mathcal{N}(t_0) - \mathcal{N}(t) = \log a(t_0)/a(t) = \int_t^{t_0} \mathrm{d}t' H(t'). \tag{2.18}$$

The number of e-folds (like tenfolds, but with base e) is the number of times the universe has expanded by a factor of e in the time from t to t_0. Note the relation between time and number of e-folds:

$$\mathrm{d}\mathcal{N} = -H\,\mathrm{d}t. \tag{2.19}$$

2.2.2 Cosmological constant

One thing that can be seen from the FLRW equations in Hubble form is that to have exactly exponential expansion, the matter contents of the universe needs to obey

$$p = -\rho. \tag{2.20}$$

This is the specific instance of $w = -1$ of a perfect fluid obeying the cosmological *equation of state*:

$$p = w\rho, \quad w \in \mathbb{R}. \tag{2.21}$$

The equation of state with $w = -1$ is associated to the cosmological constant, typically denoted by Λ. It is constant because ρ and p are constant, either by the Friedmann equations or directly by the conservation equation (2.11) which reduces to $\dot{\rho} = 0$. It is equivalent to adding a cosmological constant term to the action:

$$S_\Lambda = \int \mathrm{d}^4 x \sqrt{-g}\Lambda, \tag{2.22}$$

for which the stress-energy tensor is

$$T_\Lambda{}^{\mu\nu} = \Lambda g^{\mu\nu}, \tag{2.23}$$

from which, using eq. (2.7), we can read off that $\rho = -\Lambda$ and $p = \Lambda$. The conclusion is that a cosmological constant is the unique solution that gives exact exponential (and eternal) inflation. The manifold associated to exact exponential inflation is de Sitter space-time.

2.2.3 The Hubble parameter

To see that H is indeed the Hubble parameter, consider two points on the manifold, x_1 and x_2, in the flat coordinate system used (eq. (2.6)). The definition of the Hubble constant is via the relation between the observed drift velocity, v_D, and the physical distance, d, between the points:

$$v_\mathrm{D} = Hd. \tag{2.24}$$

This is true only for points that are nearby enough such that H can be taken to be constant. Because the points are stationary in the coordinates, the physical distance and drift velocity between the two points, using eq. (2.6), is

$$d^2 = g(x_2 - x_1, x_2 - x_1) = 0 + a^2 |x_2 - x_1|^2, \tag{2.25}$$

$$d = a |x_2 - x_1|, \tag{2.26}$$

$$v_\mathrm{D} = \frac{\mathrm{d}d}{\mathrm{d}t} = \dot{a} |x_2 - x_1|. \tag{2.27}$$

Therefore, from the Hubble relation it can be seen that H is indeed the Hubble parameter:

$$v_\mathrm{D} = \frac{\dot{a}}{a}d \quad \Rightarrow \quad H = \frac{\dot{a}}{a} \text{ is the Hubble parameter.} \tag{2.28}$$

2.2.4 Arguments for inflation

Inflation solves (at least) three 'problems' in hot big big bang cosmology, making it a plausible theory. Observations of the CMB and other data favour a model of the universe that is spatially flat, or nearly flat. If the universe is spatially flat now, then, following backwards the thermal evolution of the universe, at earlier times the universe must have been even more flat. This is not necessarily a problem; the universe could have started out as very flat, but a natural mechanism that made the universe become flat is preferred over demanding it to start out as such. If inflation lasted a minimum of 17-68 e-folds[25], then it will have flattened out the universe enough to be consistent with current data; the wide range of 17-68 comes about from minimal and maximal estimates of the energy density at the end of inflation.

The second problem has to do with horizons. The high degree of isotropy in the observed CMB cannot be explained by the thermal history of the universe, as calculations show that regions of space that we see separated by more than a couple of degrees in the sky were never in causal contact with each other and therefore could not have thermalised to the exact same background temperature. Initial inhomogeneities would have caused different regions of the sky to have a different temperature. Inflation solves this problem by allowing these different regions of the sky to start out as being in causal contact at some point during inflation; to solve the horizon problem, inflation needs to last a minimum, again, of 17-68 e-folds. We will revisit the horizon problem using a more qualitative picture when discussing perturbation modes, in section 3.1.2 and using figure 3.1.

The third problem is about topological defects, such as magnetic monopoles, which are thought to be created in high-energy gauge theories when a simple symmetry group is spontaneously broken to the gauge symmetry of the Standard Model. In all such cases, magnetic monopoles (magnetic charge) are created and it is believed that no continuous processes exist that can smooth these out to undetectable levels. One possible solution is that a period of inflation causes enough expansion to reduce the magnetic monopole ratio such that its presence would not be observable today. This 'problem' is still purely speculative, though.

Of these three, the horizon problem is the most convincing argument for inflation. Finally, as mentioned, there is also the idea that inflation is a natural candidate for creating the cosmological perturbations that we observe in the CMB data and in seeding the formation of large scale structure. The natural quantum mechanical perturbations of the fields during inflation become stretched out to cosmological scale, giving the observed inhomogeneities.

2.2.5 Slow-roll inflation

For inflation to solve the aforementioned problems it must last for at least $\mathcal{O}(50)$ e-folds. The proposed model for this is a scalar field, or multiple scalar fields, evolving under a potential, where the scalar fields slowly roll down the the potential. This causes a slowly decreasing Hubble parameter, thus giving a period of exponential expansion, and eventually ending inflation when the field reaches the minimal of the potential. We will discuss the scalar field model in section 2.3. It is now useful to have a characterisation of the 'slowness' of H, i.e. of the slowness of the slow-roll, in very general terms without referring to

any particular model. This characterisation can be made at the level of H and its derivatives, by the Hubble slow-roll parameters:

$$H = \frac{\dot{a}}{a}, \tag{2.29}$$

$$\epsilon = -\frac{\dot{H}}{H^2}, \tag{2.30}$$

$$\eta = \frac{\dot{\epsilon}}{\epsilon H}, \tag{2.31}$$

$$\xi = \frac{\dot{\eta}}{\eta H}. \tag{2.32}$$

Each slow-roll parameter is the fractional and dimensionless change of the previous. The change is expressed via a derivative in time, but this has dimension of $[s^{-1}]$, so by dividing by the time scale of inflation, H $[s^{-1}]$, it is made dimensionless. For instance, ϵ gives the average relative change in H during one *Hubble time*; a Hubble time is $\Delta t = 1/H$. If ϵ is small, then this means that H does not vary much during one Hubble time.

Slow-roll inflation is the name given to (the period of) inflation that not only has a slowly varying Hubble parameter, in the sense that $\epsilon \ll 1$, but where also ϵ is slowly varying, in the sense that $|\eta| \ll 1$. Typically, all slow-roll parameters are taken to be much smaller than 1 during the entirety of slow-roll. However, it is admissible for the parameters after η, starting with ξ, to take on sizable values for a short period, such as when it is oscillating.

Energy conditions: $\dot{H} < 0$

From the FLRW eqs. (2.15) and (2.16) it follows that

$$\dot{H} = -\frac{1}{2M_{\text{Pl}}^2}(p + \rho), \tag{2.33}$$

$$p + \rho > 0 \quad \Rightarrow \quad \dot{H} < 0, \tag{2.34}$$

where we have already seen the edge case of $\rho + p = 0 \Leftrightarrow w = -1 \Leftrightarrow \dot{H} = 0$ for a cosmological constant. All energy conditions give $\rho + p \geq 0$, so we will assume for the rest that $\dot{H} \leq 0$, so H is decreasing. This explains the 'odd' choice of sign for the definition of ϵ, which is now seen to fix ϵ to be positive. The other parameters are not necessarily bounded to be positive or negative. From the requirement of accelerated expansion, $\ddot{a} > 0$, it follows from

$$\dot{H} = \frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2},$$

that

$$\epsilon = -\frac{\dot{H}}{H^2} = 1 - \frac{\ddot{a}}{aH^2}, \tag{2.35}$$

so accelerated expansions occurs precisely while $\epsilon < 1$; accelerated expansion is a minimal requirement for inflation.

2.3 Canonical Scalar Field

The single canonical scalar field offers a simple model for inflation. The action for the canonical scalar field model, minimally coupled to gravity and with an arbitrary potential, V, is

$$S = \int \mathrm{d}^4 x \sqrt{-g} \left[\frac{M_{\mathrm{Pl}}^2}{2} R - \frac{1}{2} (\nabla \phi)^2 - V(\phi) \right]. \tag{2.36}$$

Here, again, ∇ denotes the covariant derivative, which for a scalar coincides with the regular coordinate derivative.

    The next step is to examine the space-time dynamics generated by this scalar field, by solving the Friedmann equations for ϕ. The stress-energy tensor, defined in eq. (2.5), for ϕ is given by

$$T_{\mu\nu} = \left[-\frac{1}{2} g^{\alpha\beta} \partial_\alpha \phi \partial_\beta \phi - V(\phi) \right] g_{\mu\nu} + \partial_\mu \phi \partial_\nu \phi. \tag{2.37}$$

We look for solutions that give an isotropic space-time, which requires $T_{\mu\nu}$ to take the perfect-fluid form, therefore $\phi(x,t)$ must be a function of only time $\phi(t)$, and the metric becomes the FLRW (flat) metric, eq. (2.6), such that

$$T_{\mu\mu} = \left[\frac{1}{2} \dot{\phi}^2 - V(\phi) \right] g_{\mu\mu} + \delta_{\mu 0} \dot{\phi}^2,$$
$$T_{\mu\nu} = 0, \text{ when } \mu \neq \nu. \tag{2.38}$$

This can be written to match the perfect fluid form, eq. (2.7),

$$T^{00} = \frac{1}{2} \dot{\phi}^2 + V(\phi), \quad T^{0i} = 0, \quad T^{ij} = a(t)^{-2} \left(\frac{1}{2} \dot{\phi}^2 - V(\phi) \right) \delta^{ij}, \tag{2.39}$$

from which we can read off that

$$\rho = \frac{1}{2} \dot{\phi}^2 + V(\phi), \quad p = \frac{1}{2} \dot{\phi}^2 - V(\phi). \tag{2.40}$$

The FLRW equations, eqs. (2.15), (2.15), and the conservation equation, eq. (2.11), become:

$$\dot{H} = -\frac{1}{2M_{\mathrm{Pl}}^2} \dot{\phi}^2, \tag{2.41}$$

$$\dot{H} + 3H^2 = \frac{1}{2M_{\mathrm{Pl}}^2} V(\phi), \tag{2.42}$$

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0. \tag{2.43}$$

Unfortunately, it's not possible to give a general solution to these equations, of which only two are independent, for a general potential V. Eq. (2.43) can be written completely in terms of ϕ and V, after which the solution for ϕ can be found for a given potential and from this H can be determined. The usual approach is to either consider specific potentials to solve for, or to say that the solution to these equations can be described well enough by a few slow-roll parameters and then calculate using the presumed solutions for H, ϵ, η, etc. In this thesis the second route is taken, where the slow-roll parameters are kept as unknown functions that can in principle be determined, and practically can be determined for specific potentials. Results will be expressed in terms of these unknown functions, with the slow-roll assumption giving bounds on their size and variation.

2.3.1 Slow-roll inflation

The slow-roll conditions should be imposed on the model such that it causes inflation. In imposing these restrictions, the connection between the slow-roll conditions and the dynamics can be understood. Slow-roll inflation requires that H changes little in one Hubble time and this is directly expressed via ϵ. For this model, this translates to the velocity $\dot{\phi}/H$ being small compared to the Planck mass, from eq. (2.41):

$$\epsilon = \frac{\dot{\phi}^2}{2M_{\mathrm{Pl}}^2 H^2} \ll 1. \tag{2.44}$$

The inflaton field should also roll down the potential in a gradual manner, such that the acceleration is small compared to the velocity. This is not a requirement of inflation per se, but having sharp features in the potential will introduce time-dependence into the generated quantum fluctuations. This would be observable in the CMB as a scale dependence, as we shall see later when we discuss what happens to the quantum fluctuations (perturbations) after inflation and when we look at the scale-dependence (running) of the perturbations, section 3.3. The observed scale dependence is very small, so relevant models of inflation should not generate strong scale dependence. In eq. (2.43) the size of the acceleration, $\ddot{\phi}/H^2$, should be constrained relative to the velocity $\dot{\phi}/H$; taking the derivative of ϵ and then using the definition for η gives

$$|\eta| \ll 1 \quad \Rightarrow \quad \frac{\ddot{\phi}}{H\dot{\phi}} = \frac{\eta}{2} - \epsilon \ll 1. \tag{2.45}$$

Therefore, to have a feasible model of inflation via a single scalar field we must demand that $\epsilon, |\eta| \ll 1$ are both slow-roll constrained. We shall see in section 3.3 that for this model a small scale-dependence also exactly requires $|\eta| \ll 1$. The slope of the potential now is the only thing that determines the velocity of ϕ:

$$3H\dot{\phi} + V'(\phi) \approx 0. \tag{2.46}$$

From $\epsilon \ll 1$ the velocity of ϕ is small and from $|\eta| \ll 1$ it follows that ϕ does not have a velocity beyond that directly induced by the potential slope, so ϕ is slowly rolling.

Potential slow-roll parameters

The slow-roll conditions were formulated by restricting ϕ, but the potential $V(\phi)$ determines the evolution of ϕ, so actually the slow-roll conditions are restrictions on the shape of the potential. For this reason, in literature, often different slow-roll parameters, ϵ_V and η_V, are used, which are the potential slow-roll parameters defined as

$$\epsilon_V = \frac{M_{\mathrm{Pl}}^2}{2} \left(\frac{V'}{V}\right)^2, \tag{2.47}$$

$$\eta_V = M_{\mathrm{Pl}}^2 \frac{V''}{V}, \tag{2.48}$$

and related to the Hubble slow-roll parameters by[3]

$$\epsilon_V = \epsilon \left(\frac{3 + \eta/2 - \epsilon}{3 - \epsilon} \right)^2 = \epsilon + \mathcal{O}(\epsilon^2), \tag{2.49}$$

$$\eta_V = \frac{6\epsilon - 3\eta/2 - 2\epsilon^2 + 5\epsilon\eta/2 - (\eta/2)^2 - \eta/2 \cdot \xi}{3 - \epsilon} = 2\epsilon - \frac{\eta}{2} + \mathcal{O}(\epsilon^2). \tag{2.50}$$

We prefer to use the Hubble slow-roll parameters, as these are more general and well-defined for any model. The Hubble slow-roll parameters ϵ and η express the core of slow-roll inflation in a model-independent way.

Model of inflation

The scalar field can offer a period of isotropic inflation, for a set of potentials that obey the slow-roll conditions, and inflation ends when the scalar field reaches the minimum of the potential. What has not been shown is that it is possible to reach a given number of e-folds of inflation with this model, but in practice it turns out that demanding η to be slow-roll will ensure a long enough period of inflation. The model is relatively simple to understand and it gives a good basis on which to build a quantum perturbation model for scalar inflation. This will be the subject of the next chapters.

---

[3]For the relation between ϵ_V and ϵ, see [15]. An interesting observation is that it seems that $\eta/2$ is the 'natural' quantity appearing in most calculations.

Chapter 3

Perturbations

3.1 Perturbations

In the previous chapter we described an isotropic model of inflation. The next step is to consider the inhomogeneities, because everything interesting about the universe, such as large scale structures, stars, the solar system, us, is certainly not homogeneous. We extend the model to include small perturbations that violate isotropy and homogeneity, which can become the source of the cosmological inhomogeneities observed today. Isotropy will become statistical isotropy and the perturbation can be quantised to become quantum perturbations, which then become a natural generator for the inhomogeneities. It is, of course, possible to start inflation with an initial condition that already contains the necessary inhomogeneities (at tiny scales). These tiny inhomogeneities are then stretched to cosmological scales during inflation, but then the question becomes "what caused these initial inhomogeneities?". This is what would happen if only classical perturbations were considered, because classical evolution does not allow for the creation of perturbations. On the other hand, a quantum model would mean that perturbations are naturally generated, independent of whatever the initial conditions are. This does not free us completely from initial conditions, because even with a quantum model the initial state has to be chosen, but this does give more freedom in this choice. Without quantum fluctuations during inflation, all anisotropies observed in the CMB need to be exactly included into the initial state. With a quantum model the 'correct' fluctuations are generated during inflation. With both the classical and quantum model, inhomogeneities in the initial state are smoothed out, which gives freedom in choosing the inhomogeneities in the initial state at scales that evolve to be outside the Hubble horizon today. Indeed, quantum fluctuations during inflation have a really good chance of being the source of the observed inhomogeneities today; being created during inflation, which is a slow-roll deviation from de Sitter space-time, they naturally acquire scale-invariance up to slow-roll corrections. This is the prediction of inflation models for the properties of the CMB anisotropies mentioned earlier.

3.1.1 Linear perturbation theory

The first step is to understand perturbations at the classical level. This will be the subject of this section and the next. After the classical picture has been formulated, the perturbations can be quantised and the effects on the CMB can be calculated. The inhomogeneities are very small, so they can be modelled as (linear) perturbations to the isotropic background. The metric $g_{\mu\nu}$ can be perturbed by a small quantity, $h_{\mu\nu}$,

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}, \tag{3.1}$$

where $\bar{g}_{\mu\nu}$ is the isotropic FLRW metric. Similarly, the stress-energy tensor becomes

$$T_{\mu\nu} = \bar{T}_{\mu\nu} + \delta T_{\mu\nu}, \tag{3.2}$$

The particular choice of the background metric/stress-energy tensor breaks explicit time-translation symmetry and Lorentz boosts, but because it is isotropic all spatial symmetries (translations and rotations) are still explicitly intact for the background. These symmetries allow the metric and stress-energy tensor

perturbations to be decomposed into scalars, vectors and tensor with respect to spatial symmetry, which is the so-called *scalar-vector-tensor decomposition.* This can be seen as an extension of the Helmholtz decomposition for a vector into a rotation-free and a divergence-free part:

$$V_i = \partial_i \Phi + B_i, \tag{3.3}$$

where B_i is divergence-free, i.e. $\vec{\partial} \cdot \vec{B} = \partial^i B_i = 0$ which also means it can be written as the rotation of a vector, $\vec{B} = \vec{\partial} \times \vec{A}$, and $\partial_i \Phi$ is rotation-free, by $\vec{\partial} \times \vec{\partial} f = 0$ for any function $f(\vec{x})$. Divergence-free is also referred to as *transverse*, because in Fourier space the vector is seen to be perpendicular (transverse) to the wavevector. Intuitively, h_{00} should transform as a scalar, h_{0i} should transform as a vector, and h_{ij} should transform as a tensor, so these degrees of freedom don't mix under the spatial symmetries. The vector can be Helmholtz decomposed and the tensor can be split into a trace plus a traceless part, where the traceless part has an analogous 'Helmholtz'-like decomposition. The result of this for the perturbed metric, following Weinberg [25], is

$$h_{00} = -E, \tag{3.4}$$
$$h_{0i} = a\left(\partial_i F + G_i\right), \tag{3.5}$$
$$h_{ij} = a^2 \left(A\delta_{ij} + \partial_i\partial_j B + \partial_i C_j + \partial_j C_i + D_{ij}\right), \tag{3.6}$$

where D_{ij} is symmetric, traceless and 'transverse' and the vectors, C_i and G_i, are transverse:

$$\partial^i C_i = 0, \quad D_{ij} = D_{ji}, \quad D_i{}^i = 0, \quad \partial^i D_{ij} = 0, \quad \partial^i G_i = 0. \tag{3.7}$$

The 10 degrees of freedom in the metric (4×4 symmetric matrix) are decomposed into four scalars, two transverse vectors, and one symmetric, traceless, 'transverse' 3D-2-tensor, which adds up to $4 \times 1 + 2 \times 2 + 1 \times (6 - 1 - 3) = 10$ degrees of freedom. To find a similar decomposition for the perturbed stress-energy tensor is bit more involved, but gives[25]

$$\delta T_{00} = -\bar{\rho}h_{00} + \delta\rho, \tag{3.8}$$
$$\delta T_{0i} = \bar{p}h_{i0} - (\bar{\rho} + \bar{p})\left(\partial_i \delta u + \delta u_i^V\right), \tag{3.9}$$
$$\delta T_{ij} = \bar{p}h_{ij} + a^2 \left(\delta p\delta_{ij} + \partial_i\partial_j \pi^S + \partial_i \pi_j^V + \partial_j \pi_i^V + \pi_{ij}^T\right), \tag{3.10}$$

with the equivalent properties,

$$\partial^i \pi_i^V = 0, \quad \pi_{ij}^T = \pi_{ji}^T, \quad \pi_i^{Ti} = 0, \quad \partial^i \pi_{ij}^T = 0, \quad \partial^i \delta u_i^V = 0. \tag{3.11}$$

These formula's can be taken as the definition of the quantities $\delta\rho$, δp and $\delta u_i := \partial_i \delta u + \delta u_i^V$, which represent the perturbation to the perfect fluid form quantities. The quantity u_i (and its perturbation δu_i) we haven't mentioned yet, but it is the velocity vector of the perfect fluid,

$$T_{\mu\nu} = pg_{\mu\nu} + (\rho + p)u_\mu u_\nu. \tag{3.12}$$

So far, we have worked in the rest frame of the perfect fluid (by assuming isotropy), $u_\mu = \mathrm{diag}\{-1, 0, 0, 0\}$. This decomposition can also be taken as the definition of the anisotropic inertia quantities π^S, π^V, and π^T, which characterise the departure from the perfect fluid form. The expression for $\delta T_{\mu\nu}$ is slightly

less clean than that of $h_{\mu\nu}$, as it mixes stress-energy tensor and metric quantities in the same line. The advantage of this split is that the Einstein equations and conservation equation decouple into three sets of equations, one for the scalars, one for the vectors and one for the tensors. In linear perturbation theory, there is no coupling between the scalar mode and the tensor mode, the scalar mode and the vector mode, and the vector mode and the tensor mode, which simplifies the calculations.

3.1.2   Modes

We used the word 'mode' here, without defining what that is. When the equations for the perturbed scalars, vectors, or tensors are Fourier transformed, then to first order in perturbations different wave numbers do not mix. Each wave number corresponds to a Fourier transformed degree of freedom, a *mode*, which can be solved independently from the other Fourier components. We are most interested in the scalar modes in this thesis, because these are currently observable in the CMB. Tensor modes might become observable in the near future, by measurement of the polarisation of light from the CMB; tensor modes are associated to propagating gravitational waves. Finally, vector modes tend to decay as $1/a^2$, so we will not be able to see them in the CMB. The mode evolves in the period between the creation of a perturbation during inflation and the time that it is measured in the CMB. The relevant physics of the universe are not precisely known for the entire history of the universe. In particular, the period just after inflation, such as the era of reheating and cold dark matter decoupling, is not fully understood. So then, how can measurements of anisotropies in the CMB be related to perturbations created during inflation?

Conservation outside the horizon

Weinberg [25] showed that there are always two (independent) scalar quantity solutions and one tensor quantity solution to the perturbation equations that are conserved outside of the horizon. Outside the horizon means that the mode of the quantity with associated wave number k has $k \ll aH$. Conserved means that, no matter the constituents of the universe, the time evolution of the mode goes to zero as $k \ll aH$. The wave number k is the comoving wave number and k/a is the physical wave number, so conservation is related to the domain where the physical wavelength is larger than the Hubble horizon, $R = 1/H$. The inverse comoving Huble horizon, aH, evolves as

$$\frac{\mathrm{d}(aH)}{\mathrm{d}t} = \frac{a}{H^2}\left(1 - \epsilon\right), \tag{3.13}$$

so it grows exponentially in time during inflation ($0 \le \epsilon \ll 1$). More generally, it is a growing quantity during accelerated expansion and a decreasing quantity otherwise. If a mode with a given wave number k started inside the horizon, $k \gg aH$, then if enough time of inflation passes then this mode will be outside of the horizon $k \ll aH$. The moment where the mode k crosses the horizon is called *horizon crossing* and happens when $k \approx aH$. When a mode crosses the horizon and becomes conserved, we say that the mode *freezes out*. The terms seems appropriate, because inflationary modes freeze out at order $\mathcal{O}(k/aH)$, which becomes small exponentially fast after horizon crossing. Within only a few

Hubble times as the mode crosses the horizon, its time evolution vanishes, so it truly 'freezes out'. These conserved scalar modes are also called *adiabatic modes*, because they affect/perturb all constituents of the universe equally ($\delta\rho/\dot{\bar{\rho}}$ is the same for every species). Another term to denote the scalar modes is *comoving curvature perturbations*, because, as we shall see when we discuss gauges, in the comoving gauge the conserved scalar quantity appears directly in the metric as a perturbation to a.

Relating perturbations to the CMB

Now to answer the question of relating measured perturbations to primordial perturbations and CMB anisotropies: the modes that freeze out during inflation, that are outside the horizon during the period between inflation and recombination, and that have fully reentered the horizon before now are the ones we can (potentially) observe in the CMB. Figure 3.1 shows how a mode is inside/outside the horizon during the evolution of the universe. All modes that have a chance of getting to us conserved need to at least freeze out during inflation. They need to be unaffected by the evolution of the universe during thermalisation and when we don't know the physics, so they need to be outside the horizon until after recombination. They need to observable, so this means that they must have fully reentered the horizon by the time they reach us; wave lengths that are of the order of Hubble or larger cannot be measured in practice. Finally, they must have free-streamed from the surface of last scattering to now. This gives a range of observable wave lengths for primordial perturbations that reach us from the surface of last scattering that contain direct information from the period of inflation; they can rightfully be called 'primordial'. Despite the CMB having thermalised to a near-perfect blackbody spectrum, these perturbations were conserved and are observable as anisotropies in the CMB. This conservation is critical for relating CMB observations to inflation, but of course conservation is not required for the perturbations to be the seeds of inhomogeneities in our universe.

As a final note, figure 3.1 can also be used to explain qualitatively how inflation solves the horizon problem, when taking λ_{phys} to be the distance between patches of the universe. If the patches start out as further away than the Hubble horizon, then the λ_{phys}-line starts out as being above the horizon during matter domination. If inflation lasts long enough, that is if it extends enough to the left, then λ_{phys} will intersect with the horizon during inflation and the scale between the patches will have been inside the horizon during inflation.

3.1.3 Gauge freedom

There are more degrees of freedom in the background plus perturbation equations than there are physical degrees of freedom. The coordinate system can be freely chosen and doing so does not change anything about the physics (covariance). However, while rotations and translations preserve the background, general coordinate transformations do not; for instance, moving in an isotropic fluid makes it non-isotropic in the new rest frame. Rather than considering general coordinate changes as coordinate changes, they can be considered (gauge) transformations of the fields. For instance, a scalar field $\phi(x)$ under a change of
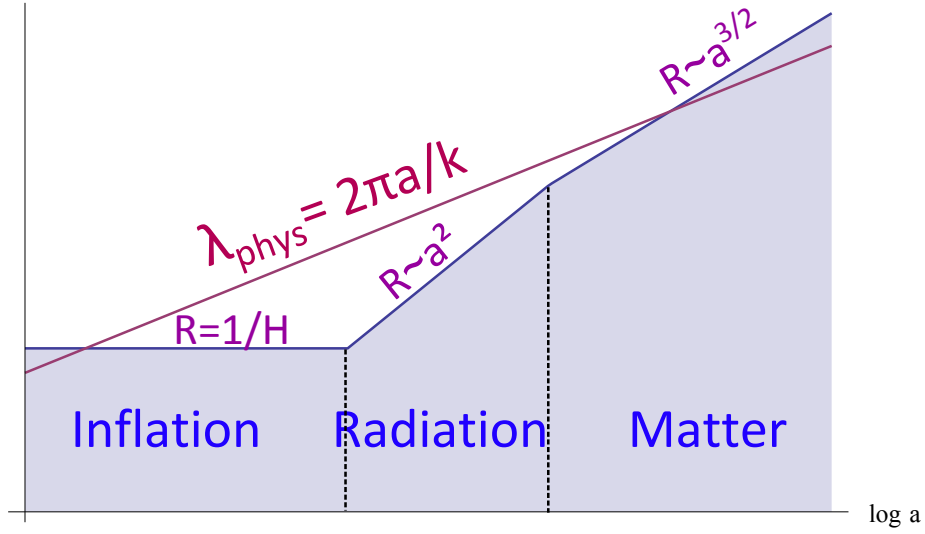
Figure 3.1: The state of a mode (inside/outside horizon) during the evolution of the universe. The physical wavelength evolves as a. During inflation, the Hubble radius is constant and it grows during matter and radiation domination.

coordinates $x \to x'$ is

$$\phi(x) \to \phi'(x') = \phi(x(x')), \tag{3.14}$$

which can be written, using the infinitesimal form of the coordinate change, $x \to x' = x + \epsilon(x)$, as a transformation of the field while leaving the coordinates untouched,

$$\phi(x) \to \phi(x) - \epsilon(x)\partial_x\phi(x). \tag{3.15}$$

The equivalence can be seen by filling in the infinitesimal transformation in eq. (3.14):

$$\phi'(x') = \phi(x' - \epsilon(x')) = \phi(x') - \epsilon(x')\partial_x\phi(x'). \tag{3.16}$$

Physicists often prefer to consider coordinate changes as field transformation because it unifies transformations of the coordinates with other potential transformations of the fields. For instance, consider the infinitesimal field transformation,

$$\phi(x) \to \phi(x) - \epsilon\partial_x^2\phi(x).$$

This transformation of the field does not correspond to any change of coordinates, because eq. (3.15) gives how any scalar must transform, but using the transformation of fields, both coordinate changes and general field transformations can be written in the same way. The isotropy of the background when considering the coordinate freedom can be preserved by using the field transformations picture and attributing all change in the total metric (and the total stress-energy tensor) to the perturbations. By doing so, the background is kept constant and the perturbations become gauge fields that transform in a particular way, representing coordinate changes. The gauge freedom explicitly parametrises the unphysical degrees of freedom in the choice of coordinates.

A second freedom comes from the full stress-energy tensor sourcing the full metric, but the division between background and perturbations not being fixed.

There is room to shift the perturbations by a small time-dependent constant which can be absorbed into the background; this changes the exact division between background and perturbations but leaves the physics unaltered. In practice, this freedom can be ignored as the choice will either arise naturally or the split can be kept formal without requiring specification.

Gauge freedom can obscure physical solutions and complicate quantisation. It is often desirable to choose a suitable gauge to fix the gauge freedom and we will do this for the scalar field model.

3.2 Scalar field perturbations

The previous chapter formally defined perturbations. For this chapter, we take a step back and start from the canonical scalar field, eq. (2.36), and derive an action for the perturbations; we will follow Maldacena [20] and Chen [9]. The scale factor $a(t)$ is a monotonically increasing function, so t and a can be used interchangeably to denote time ($\mathcal{N}$, the number of e-folds, can also be used). Choosing a background metric amounts to choosing $a(t)$, which is equivalent to choosing t, which in turn amounts to choosing a preferred time-slicing of the manifold. Choosing a time-slicing is giving the hyper-surfaces of constant time. By doing so, explicit Poincaré symmetry is broken to explicit spatial symmetry and implicit time symmetry. The reason for doing this explicit time slicing is because for quantum field theory it is desirable to have a Hamiltonian formulation of the field theory to facilitate quantisation; a Hamiltonian formalism requires breaking explicitly time symmetry by choosing a preferred time coordinate. For the scalar field, it is convenient to use the ADM formalism, developed by Arnowitt, Deser, and Misner, to effect a Hamiltonian description of the system from the current Lagrangian description.

3.2.1 ADM formalism

Consider again the action for a scalar field, eq. (2.36), but now without assuming isotropy, such that $\phi(x, t)$ is a function of both space and time,

$$ S = \int \mathrm{d}^4 x \sqrt{-g} \left[\frac{M_{\mathrm{Pl}}^2}{2} R - \frac{1}{2} (\nabla \phi)^2 - V(\phi) \right]. $$

Using the ADM metric,

$$ \mathrm{d}s^2 = -N^2 \, \mathrm{d}t^2 + h_{ij} \left(\mathrm{d}x^i + N^i \, \mathrm{d}t \right) \left(\mathrm{d}x^j + N^j \, \mathrm{d}t \right), \tag{3.17} $$

where N is the *lapse function*, $N > 0$, giving the elapsed proper time between two hyper-surfaces as $\mathrm{d}s = -N \, \mathrm{d}t$, and N^i is the *shift vector*, which gives the spatial direction that a fixed point is shifted to when going from one hyper-surface to the next under coordinate-time evolution. Together, the lapse and the shift describe how the different hyper-surfaces are 'glued' together. The determinant becomes

$$ \sqrt{-g} = \sqrt{h} N, \tag{3.18} $$

and the four-dimensional Ricci scalar R is related to the three-dimension Ricci scalar via the Gauss-Codazzi equation,[1]

$$ R = {}^{(3)}R + \frac{1}{N^2} \left(E_{ij} E^{ij} - E^2 \right), \tag{3.19} $$

where

$$ E_{ij} = \frac{1}{2} \left(\dot{h}_{ij} - \nabla_i N_j - \nabla_j N_i \right), \tag{3.20} $$

$$ E = E_i{}^i. \tag{3.21} $$

---

[1]What is shown is, in fact, not the full Gauss-Codazzi equation, but the Gauss-Codazzi equation plus the Gibbons-Hawking-York boundary term; taken together, terms cancel to give this result for R.

The action in the ADM formalism becomes[20]

$$S = \int \mathrm{d}^3x \, \mathrm{d}t \sqrt{h} N \Big[\frac{M_{\mathrm{Pl}}^2}{2} {}^{(3)}R + \frac{M_{\mathrm{Pl}}^2}{2N^2} \left(E_{ij} E^{ij} - E^2 \right)$$
$$+ \frac{1}{2N^2} \left(\dot\phi - N^i \partial_i \phi \right)^2 - h^{ij} \partial_i \phi \partial_j \phi - V(\phi) \Big]. \tag{3.22}$$

The tensor E_{ij} is related to the extrinsic curvature K_{ij} of the slices:

$$K_{ij} = \frac{1}{N} E_{ij}. \tag{3.23}$$

In this new form of the action, h_{ij} and ϕ are the dynamical fields, while N and N^i do not have dynamical terms (no time derivatives) and thus play the role of Lagrange multipliers: their equations of motion are constraint equations.

Comoving gauge

A convenient gauge for the scalar field is the comoving gauge, which has the remarkable feature that all perturbations become absorbed into the metric and there are no perturbations to the scalar field in this gauge:

$$\phi(x,t) = \phi(t), \quad \delta\phi = 0, \quad h_{ij} = a^2 e^{2\mathcal{R}} \hat{h}_{ij},$$
$$\det \hat{h} = 1, \quad \hat{h}_{ij} = \delta_{ij} + \gamma_{ij} + \frac{1}{2} \gamma_{il} \gamma_{lj} + \dots, \tag{3.24}$$

where the gauge choice has been shown up to second order; $\mathcal{R}$ and γ_{ij} are first order perturbation quantities. The *comoving curvature perturbations* $\mathcal{R}$ are seen to be direct perturbations to the scale factor. With this gauge the action and constraint equations are

$$S = \int \mathrm{d}^3x \, \mathrm{d}t \, a^3 e^{3\mathcal{R}} N \Big[\frac{M_{\mathrm{Pl}}^2}{2} {}^{(3)}R - V(\phi) + \frac{M_{\mathrm{Pl}}^2}{2N^2} \left(E_{ij} E^{ij} - E^2 \right) + \frac{1}{2N^2} \dot\phi^2 \Big], \tag{3.25}$$

$$\frac{1}{2} M_{\mathrm{Pl}}^2 {}^{(3)}R - V(\phi) - \frac{M_{\mathrm{Pl}}^2}{2N^2} \left(E_{ij} E^{ij} - E^2 \right) - \frac{1}{2N^2} \dot\phi^2 = 0, \tag{3.26}$$

$$\nabla_i \left[\frac{1}{N} E^i{}_j - \frac{1}{N} E \delta^i{}_j \right] = 0. \tag{3.27}$$

We also used that $\det h = \left(a^2 e^{2\mathcal{R}} \right)^3$. To get the action for the perturbations, first the constraints need to be solved (possibly in a perturbative manner, order by order) and then the solution to the constraints can be plugged back into the action to give an action for perturbations.

Zeroth order constraint

Looking at the constraint equations order by order, at zeroth order in perturbations the constraints give the original isotropic background equations. This can be seen by considering all fields at zeroth order and solving the resulting equations. By comparing the ADM metric with the FLRW metric it should be solved by $N = 1$, $N^i = 0$, so we assume that N^i starts at first order. The Ricci-3-scalar contains first and second-order spatial derivatives of h_{ij}, but not terms

without any spatial derivatives, and $\partial_k h_{ij}$ starts at first order in perturbations, so the entire $^{(3)}R$ starts at first order. The second term, containing E's, can be written at first order by

$$h_{ij} = a^2 \delta_{ij}, \quad h^{ij} = a^{-2} \delta^{ij},$$
$$\dot{h}_{ij} = 2a^2 H \delta_{ij} \quad \Rightarrow \quad E_{ij} = a^2 H \delta_{ij},$$
$$\dot{h}^{ij} = -2a^{-2} H \delta^{ij} \quad \Rightarrow \quad E^{ij} = -a^{-2} H \delta^{ij},$$
$$E = h_{ij} E^{ij} = -3H.$$

This gives the zeroth order constraint from N:

$$3H^2 = \frac{1}{M_{\text{Pl}}^2} \left(N^2 V(\phi) + \frac{1}{2} \dot{\phi}^2 \right), \tag{3.28}$$

which is equivalent to eq. (2.41)-(2.42) for $N = 1$. The freedom to choose N different from 1 by small amounts and thereby tuning the background solution for ϕ is the additional freedom in splitting perturbations from the background that we mentioned earlier. The way we have chosen to do perturbation theory forces the choice $N = 1$ in order to be consistent.

3.2.2 Conservation outside the horizon

For a canonical scalar field, Weinberg [25] showed that the solution for the perturbations is always in the conserved mode, so the solutions to the equations of motion for the perturbations from the scalar field are conserved perturbations. It is also possible to see this now, directly from the action, as shown by Maldacena [20].

Proving conservation

We want to show that $\mathcal{R} = \text{constant}$ and $\gamma_{ij} = \text{constant}$ is a solution to the equations of motion outside the horizon. That is, we want to show that for any wave number k with $k \ll aH$ we have (with f denoting the collection of fields $\mathcal{R}$ and γ) that

$$\delta S = \int \mathrm{d}^4 x \sqrt{-g} \frac{\delta \mathcal{L}}{\delta f} \bigg|_{f=\text{const.}} \delta f = 0 + \mathcal{O}\left(\frac{k}{aH}\right)^2. \tag{3.29}$$

We want this up to second order in k/aH to get that the solution approaches a constant quickly after horizon crossing. The way we will show this is by proving that outside the horizon $\mathcal{L}$ can be written as the sum of a total derivative in time and a part that vanishes on the constant solution,

$$S = \int \mathrm{d}^4 x \, (\partial_t W + T) \tag{3.30}$$

with $\dfrac{\delta T}{\delta f} \bigg|_{f=\text{const.}} = 0$, because then the desired conclusion directly follows:

$$\delta S_W = \int \mathrm{d}^4 x \, \partial_t \left(\frac{\delta W}{\delta f} \delta f \right) = \left[\frac{\delta W}{\delta f} \delta f \right]_{t=t_0}^{t=t_1} = 0, \tag{3.31}$$

so any total derivative terms play no role in the equations of motion and

$$\delta S = \int \mathrm{d}^4 x \left. \frac{\delta T}{\delta f} \right|_{f=\text{const.}} \delta f = 0, \tag{3.32}$$

then gives that constant f is a solution.

Any terms in the action that have a spatial derivative acting on f will give, outside the horizon, a contribution of $\mathcal{O}(k/aH)$, because a spatial derivative becomes multiplication by the wave number in Fourier space.[2] To prove conservation up to the desired order in k/aH it is enough to keep only terms that come with zero or one spatial derivatives. For terms with time derivatives, we need to determine whether or not the term belongs to T. Trivially, terms without any time derivatives on f do not belong to T, so these must be shown to be in $\partial_t W$. Any term in T must have at least two fields with a time derivative acting on it, in order to give a term proportional to $\partial_t f \delta f$ in δT, such that filling in $f = \text{constant}$ sets the term to 0 by hitting at least one $\partial_t f$. Only one field having a time derivative is not sufficient, as $(\ldots)\delta \partial_t f$ after partial integration can have ∂_t hitting a background quantity, such as a, rather than a field, giving $\dot{a}(\ldots)\delta f$ which is not (necessarily) zero on the constant solution. In the same way, it is not sufficient to have multiple time derivatives acting on a single field. The theory we are working with, however, only has zero or one time derivatives[3] working on a field (barring partial integrations which can be performed).

Total derivative term

Given this, it is enough to take the action up to first order in spatial derivatives in fields, first order in time derivatives in fields, but to all order in fields for the rest and show that it is a total derivative of time; the remaining (neglected) terms are then in T or well outside the horizon, after which it follows that any solution f to the equations of motion goes to a constant and thus is conserved, outside the horizon.

To show that all the relevant terms form a total derivative, let the constraint for N be solved completely, then plug the constraint back into the action, after which the action becomes very simple looking:

$$S = \frac{1}{2} \int \mathrm{d}^3 x \, \mathrm{d}t \, a^3 e^{3\mathcal{R}} N \left(M_{\text{Pl}}^2 {}^{(3)}R - 2V(\phi) \right). \tag{3.33}$$

Assume now that to first order in derivatives N is solved by $N = 1 + \delta N$ (i.e. δN solves the constraints in first-order in derivatives along with a suitable choice for N^i) and assume that N^i starts at zeroth order in derivatives, then $\nabla_i N^j$ starts at first order in derivatives. The Ricci scalar, ${}^{(3)}R$, is of second order in derivatives, so it can be neglected. From the constraint equation for N,

$$2V\delta N = 2M_{\text{Pl}}^2 H (3\dot{\mathcal{R}} - \nabla_i N^i), \tag{3.34}$$

so

$$S = -\frac{1}{2} \int \mathrm{d}^3 x \, \mathrm{d}t \, a^3 e^{3\mathcal{R}} \left(2V(\phi) + 2M_{\text{Pl}}^2 H (3\dot{\mathcal{R}} - \nabla_i N^i) \right). \tag{3.35}$$

---

[2]It is $\mathcal{O}(k/aH)$ and not just $\mathcal{O}(k)$, because aH is the relevant length scale in the problem; all derivatives in the action come with a relative factor of $1/a$.

[3]Because of the Gibbons-Hawking-York boundary term which cancels the second-order time derivatives of h_{ij} in the ADM formalism.

Using the background equations[4] for V, eq. (2.42),

$$\dot{H} + 3H^2 = \frac{1}{2M_{\text{Pl}}^2} V(\phi),$$

the action becomes

$$S = -\frac{1}{2} M_{\text{Pl}}^2 \int \mathrm{d}^3x \, \mathrm{d}t \, a^3 e^{3\mathcal{R}} \left(2\dot{H} + 6H^2 + 6H\dot{\mathcal{R}} - 2H\nabla_i N^i \right). \qquad (3.36)$$

The last term can be written as a total derivative and thus ignored[5], using the following identity for the covariant derivative of a vector:

$$\nabla_i V^i = \frac{1}{\sqrt{h}} \partial_i \left(\sqrt{h} V^i \right). \qquad (3.37)$$

The action is a total derivative, namely it is equal to

$$S = -M_{\text{Pl}}^2 \int \mathrm{d}^3x \, \mathrm{d}t \, \partial_t \left(a^3 e^{3\mathcal{R}} H \right). \qquad (3.38)$$

Therefore, we conclude that the solutions to the equations of motion are conserved outside of the horizon.

3.2.3   Action for perturbations

To get the action for perturbations, the constraints need to be solved. This is a complicated task and therefore it is done order by order in powers of (perturbation) fields; expanding all quantities order by order and then plugging in the constraints gives the action order by order. The quadratic action gives the equations of motion and the higher order actions give the interactions. We know from perturbation theory that at first order in perturbations the scalar and tensor decouple, so the quadratic actions for the scalar and tensor decouple. At cubic and higher order, there will be terms that mix scalar and tensor degrees of freedom. From this point onwards we focus only on the scalar field, $\mathcal{R}$, and omit the tensor, γ_{ij}, and all mixing between scalars and tensors. The way the quantum field theory calculations will be set up for calculating correlation functions (the 'in-in formalism') makes this possible: the tensor terms will not mix with the calculated answers for the scalar, but instead give additional contributions only. Therefore, tensors can be omitted and we can still be sure that the answers we get for scalars do not need to be modified when tensors are included. For the rest of the calculations, we set $M_{\text{Pl}} = 1$ for convenience of notation.

To get S_2, the action quadratic in $\mathcal{R}$, it is enough to solve the constraints up to first order, because a second order term in N would multiply the constraint, $\delta\mathcal{L}/\delta N$, evaluated at zeroth order, which is 0. A similar argument goes for N^i.[20] To this order, the constraints are solved by

$$N = 1 + \frac{\dot{\mathcal{R}}}{H}, \quad N^i = \partial_i \left(-\frac{R}{a^2 H} + \tilde{\chi} \right), \quad \partial^2 \tilde{\chi} = \frac{1}{2} \epsilon \dot{\mathcal{R}}. \qquad (3.39)$$

---

[4]More accurate is to say: using the zeroth order (in fields) solution to the N constraint.

[5]It vanishes either because a manifold without boundary is used and all quantities are assumed to be square-integrable, or simply because we are currently only concerned with the equations of motions, to which boundary terms have been shown to not contribute.

Note that at first order, there is no 'true' vector to source N^i, as the first object with an index is γ_{ij}, but we know this to decouple from $\mathcal{R}$ at this order, so that only leaves partial derivatives ∂_i of a scalar. The quadratic action for scalars is

$$S_2 = \int \mathrm{d}^4 x \, \epsilon a^3 \left(\dot{\mathcal{R}}^2 - \frac{1}{a^2} (\partial \mathcal{R})^2 \right).$$
(3.40)

In a similar way, the cubic action, which describes the leading order interactions, can be found using the same order in constraints and is given by:[9]

$$
\begin{aligned}
S_3 = \int \mathrm{d}^4 x \bigg(& a^3 \epsilon^2 \mathcal{R} \dot{\mathcal{R}}^2 + a \epsilon^2 \mathcal{R} (\partial \mathcal{R})^2 - 2 a \epsilon^2 \dot{\mathcal{R}} (\partial \mathcal{R})(\partial \chi) \\
& + \frac{a^3 \epsilon \dot{\eta}}{2} \mathcal{R}^2 \dot{\mathcal{R}} + \frac{\epsilon^3}{2a} (\partial \mathcal{R})(\partial \chi) \partial^2 \chi + \frac{\epsilon^3}{4a} (\partial^2 \mathcal{R})(\partial \chi)^2 \\
& + f(\mathcal{R}) \frac{\delta L}{\delta \mathcal{R}} \Big|_1 \bigg).
\end{aligned}
$$
(3.41)

The function $f(\mathcal{R})$ is quadratic in $\mathcal{R}$ and its derivatives. We also modified the definition of χ to be different from both Chen [9]'s and Maldacena [20]'s definition:

$$\chi = a^2 \partial^{-2} \dot{\mathcal{R}},$$
(3.42)

where the difference is a factor of $2a^2/\epsilon$ from Maldacena and a factor of $1/\epsilon$ from Chen. Every spatial derivative naturally comes with a factor of a^{-1}, making it a derivative of physical distance, and this definition makes that explicit; every term in the integral has a common factor of a^3, with spatial derivatives modifying this power accordingly (maybe it would be more accurate to say that the natural integral measure is $a^3 \, \mathrm{d}^3 x$ of physical volume). Finally, we prefer to have the slow-roll parameters explicitly expressed in the integral, to be able to tell at a glance to which order in slow-roll a given term is. Typically, $\epsilon \dot{\eta}$ is of third order in slow-roll, but it can vary a lot more on small timescales than ϵ^3, being then closer to $\mathcal{O}(\epsilon)$ for a short time.

Field shift

The function $f(\mathcal{R})$ multiplies the first order equations of motion (from S_2) and can be removed by performing a field shift to a field $\mathcal{Q}$:

$$\mathcal{R} \to \mathcal{Q} - f(\mathcal{Q}) = \mathcal{Q} + \frac{\eta}{4} \mathcal{Q}^2 + \dots,$$
(3.43)

where '...' denotes the terms in f which have derivatives on $\mathcal{Q}$; at the order in power of perturbations and slow-roll parameters considered here, all these terms with derivatives give no contribution to the correlation functions of interest, therefore they can be ignored. In Fourier space, the field shift is given by

$$\mathcal{R}(k) \to \mathcal{Q}(k) + \frac{\eta}{4} \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \mathrm{d}^3 q \, \mathcal{Q}(q) \mathcal{Q}(k-q) + \dots$$
(3.44)

Note that the multiplication in real space became convolution in Fourier space; convolution is denoted by a $\star$:

$$[f \star g](k) := \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \mathrm{d}^3 q \, f(q) g(k-q).$$
(3.45)

This field shift leaves the quadratic action unchanged, removes the $f(\mathcal{R})$ term from the cubic action and only further modifies the action at order S_4 and beyond. The calculation presented here goes up to S_3, so the only side-effect will be a fixed contribution that comes from translating correlation functions from $\mathcal{Q}$ to $\mathcal{R}$.

Tensor perturbations

Finally, although we will not discuss these in much detail, the quadratic action for tensor perturbations is

$$S_2^{\gamma} = \frac{1}{8} \int \mathrm{d}^4 x \, a^3 \left(\dot{\gamma}_{ij} \dot{\gamma}_{ij} - \frac{1}{a^2} \partial_l \gamma_{ij} \partial_l \gamma_{ij} \right). \tag{3.46}$$

Here, summation is implied without the use of the metric, which would have been the case when writing raised indices. This is essentially the action for two massless scalar fields, labelled by polarisation.[20][5]

We now have the complete classical field theory setup to perform quantisation on and calculate what quantum fluctuations do for the observables of inflation, such as the CMB. First we take a small detour to talk about the relevant expectation values of powers of the perturbations, which are the primary observables of the quantum field theory of the canonical scalar field. After we have explained what the observables are and how they are of interest to us, we will proceed with quantising the field theory and showing how to calculate the correlators of perturbations.

3.3 Correlators of perturbations

The primary observables of the quantum field theory of the canonical scalar field are the correlation functions of the perturbations. It is these perturbations that seed the inhomogeneities and their correlators are directly related to observations of the CMB. Assume for the moment that the perturbations have been quantised, giving operators $\hat{\mathcal{R}}$ and $\hat{\gamma}_{ij}$; these are the primary observables of the theory. The background evolution of the field, which determines the scale factor a and the slow-roll parameters, can be indirectly measured via these observables, because their expectation values depend on the value of the background field. Because of the conservation outside of the horizon, the value of the background is imprinted onto the mode as it leaves the horizon. It is most convenient to express things in Fourier space for this reason, because each mode k leaves the horizon at a time $aH \approx k$, so the background is imprinted at the time t_k when $a(t_k)H(t_k) = k$, which is determined per mode, so for each Fourier transformed $\mathcal{R}(k,t)$ rather than $\mathcal{R}(x,t)$.[6]

3.3.1 Gaussian random variables

A Gaussian distribution variable X can be fully characterised by its mean and variance,

$$\mu = E(X), \tag{3.47}$$

$$\sigma^2 = E(X^2) - E(X)^2, \tag{3.48}$$

$$E((X-\mu)^n) = \begin{cases} 0 & n \text{ odd} \\ \sigma^n(n-1)!! & n \text{ even} \end{cases}. \tag{3.49}$$

The expectation values $E((X-\mu)^n)$ are called the *central moments* of X and for a Gaussian variable these are completely determined by the second central moment and the mean (the first moment). Note that $n!!$ is the double factorial (the product of all numbers up to n with the same parity as n). If the perturbations are completely Gaussian then all the information on its statistics is contained in the second central moment; taking other moments offers no additional information.

Because the perturbations are small, higher moments are much smaller than lower moments and become increasingly difficult to detect. For this reason, measurements determine the second and third central moments of the perturbation; the first moment is zero by definition of 'perturbation', so central moment and moment can be used interchangeably. The measured perturbations in the CMB are Gaussian to a very high degree, but not perfectly, so in particular that means that the third moment will be non-zero and this is a direct measurement of the non-Gaussianity of the perturbations; in fact, the deviation of the third moment from zero can be taken as the definition of 'non-Gaussianity' for the perturbations.

The simplest way to get an approximately Gaussian distribution is to take a Gaussian distribution X, with $\mu = 0$, and make it non-Gaussian by adding to

---

[6]We will be a bit sloppy about distinguishing between a function and its Fourier transformed, letting either the context or the argument (e.g. x or k) determine which one is meant, but this should not cause ambiguity.

it a non-linear term:

$$X \rightarrow X' = X + \frac{3}{5} f_{\mathrm{NL}} \left(X^2 - E(X^2) \right). \tag{3.50}$$

The parameter f_{NL} is a real number that signifies the departure from Gaussianity; 'NL' stands for non-linear and the factor of $3/5$ (or $-3/5$) is by convention[4]. The (central) moments for this shifted variable become

$$\mu' = E(X') = 0, \tag{3.51}$$

$$\sigma'^2 = E(X'^2) = \sigma^2 + \frac{18}{25} f_{\mathrm{NL}}^2 \sigma^4, \tag{3.52}$$

$$E(X'^3) = \frac{18}{5} f_{\mathrm{NL}} \sigma^4 + \frac{216}{125} f_{\mathrm{NL}}^3 \sigma^6, \tag{3.53}$$

For $|f_{\mathrm{NL}}| \ll \sigma^{-1}$ the induced non-Gaussianity does not significantly deform the second moment and in this limit,

$$\mu' = 0, \tag{3.54}$$

$$\sigma'^2 = \sigma^2, \tag{3.55}$$

$$E(X'^3) = \frac{18}{5} f_{\mathrm{NL}} \sigma^4, \tag{3.56}$$

so the f_{NL}-induced non-Gaussianity linearly shifts the third moment away from zero. The standard deviation σ is the measure of the size of the perturbations and we see that

$$E(X'^3) \sim f_{\mathrm{NL}} E(X'^2) E(X'^2), \tag{3.57}$$

so the third moment is proportional to the square of the second moment, with a small proportionality constant (in units of σ^{-1}). In general, of course, the non-Gaussianity can take any form and then measurements of higher moments are needed to determine its statistics; because we currently do not have access to higher moments in the CMB, we are forced to pick a presumed form of the non-Gaussianity and then we can express our measurements in terms of the presumed form.

3.3.2 Symmetries

Translation

For the mode functions, $\mathcal{R}(\vec{k}, t)$, the 'Gaussian' random variables are not directly $\mathcal{R}(\vec{k}, t)$ for each $\vec{k}$. Because of homogeneity, the expectation values,

$$\langle \mathcal{R}(\vec{x}_1, t) \mathcal{R}(\vec{x}_2, t) \cdots \mathcal{R}(\vec{x}_n, t) \rangle, \tag{3.58}$$

must be invariant under spatial translations, which in Fourier space leads to that the expectation value must have a momentum-conservation delta function

$$\left\langle \mathcal{R}(\vec{k}_1, t) \cdots \mathcal{R}(\vec{k}_n, t) \right\rangle = (2\pi)^3 \delta^{(3)}(\vec{k}_1 + \ldots + \vec{k}_n) \left\langle \mathcal{R}(\vec{k}_1, t) \cdots \mathcal{R}(\vec{k}_n, t) \right\rangle', \tag{3.59}$$

where the prime indicates the expectation value with the momentum conservation part stripped off (the *stripped correlator*). From this it can directly be seen that $\mathcal{R}(k, t)$ is not a Gaussian variable on its own, by momentum conservation:[7]

$$\left\langle \mathcal{R}(\vec{k}, t) \mathcal{R}(\vec{k}, t) \right\rangle = 0. \tag{3.60}$$

Instead, $\mathcal{R}(\vec{k}, t)$ behaves like a (near) Gaussian random variable with its adjoined $R(-\vec{k}, t)$:

$$\left\langle \mathcal{R}(\vec{k}, t) \mathcal{R}(-\vec{k}, t) \right\rangle \neq 0. \tag{3.61}$$

Rotation

Because of momentum-conservation, the non-trivial momentum configurations of n momenta form an n-sided polygon in three-dimensional space. Rotational invariance further restricts the freedom in expectation values by imposing that only the shape and size of the n-sided polygon matters, but not its orientation (its embedding into three-dimensional space). There are naïvely $3n$ degrees of freedom; momentum conservation reduces this to $3n - 3$. Rotational invariance allows us to align the axis of the coordinate system with, say, the first momentum vector $\vec{k}_1$, losing another 2 degrees of freedom. For $n = 2$ this leaves $3n - 3 - 2 = 1$ degree of freedom, for instance the length of k_1. For $n > 2$, we can then choose our last rotational degree of freedom to rotate the coordinate system around the k_1 axis (our x-axis) and make, say, the y-component of k_2 zero. After this, the rotational degrees of freedom are expended. This leaves $3n - 3 - 2 - 1 = 3n - 6$ degrees of freedom for $n \geq 3$. In the most relevant cases, this means that a correlator of two fields has one degree of freedom, the size of the momentum, k; a correlator of three fields has three degrees of freedom, for instance the shape of the triangle (two angles) and its size (size of any side), or one angle and the size of two sides, or the size of three sides (k_1, k_2, and k_3), or the overall size $K = k_1 + k_2 + k_3$ plus the size of two sides, etc.

Approximate scale-invariance

The dependence of the expectation values on the scale of k is typically small, so it is convenient to work with expressions that are scale-independent whenever the expectation value is scale independent. From the assumption that a general expectation value is scale independent, for $\vec{x} \rightarrow \lambda \vec{x}$,

$$\langle \mathcal{R}(\lambda \vec{x}_1, t) \cdots \mathcal{R}(\lambda \vec{x}_n, t) \rangle = \langle \mathcal{R}(\vec{x}_1, t) \cdots \mathcal{R}(\vec{x}_n, t) \rangle, \tag{3.62}$$

it follows that in momentum space $\vec{k}$ scales as $\vec{k} \rightarrow \vec{k}/\lambda$ and therefore

$$\left\langle \mathcal{R}\left(\frac{\vec{k}_1}{\lambda}, t\right) \cdots \mathcal{R}\left(\frac{\vec{k}_n}{\lambda}, t\right) \right\rangle = \lambda^{3n} \left\langle \mathcal{R}(\vec{k}_1, t) \cdots \mathcal{R}(\vec{k}_n, t) \right\rangle, \tag{3.63}$$

so to get scale-invariant quantities from an expectation value of n fields we should multiply the expectation value by k^{3n}:

$$\left\langle \mathcal{R}(\vec{k}_1, t) \cdots \mathcal{R}(\vec{k}_n, t) \right\rangle = \frac{S_n(\vec{k}_1, \ldots, \vec{k}_n; t)}{k_1^3 \cdots k_n^3}. \tag{3.64}$$

---

[7]The result is proportional to $\delta^{(3)}(k)$ and we assume that the expectation values are smooth functions, so the result is 0.

This gives S_n as a scale-invariant quantity whenever the expectation values are scale-invariant. It is also useful to know how the delta-function scales, namely as

$$\delta^{(3)}\left(\frac{\vec{k}}{\lambda}\right) = \lambda^3 \delta^{(3)}(\vec{k}), \tag{3.65}$$

so to define a scale invariant quantity S_n' from the stripped correlator the full correlator must be multiplied by k^{3n-3}:

$$\left\langle \mathcal{R}(\vec{k}_1, t) \cdots \mathcal{R}(\vec{k}_n, t) \right\rangle = \frac{S_n'(\vec{k}_1, \ldots, \vec{k}_n; t)}{k_2^3 \cdots k_n^3} (2\pi)^3 \delta^{(3)}(\vec{k}_1 + \ldots + \vec{k}_n). \tag{3.66}$$

To get a scale-invariant quantity, it is enough to multiply by any combination of momenta that gives the appropriate power; we have chosen a democratic combination for the full correlator and a notationally convenient combination for the stripped correlator. For the stripped correlator of three field, a better (democratic) choice is available, as we shall see shortly.

3.3.3 Power spectrum and bispectrum definition

For the perturbations, the second moment is referred to as the *two-point function*:[9]

$$\lim_{t \to \infty} \left\langle \mathcal{R}(\vec{k}, t) \mathcal{R}(\vec{k}', t) \right\rangle = \frac{(2\pi)^2 P_{\mathcal{R}}(k)}{2k^3} (2\pi)^3 \delta^{(3)}(\vec{k} + \vec{k}'). \tag{3.67}$$

The quantity $P_{\mathcal{R}}$ defined here is the *power spectrum*; note the factor of k^3 in the denominator to get a (potentially) scale-independent power spectrum.[8] A scale-independent power spectrum also directly means that $P_{\mathcal{R}}(k)$ does not depend on k. Technically, 'power spectrum' only refers to $P_{\mathcal{R}}$, but we will use it to also refer to the two-point function. The third moment is referred to as the *three-point function* or *bispectrum*:[9]

$$\lim_{t \to \infty} \left\langle \mathcal{R}(\vec{k}_1, t) \mathcal{R}(\vec{k}_2, t) \mathcal{R}(\vec{k}_3, t) \right\rangle = \frac{S(k_1, k_2, k_3)}{k_1^2 k_2^2 k_3^2} (2\pi)^4 P_{\mathcal{R}}(k_*)^2 (2\pi)^3 \delta^{(3)}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3). \tag{3.68}$$

The quantity $S(k_1, k_2, k_3)$ is the *shape function*, or shape, of the bispectrum. Note again the (democratically chosen) factor of k^6 in the denominator to make the shape (potentially) scale-invariant. We will come back to the shape function later and explain why it's so named, but for now the message is that it encodes the shape of the bispectrum. Note the direct inclusion of the power spectrum into the definition, as motivated by the relation between the third and second moments in the f_{NL} model, eq. (3.57); from this one can also see that the shape S is related to f_{NL}. The power spectrum is evaluated at a specific wave number k_*, which should be chosen to be representative for the given model/observation, in order to normalise the shape function/f_{NL}, whilst keeping all scale-dependence of the bispectrum in the shape function S. The power spectrum gives the size of the perturbations, while the bispectrum gives the size of the non-Gaussianities in the perturbations.

---

[8]One might wonder at this point why the artificial factor of $(2\pi)^2/2$ is introduced here (and also in the bispectrum); this is the convention, see e.g. [9], but the rational behind this is unclear to this author.

Tensor power spectrum definition

Similarly, a power spectrum is defined for tensor perturbations. For a given polarisation γ^s the tensor power spectrum is defined by:

$$\lim_{t \to \infty} \left\langle \gamma^s(\vec{k}, t) \gamma^{s'}(\vec{k}', t) \right\rangle = \frac{(2\pi)^2 P_\gamma^s(k)}{2k^3} (2\pi)^3 \delta_{ss'} \delta^{(3)}(\vec{k} + \vec{k}'). \tag{3.69}$$

One could also define a bispectrum for tensors, but considering that typically the power spectrum for tensors is too small to be observable in the CMB it seems pointless to do so.

3.3.4 Canonical scalar field

Power spectrum

For the canonical scalar field, the power spectrum for scalar and tensors has been calculated and are well-known in literature. For a general potential $V(\phi)$, as mentioned, the results are expanded in powers of slow-roll parameters, with the leading order in slow-roll power spectrum:[9][20][5][9]

$$P_{\mathcal{R}}(k) = \frac{H(t_k)^2}{8\pi^2 M_{\mathrm{Pl}}^2 \epsilon(t_k)}. \tag{3.70}$$

The size of the scalar perturbations from measurements is

$$P_{\mathcal{R}}(k_*) = (2.142 \pm 0.049) \cdot 10^{-9},$$

with 68% confidence level at $k_* = 0.05 \, \mathrm{Mpc}^{-1}$ (Planck (2015), [23]). The tensor power spectrum (sum of both polarisations) is given by:[20][5]

$$P_\gamma(k) = \frac{H(t_k)^2}{\pi^2 M_{\mathrm{Pl}}^2}. \tag{3.71}$$

We can directly see that $\epsilon = 0$ gives divergent $P_{\mathcal{R}}$. This is because in pure de Sitter there are no physical scalar perturbations; they are pure gauge modes which can be gauged away. This is not true for tensor perturbations, which cannot be gauged away in de Sitter. Therefore, to have physical scalar perturbation we must have $\epsilon \neq 0$, but then it immediately follows that $P_{\mathcal{R}}$ *must* have scale-dependence, because then H changes in time, so $H(t_k)$ changes with k. The slow-roll parameter ϵ is directly related to both scale-dependence of the power spectrum and the presence of physical perturbations.

The ratio of the tensor power spectrum to scalar power spectrum is known as the *tensor-to-scalar ratio*, denoted by

$$r = \frac{P_\gamma(k_*)}{P_{\mathcal{R}}(k_*)} = 8 \, \epsilon(k_*), \tag{3.72}$$

and has not been measured yet, but the current constraints on it are $r < 0.07$ (95% CL) at $k_* = 0.05 \, \mathrm{Mpc}^{-1}$ (BICEP2/Keck, [2]). This also gives a direct bound on ϵ for the scalar field model, $\epsilon < 0.01$.

---

[9]To compare with other references, note that $\dot{\phi}^2 = -M_{\mathrm{Pl}}^2 \dot{H} = M_{\mathrm{Pl}}^2 \epsilon H^2$; there are factors of 2 differences stemming from disagreement about the exact definition of $P_{\mathcal{R}}$.

Spectral index and running

The quantity (number) that characterises the scale-dependence of the (scalar) power spectrum is the *spectral index*, n_s, defined as

$$n_\mathrm{s} - 1 := \left. \frac{\mathrm{d} \log P_\mathcal{R}}{\mathrm{d} \log k} \right|_{k_*}. \tag{3.73}$$

The quantity n_s is the spectral index and we call $n_\mathrm{s}-1$ the *spectral tilt*, or just tilt if it's clear from the context. The value of $n_\mathrm{s} - 1$ obviously depends on the value of k_* chosen, which is referred to as the *pivot*. Two of the most common choices are $n_\mathrm{s} = 0.968 \pm 0.012$ (68% CL) at $k_* = 0.002\,\mathrm{Mpc}^{-1}$ (WMAP-7year (2011), [18]) and $n_\mathrm{s} = 0.9667 \pm 0.0040$ (68% CL) at $k_* = 0.05\,\mathrm{Mpc}^{-1}$ (Planck (2015), [23]). The same pivot value should be used everywhere, e.g. for the measured power spectrum $P_\mathcal{R}(k_*)$ the same pivot appears. With 99.5% confidence level (CL), n_s is smaller than 1, so $n_\mathrm{s} - 1$ is negative. The spectral tilt $(n_\mathrm{s} - 1)$ gives the tilt of the power spectrum as a function of $\log k$, which would otherwise be a flat horizontal line. There is also the possibility that the tilt of the power spectrum has a running, that is that it is slightly curved rather than straight, characterised by the *running* of the spectral index, α (or sometimes α_s), defined as

$$\alpha := \left. \frac{\mathrm{d}^2 \log P_\mathcal{R}}{\mathrm{d} \log k^{\,2}} \right|_{k_*}. \tag{3.74}$$

A running of the spectral index has not been detected, so far, but its value is constrained to $\alpha = -0.002 \pm 0.013$ (95% CL) at $k_* = 0.05\,\mathrm{Mpc}^{-1}$ (Planck (2015), [23]), compatible with zero running.

Definition of spectral index

The way these quantities are defined might seem slightly arbitrary, but there is reason behind these particular definitions. First, the use of both log's makes the tilt and running inherently dimensionless numbers, which is a nice property. The power spectrum can be split into a scale-independent factor and a part that captures the scale-dependence,

$$P_\mathcal{R}(k) = P_\mathcal{R}^* \cdot \left(\frac{k}{k_*} \right)^{n(k)}, \tag{3.75}$$

where $P_\mathcal{R}^* = P_\mathcal{R}(k_*)$ and $n(k)$ is some index function, $n \ll 1$ for a small scale-dependence. To get n and its derivatives out of this, we first take the log,

$$\log P_\mathcal{R}(k) = \log P_\mathcal{R}^* + n(k) \log \left(\frac{k}{k_*} \right), \tag{3.76}$$

and then Taylor-expand $P_\mathcal{R}$ to get a Taylor-expansion of n, but from the formula we can see that a natural variable for n is not k, but rather $\log k$, such that if we expand n in terms of $\log k$ around $k = k_*$ we get

$$\log P_\mathcal{R}(k) = \log P_\mathcal{R}(k_*) + n(k_*) \log \left(\frac{k}{k_*} \right) + \frac{\mathrm{d}n(k_*)}{\mathrm{d} \log k} \log \left(\frac{k}{k_*} \right)^2 + \mathcal{O}\left(\log \left(\frac{k}{k_*} \right)^3 \right), \tag{3.77}$$

which is a much cleaner expansion than for expansion in k. From this, the natural choice is to let $n_{\mathrm{s}} - 1$ be the first Taylor coëfficient of Taylor-expanding $\log P_{\mathcal{R}}$ in $\log k$ and α the second Taylor coëfficient, so

$$\log P_{\mathcal{R}}(k) = \log P_{\mathcal{R}}(k_*) + (n_{\mathrm{s}} - 1)\log\left(\frac{k}{k_*}\right) + \frac{1}{2}\alpha\log\left(\frac{k}{k_*}\right)^2 + \mathcal{O}\left(\log\left(\frac{k}{k_*}\right)^3\right),$$
(3.78)

which gives the definitions for $n_{\mathrm{s}} - 1$ and α as before, and the relation to $n(k)$ as

$$n(k) = (n_{\mathrm{s}} - 1) + 2\alpha\log\left(\frac{k}{k_*}\right) + \mathcal{O}\left(\log\left(\frac{k}{k_*}\right)^2\right).$$
(3.79)

As a final technical note, the Taylor expansion employed is not in the smallness of $\log k/k_*$, but rather in the smallness of the coëfficients, relative to $\log k/k_*$, of the Taylor series.

Formula for spectral index

For the scalar field model, the tilt and running can be expressed in terms of the slow-roll parameters, using the expression for the power spectrum from eq. (3.70) and the definitions of n_{s} and α, eqs. (3.73) and (3.74). In eq. (3.70) time is evaluated at horizon crossing for the mode k, at time t_k such that $a(t_k)H(t_k) = k$, so this gives

$$\mathrm{d}\log k = \frac{\mathrm{d}k}{k} = \frac{\mathrm{d}(aH)}{aH} = \frac{\dot{a}}{a}\,\mathrm{d}t + \frac{\dot{H}}{H}\,\mathrm{d}t = H(1 - \epsilon)\,\mathrm{d}t,$$
(3.80)

or equivalently

$$\frac{\mathrm{d}}{\mathrm{d}\log k} = \frac{1}{H(1 - \epsilon)}\frac{\mathrm{d}}{\mathrm{d}t} = \frac{1}{H}(1 + \epsilon + \epsilon^2 + \ldots)\frac{\mathrm{d}}{\mathrm{d}t}.$$
(3.81)

We only need the zeroth order expression in slow-roll here, because the power spectrum as given by eq. (3.70) is only given to leading-order in slow-roll, so we use

$$\frac{\mathrm{d}}{\mathrm{d}\log k} = \frac{1}{H}\frac{\mathrm{d}}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}\mathcal{N}}.$$
(3.82)

The spectral tilt of the scalar field is:

$$n_{\mathrm{s}} - 1 = -\frac{\mathrm{d}}{\mathrm{d}\mathcal{N}}\left(2\log H - \log\epsilon\right) + \mathcal{O}(\epsilon^2) = -2\epsilon - \eta + \mathcal{O}(\epsilon^2).$$
(3.83)

The parameters ϵ and η should be evaluated at k_*. The tilt for the model is negative and small, which is in line with the value from measurements, if both ϵ and η are taken to be constrained by slow-roll. We mentioned before that η should be small for the field to roll slowly down the potential, but that was of course a slightly artificial constraint; here we see the more physical reason, namely that not taking η to be small would introduce significant scale-dependence that is not in line with data. From the measurement result, $2\epsilon + \eta \approx 0.04$ and $\epsilon < 0.01$, so an estimate for η is $\eta \lesssim 0.02$.

Running of the spectral index

The running of the spectral index is calculated in the same way to give

$$\alpha = -2\epsilon\eta - \eta\xi + \mathcal{O}(\epsilon^3). \tag{3.84}$$

Again, the parameters should be evaluated at k_*. Remember that $\xi = -\partial_{\mathcal{N}} \log \eta$ is the slow-roll parameter after η, but its value is not constrained to always be small, so locally (for a short time) the second term, $\eta\xi$, could be of order ϵ or larger. For a discussion of models where the potential has features that give rise to such running, see for example Flauger and Pajer [15] or Chen, Easther and Lim [11][10]. The order estimate of $\mathcal{O}(\epsilon^3)$ should be taken to mean $\mathcal{O}(\epsilon^3)$ on average, but slow-roll parameters after η may vary becoming of order $\mathcal{O}(\epsilon^2)$–$\mathcal{O}(\epsilon)$ for short times. On average, the running is of order $\mathcal{O}(\epsilon^2)$, which is small and compatible with measurements.

3.3.5 Bispectrum

For the bispectrum, because there are three degrees of freedom, two if one assumes scale-invariance, just giving the size doesn't contain all the information. In the example non-Gaussianity model, the parameter f_{NL} parametrises the amount of non-Gaussianity and for this model it gave the size of the third moment. We want to do an analogous thing for the bispectrum: give a number for f_{NL} that indicates the degree of non-Gaussianity. The shape function, S, encodes both shape and size, so to get a size number a specific shape (momentum configuration) must be picked and the value of S for that shape is the 'size'. Of course, different shapes may have their maximum in different limits and certain shapes may have divergences for certain limits of momentum configuration, so there is no 'one size fits all' number that can be given. There are, however, some common shapes that come from models and that can be calculated from the data, so typically to give the amount of non-Gaussianity one gives the presumed (approximate) shape plus the size of that shape, f_{NL}. The shape of the bispectrum is the dependence of S on the momentum ratio's k_2/k_1 and k_3/k_1 whilst keeping the momentum scale $K = k_1 + k_2 + k_3$ fixed. The bispectrum can also have a running, which is the dependence of S on the total momentum K; if the bispectrum is scale-invariant then S does not depend on K. The size of the bispectrum, f_{NL}, is defined as:[9]

$$S(k,k,k) \to \frac{9}{10} f_{\mathrm{NL}}(k), \tag{3.85}$$

so as S evaluated at the equilateral triangle configuration, that is $k_1 = k_2 = k_3$ $(= K/3)$. If the bispectrum is scale-invariant, then $f_{\mathrm{NL}}(k)$ is independent of k.

Shapes

The most common shapes are the *equilateral shape*, *local shape*, and *orthogonal shape*. The local shape is

$$S_{\mathrm{local}}(k_1, k_2, k_3) = \frac{1}{3}\left(\frac{k_1^2}{k_2 k_3} + 2 \text{ perm.}\right), \tag{3.86}$$

where '+2 perm.' stands for the two other permutations of k_1, k_2, k_3 and is characterised by the fact that it peaks, divergently, in the *squeezed limit*, that is when one momentum is much smaller than the other two momenta that are roughly the same size, e.g. $k_1 \ll k_2 \approx k_3$:

$$S_{\text{local}}(k_1, k_2, k_3) \to \frac{2}{3} \frac{k_2}{k_1}. \tag{3.87}$$

The simple non-Gaussianity model we introduced has this shape and it's local in real space, hence the name 'local'. The equilateral shape is

$$S_{\text{equil}}(k_1, k_2, k_3) = 6 \left(\frac{k_1}{k_2} + 5 \text{ perms.} \right) - 6 \left(\frac{k_1^2}{k_2 k_3} + 2 \text{ perms.} \right) - 12. \tag{3.88}$$

It peaks in the equilateral triangle configuration. The *orthogonal shape* is a shape that is orthogonal to both the local and equilateral shape, for a properly defined inner product on shapes (see e.g. [15][9]), and is given by

$$S_{\text{ortho}}(k_1, k_2, k_3) = -18 \left(\frac{k_1^2}{k_2 k_3} + 2 \text{ perms.} \right) + 18 \left(\frac{k_1}{k_2} + 5 \text{ perms.} \right) - 48. \tag{3.89}$$

From the Planck (2015) data[24], the constraints on f_{NL} for the above shapes are at 68% confidence level: $f_{\text{NL}}^{\text{local}} = 0.8 \pm 5.0$, $f_{\text{NL}}^{\text{equil}} = -4 \pm 43$, $f_{\text{NL}}^{\text{ortho}} = -26 \pm 21$. We see that typically the bispectrum is at the order of the power spectrum squared, because $f_{\text{NL}} \sim \mathcal{O}(1) - \mathcal{O}(10)$, so it's very small.

Scalar field

For the scalar field model, the bispectrum is a combination of the local shape, with $f_{\text{NL}}^{\text{local}} \sim \eta$, and new kind of shape, with $f_{\text{NL}} \sim \epsilon$, given up to first order in slow-roll by:[20][9]

$$\begin{aligned} S = &\frac{\epsilon}{8} \left[\left(\frac{k_1}{k_2} + 5 \text{ perms.} \right) - \left(\frac{k_1^2}{k_2 k_3} + 2 \text{ perms.} \right) + \frac{8}{K} \left(\frac{k_1 k_2}{k_3} + 2 \text{ perms.} \right) \right] \\ &+ \frac{\eta}{8} \left(\frac{k_1^2}{k_2 k_3} + 2 \text{ perm.} \right). \end{aligned} \tag{3.90}$$

Thus, for the scalar field model, $f_{\text{NL}}^{\text{scalar}} = \frac{10}{72}(11\epsilon + 3\eta) = \mathcal{O}(\epsilon) \sim 0.03$. Non-linear effects in the CMB evolution will generate non-Gaussianity with $f_{\text{NL}} \sim \mathcal{O}(1)$, so the bispectrum from a canonical scalar field is unlikely to be detectable in the CMB.[9]

Chapter 4

(Bi)spectrum calculation

4.1 Quantisation of the scalar field

In the previous chapter we discussed the observables of a quantum field theory for inflation and showed the literature results for the canonical scalar field. Now, we turn our attention to how the results for the scalar field were calculated. The first step is quantising the field theory and after that the method of calculation is detailed. Quantisation is done using a semi-classical approximation for a quantum field theory in curved space time, where the background field (and the resulting metric) are kept classical and the perturbations are quantised.

The starting point is the action for the perturbations, the S_2 and S_3 actions, given by eqs. (3.40) and (3.41):

$$S_2 = \int \mathrm{d}^4x \, \epsilon a^3 \left(\dot{\mathcal{R}}^2 - \frac{1}{a^2}(\partial \mathcal{R})^2 \right),$$

$$S_3 = \int \mathrm{d}^4x \bigg(a^3 \epsilon^2 \mathcal{R} \dot{\mathcal{R}}^2 + a \epsilon^2 \mathcal{R} (\partial \mathcal{R})^2 - 2a \epsilon^2 \dot{\mathcal{R}} (\partial \mathcal{R})(\partial \chi)$$

$$+ \frac{a^3 \epsilon \dot{\eta}}{2} \mathcal{R}^2 \dot{\mathcal{R}} + \frac{\epsilon^3}{2a}(\partial \mathcal{R})(\partial \chi) \partial^2 \chi + \frac{\epsilon^3}{4a}(\partial^2 \mathcal{R})(\partial \chi)^2$$

$$+ f(\mathcal{R}) \frac{\delta L}{\delta \mathcal{R}} \bigg|_1 \bigg),$$

where we remind ourselves that the $f(\mathcal{R})$ term can be removed via the field shift $\mathcal{R} \to \mathcal{Q} - f(\mathcal{Q})$, eq. (3.43).

4.1.1 In-In formalism

The quadratic action looks 'simple', whereas the cubic action (and beyond) looks 'hard', so this motivates to use the interaction picture: the calculation of the expectation values will be done in the 'in-in' formalism of quantum mechanics; for a more detailed discussion of this and the quantisation procedure, see the excellent review by Chen [9]. The procedure works by putting the part of the Hamiltonian whose Schrödinger equation can be solved, denoted by H_0, in the Heisenberg picture and the rest, denoted by H_I, into the Schrödinger picture. Operators gain an implicit time dependence via H_0 and the states gain an implicit time dependence via H_I; for a (time-independent) operator $\hat{X}$ and a (time-independent) state $|\Psi\rangle$ the time dependence is:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{X}_\mathrm{I}(t) = i\left[\hat{H}_0(t), \hat{X}_\mathrm{I}(t)\right], \qquad\qquad \hat{X}_\mathrm{I}(t_0) = \hat{X}, \qquad (4.1)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}|\Psi(t)\rangle = -i\hat{H}_\mathrm{I}(t)|\Psi(t)\rangle, \qquad\qquad |\Psi(t_0)\rangle = |\Psi\rangle. \qquad (4.2)$$

This representation is very suited to calculating vacuum expectation values, because it pairs the 'difficult' time dependence with the 'simple' state. The expectation value of an operator $\hat{X}$ for a state $|\Psi\rangle$ can be written as

$$\langle \hat{X} \rangle_\Psi(t) = \langle \Psi(t)| \hat{X}_\mathrm{I}(t) |\Psi(t)\rangle$$

$$= \langle \Psi| \left[\bar{T} \exp\left(i \int_{t_0}^{t} \mathrm{d}t' \hat{H}_\mathrm{I}(t') \right) \right] \hat{X}_\mathrm{I}(t) \left[T \exp\left(-i \int_{t_0}^{t} \mathrm{d}t' \hat{H}_\mathrm{I}(t') \right) \right] |\Psi\rangle.$$

$$(4.3)$$

Here, T stands for time-ordering (of the exponential) and $\bar{T}$ for anti-time ordering. Note that a product of operators goes into a product of interaction picture operators as

$$O_1 \cdots O_n \to U_0 O_1 \cdots O_n U_0^{-1} = O_{\mathrm{I},1} \cdots O_{\mathrm{I},n}, \tag{4.4}$$

because U_0, the time-evolution operator of H_0, has the property $U_0^{-1} U_0 = \mathbb{1}$.

Power and bispectrum

Applied to the current case, H_0 comprises of S_2, that is all terms of second order in $\mathcal{R}$, and H_{I} comprises of S_3 and onwards, that is all terms of third or higher order in $\mathcal{R}$. The expectation values of interest are the power spectrum (two-point function) and bispectrum (three-point function). These correspond to the moments of $\mathcal{R}$ at a time t during inflation, evolved to outside the horizon $(t \to \infty)$, where they become frozen. The quantum state of the universe during inflation is denoted by $|\Psi\rangle = |0\rangle$ in the infinite past $(t_0 \to -\infty$; the start of inflation); the state $|0\rangle$ is referred to as the Bunch-Davies vacuum and is the equivalent of the Minkowski vacuum in de Sitter. The exact definition of the vacuum will be given in section 4.1.5 where the time-solution of $\mathcal{R}_{\mathrm{I}}(t)$ in the interaction picture is determined. It is convenient to work in Fourier space for the operators $\mathcal{R}_{\mathrm{I}}(x, t)$, so instead of x it becomes a function of k: $\mathcal{R}_{\mathrm{I}}(k, t)$. An expansion in H_{I} in eq. (4.3) is justified, because the perturbations are very small, giving for the scalar power spectrum at a time t during inflation:

$$\langle \mathcal{R}(k_1)\mathcal{R}(k_2)\rangle(t) = \langle 0| \, \mathcal{R}_{\mathrm{I}}(k_1, t)\mathcal{R}_{\mathrm{I}}(k_2, t) \, |0\rangle + $$
$$i \int_{-\infty}^{t} \mathrm{d}t' \, \langle 0| \, [H_{\mathrm{I}}(t'), \mathcal{R}_{\mathrm{I}}(k_1, t)\mathcal{R}_{\mathrm{I}}(k_2, t)] \, |0\rangle + \mathcal{O}\!\left(\mathcal{R}_{\mathrm{I}}^8\right), \tag{4.5}$$

and for the scalar bispectrum at a time t during inflation:

$$\langle \mathcal{R}(k_1)\mathcal{R}(k_2)\mathcal{R}(k_3)\rangle(t) = \langle 0| \, \mathcal{R}_{\mathrm{I}}(k_1, t)\mathcal{R}_{\mathrm{I}}(k_2, t)\mathcal{R}_{\mathrm{I}}(k_3, t) \, |0\rangle$$
$$+ i \int_{-\infty}^{t} \mathrm{d}t' \, \langle 0| \, [H_{\mathrm{I}}(t'), \mathcal{R}_{\mathrm{I}}(k_1, t)\mathcal{R}_{\mathrm{I}}(k_2, t)\mathcal{R}_{\mathrm{I}}(k_3, t)] \, |0\rangle \tag{4.6}$$
$$+ \mathcal{O}\!\left(\mathcal{R}_{\mathrm{I}}^9\right).$$

4.1.2 Field shift

The shift $\mathcal{R} \to \mathcal{Q} - f(\mathcal{Q})$ (eq. (3.43)) gives a field theory for $\mathcal{Q}$ derived from the field theory for $\mathcal{R}$. The bispectrum (eq. (4.6)) can also be calculated in the field theory for $\mathcal{Q}$, which is more convenient, because S_3 was made simpler. However, the (bi)spectrum for $\mathcal{Q}$ does not directly correspond to the (bi)spectrum for $\mathcal{R}$, but is related via the shift. One important detail is that S_2 is the same for both theories, so the expression for $\mathcal{R}(k, t)$ is the same as for $\mathcal{Q}(k, t)$. However, the vacuum state $|0\rangle_{\mathcal{R}}$ is **not** the same as the vacuum state $|0\rangle_{\mathcal{Q}}$: for instance, $\langle 0|_{\mathcal{Q}} \, \mathcal{R}(k, t) \, |0\rangle_{\mathcal{Q}} \neq 0$. This difference is critical in relating calculations in the two field theories. In the field theory for $\mathcal{Q}$, the operator $\mathcal{R}$ is a function of $\mathcal{Q}$, via the shift: $\mathcal{R}(\mathcal{Q}) = \mathcal{Q} - f(\mathcal{Q})$. Therefore, when calculating an expectation value for $\mathcal{R}$ in the field theory for $\mathcal{Q}$ this relation needs to be applied: the (bi)spectrum of $\mathcal{R}$ can be calculated in the field theory for $\mathcal{Q}$ by applying the

shift on the right side of eq. (4.6) with all other quantities (in particular, $|0\rangle$) taken in the field theory for $\mathcal{Q}$.

To calculate the (bi)spectrum and get the order in $\mathcal{R}$ at which the calculation is valid, the following two points are relevant. The first point is that the goal is to calculate the effect of the interactions in S_3. Therefore, from

$$H_{\mathrm{I}} = -L_{\mathrm{I},3} + \mathcal{O}\big(\mathcal{R}_{\mathrm{I}}^4\big), \tag{4.7}$$

with L_3 the Lagrangian defined by

$$S_3 = \int \mathrm{d}t\, L_3(t), \tag{4.8}$$

it is sufficient to take H_{I} at third order and neglect higher orders of $\mathcal{R}$ and, moreover, $H_{\mathrm{I},3}$ only involves terms from S_3. The second point is that all odd powers of the primary field sandwiched between vacuum states vanish. This is due to the way that the vacuum state $|0\rangle$ is defined (or, rather, will be defined) in section 4.1.5. For example, in the field theory for $\mathcal{Q}$, for every odd natural number n, $\langle 0|_{\mathcal{Q}}\, \mathcal{Q}_{\mathrm{I}}^n\, |0\rangle_{\mathcal{Q}} = 0$. In other words, $\mathcal{Q}_{\mathrm{I}}$ is Gaussian, but $\mathcal{Q}$ is not because it involves interactions.

Formula for (bi)spectrum

The above discussion gives that the power spectrum is easiest to calculate in the field theory for $\mathcal{R}$:

$$\langle \mathcal{R}(k_1)\mathcal{R}(k_2)\rangle(t) = \langle 0|_{\mathcal{R}}\, \mathcal{R}_{\mathrm{I}}(k_1,t)\mathcal{R}_{\mathrm{I}}(k_2,t)\,|0\rangle_{\mathcal{R}} + \mathcal{O}\big(\mathcal{R}_{\mathrm{I}}^6\big), \tag{4.9}$$

and that the bispectrum is easiest to calculate in the field theory for $\mathcal{Q}$:

$$
\begin{aligned}
\langle \mathcal{R}(k_1)\mathcal{R}(k_2)\mathcal{R}(k_3)\rangle(t) = {} & \frac{\eta}{4} \langle 0| \,[\mathcal{Q}_{\mathrm{I}} \star \mathcal{Q}_{\mathrm{I}}](k_1,t)\mathcal{Q}_{\mathrm{I}}(k_2,t)\mathcal{Q}_{\mathrm{I}}(k_3,t)\,|0\rangle + 2 \text{ perm.} \\
& + \frac{\eta^3}{64} \langle 0| \,[\mathcal{Q}_{\mathrm{I}} \star \mathcal{Q}_{\mathrm{I}}](k_1,t)[\mathcal{Q}_{\mathrm{I}} \star \mathcal{Q}_{\mathrm{I}}](k_2,t)[\mathcal{Q}_{\mathrm{I}} \star \mathcal{Q}_{\mathrm{I}}](k_3,t)\,|0\rangle \\
& + \text{ terms that vanish outside the horizon} \\
& + \text{ terms that vanish as } t \to \infty \\
& + i \int_{-\infty}^{t} \mathrm{d}t'\, \langle 0| \,[H_{\mathrm{I},3}(t'),\, \mathcal{Q}_{\mathrm{I}}(k_1,t)\mathcal{Q}_{\mathrm{I}}(k_2,t)\mathcal{Q}_{\mathrm{I}}(k_3,t)]\,|0\rangle \\
& + \mathcal{O}\big(\mathcal{R}_{\mathrm{I}}^8\big)
\end{aligned}
\tag{4.10}
$$

The '+ 2 perm.' on the first line stands for the three different choices of associating the convolution with one of the (external) momenta k_i; the 'terms that vanish' lines refer to terms that result from the other terms in the shift $f(\mathcal{Q})$.

4.1.3 Quantisation

To quantise the scalar perturbations, the canonical quantisation procedure is used. The classical equations of motion are solved in Fourier space, giving solutions $u(\vec{k}, \tau)$, such that classically

$$\mathcal{R}_{\mathrm{I}}(\vec{x}, \tau) = \int \frac{\mathrm{d}^3\vec{k}}{(2\pi)^3}\, u(\vec{k}, \tau)e^{-i\vec{k}\cdot\vec{x}}. \tag{4.11}$$

After quantisation, $\hat{\mathcal{R}}_I(x,t)$ and $\hat{\mathcal{R}}_I(k,t)$ become operators, and we demand that $\hat{\mathcal{R}}_I(x,t)$ is an observable, which requires it to be Hermitian, such that

$$\hat{\mathcal{R}}_I^\dagger(\vec{k},\tau) = \hat{\mathcal{R}}_I(-\vec{k},\tau). \tag{4.12}$$

If we take $\hat{\mathcal{R}}_I(\vec{k},\tau) = u(\vec{k},\tau)\hat{A}_{\vec{k}}$, then this translates to

$$u(\vec{k},\tau)^*\hat{A}_{\vec{k}}^\dagger = u(-\vec{k},\tau)\hat{A}_{-\vec{k}}. \tag{4.13}$$

This is elegantly solved by

$$u(\vec{k},\tau)\hat{A}_{\vec{k}} \;\rightarrow\; u(\vec{k},\tau)\hat{a}_{\vec{k}} + u^*(-\vec{k},\tau)\hat{a}_{-\vec{k}}^\dagger, \tag{4.14}$$

such that

$$\hat{\mathcal{R}}_I(\vec{k},t) = u(\vec{k},\tau)\hat{a}_{\vec{k}} + u^*(-\vec{k},\tau)\hat{a}_{-\vec{k}}^\dagger. \tag{4.15}$$

The commutation relations of $\mathcal{R}_I(\vec{x},t)$ with it's conjugate momentum, $\Pi_{\mathcal{R}}(\vec{x},t)$ are

$$[\mathcal{R}_I(\vec{x},t), \Pi_{\mathcal{R}}(\vec{y},t)] = i\delta^{(3)}(\vec{x}-\vec{y}), \tag{4.16}$$

which gives for $a_{\vec{k}}$ and $a_{-\vec{k}}^\dagger$ that they are a pair of creation/annihilation operators:

$$\left[\hat{a}_{\vec{k}}, \hat{a}_{-\vec{q}}^\dagger\right] = (2\pi)^3\delta^{(3)}(\vec{k}+\vec{q}). \tag{4.17}$$

This also gives for the normalisation of $u(\vec{k},\tau)$ that the Wronskian condition needs to be satisfied:

$$\epsilon a^3 u(\vec{k},\tau)\dot{u}^*(\vec{k},\tau) - \text{c.c.} = i. \tag{4.18}$$

This condition will be used when normalising the solutions to the classical equations of motion. To be able continue with creating the Fock space, we need to have the solutions to the classical equations of motion.

4.1.4 Equations of motion

The equations of motion for $\mathcal{R}_I(x,t)$ and $\mathcal{Q}_I(x,t)$ are given by (classically) solving eq. (4.1). The equation of motion for $\mathcal{R}$ is given by

$$\ddot{\mathcal{R}}(\vec{x},t) + \left(3H + \frac{\dot{\epsilon}}{\epsilon}\right)\dot{\mathcal{R}}(\vec{x},t) - \frac{1}{a^2}\vec{\nabla}^2\mathcal{R}(\vec{x},t) = 0, \tag{4.19}$$

or in Fourier space

$$\ddot{\mathcal{R}}(\vec{k},t) + \left(3H + \frac{\dot{\epsilon}}{\epsilon}\right)\dot{\mathcal{R}}(\vec{k},t) + \frac{\vec{k}^2}{a^2}\mathcal{R}(\vec{k},t) = 0. \tag{4.20}$$

This looks like the equation for a damped harmonic oscillator, if the coefficients would not depend on time:

$$\frac{\mathrm{d}^2\mathcal{R}_{\vec{k}}}{\mathrm{d}t^2} + 2d_{\vec{k}}\omega_{\vec{k}}\frac{\mathrm{d}\mathcal{R}_{\vec{k}}}{\mathrm{d}t} + \omega_{\vec{k}}^2\mathcal{R}_{\vec{k}} = 0, \tag{4.21}$$

where

$$\omega_{\vec{k}} = \frac{k}{a} \tag{4.22}$$

is the undamped frequency and

$$d_{\vec{k}} = \frac{a}{2k}\left(3H + \frac{\dot{\epsilon}}{\epsilon}\right) \tag{4.23}$$

is the damping ratio. The damping ratio can be written in the more insightful form

$$d_{\vec{k}} = \frac{aH}{k}\left(\frac{3}{2} + \frac{\eta}{2}\right). \tag{4.24}$$

Directly from this it can be seen that the system is underdamped ($d < 1$) inside the horizon, where $k \gtrsim aH$, and overdamped ($d > 1$) outside the horizon, where $k \lesssim aH$, which corresponds to the fact that the modes freeze out.

Rewriting to Bessel's equation

Unfortunately, both $d_{\vec{k}}$ and $\omega_{\vec{k}}$ depend on time, but by going to conformal time the time-dependence from $\omega_{\vec{k}}$ can be removed; the differential equation (eq. (4.20)) can be written as:

$$\mathrm{L}\,u = 0, \tag{4.25}$$

$$\frac{1}{a^2}\,\mathrm{L} := \frac{\mathrm{d}^2}{\mathrm{d}t^2} + H(3 + \eta)\frac{\mathrm{d}}{\mathrm{d}t} + \frac{\vec{k}^2}{a^2}, \tag{4.26}$$

and in conformal time, with

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} = \frac{1}{a}\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{1}{a}\frac{\mathrm{d}}{\mathrm{d}\tau}\right) = \frac{1}{a^2}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2} - \frac{a'}{a^3}\frac{\mathrm{d}}{\mathrm{d}\tau} = \frac{1}{a^2}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2} - \frac{H}{a}\frac{\mathrm{d}}{\mathrm{d}\tau},$$

the expression for L is

$$\mathrm{L} = \frac{\mathrm{d}^2}{\mathrm{d}\tau^2} + aH(2 + \eta)\frac{\mathrm{d}}{\mathrm{d}\tau} + \vec{k}^2. \tag{4.27}$$

Next, aH can be expressed in terms of τ and slow-roll parameters (see appendix A):

$$\tau = \frac{-1}{aH}\left(1 + \epsilon + \epsilon^2 + \epsilon\eta + \epsilon^3 + 2\epsilon^2\eta + \epsilon\eta^2 + \epsilon\eta\xi + \mathcal{O}(\epsilon^4)\right). \tag{4.28}$$

Using the abbreviation

$$\rho := (2 + \eta)(1 + \epsilon + \epsilon\eta + \epsilon^2 + \mathcal{O}(\epsilon^3)), \tag{4.29}$$

the operator is written as

$$\mathrm{L} = \frac{\mathrm{d}^2}{\mathrm{d}\tau^2} - \frac{\rho}{\tau}\frac{\mathrm{d}}{\mathrm{d}\tau} + \vec{k}^2. \tag{4.30}$$

If ρ is a constant (does not depend on τ), then L can be reduced to Bessel's differential equation operator[25],

$$\mathrm{B}_\alpha = x^2\frac{\mathrm{d}^2}{\mathrm{d}x^2} + x\frac{\mathrm{d}}{\mathrm{d}x} + (x^2 - \alpha^2), \tag{4.31}$$

where α is a complex parameter. To go from L to B_α, the coefficient in front of the first derivative needs to become 1. This is achieved by considering a specific (well-guessed) form for the solution:

$$\mathrm{L}\,\tau^n y(\tau) = \left[\tau^n \frac{\mathrm{d}2}{\mathrm{d}\tau^2} + (2n - \rho)\tau^{n-1}\frac{\mathrm{d}}{\mathrm{d}\tau} + \left(n(n-1) - \rho n + \vec{k}^2\tau^2 \right)\tau^{n-2} \right] y(\tau). \tag{4.32}$$

From this it is seen that

$$n = \frac{\rho + 1}{2}, \quad \text{for } \rho \text{ constant,} \tag{4.33}$$

as it was also assumed in the calculation that n is a constant. Therefore, to use the solution the assumption needs to be made that ρ does not depend on time. To get the correct powers of τ, multiply the entire expression by τ^{2-n}:

$$\tau^{2-n}\,\mathrm{L}\,\tau^n y(\tau) = \left[\tau^2\frac{\mathrm{d}2}{\mathrm{d}\tau^2} + \tau\frac{\mathrm{d}}{\mathrm{d}\tau} + \left(n(n-1) - \rho n + \vec{k}^2\tau^2 \right) \right] y(\tau). \tag{4.34}$$

The final step is noting that all the differential terms are invariant under $\tau \to \lambda\tau$, where $\lambda \in \mathbb{C}$, which allows changing only the non-differential term to become of the form $\tau^2 - \alpha^2$, via

$$x = k\tau, \tag{4.35}$$

such that

$$\tau^{2-n}\,\mathrm{L}\,\tau^n y(\tau) = \mathrm{B}_{\pm\mathrm{n}}(y \circ (x \mapsto x/k))\,(k\tau), \tag{4.36}$$

$$n = \frac{\rho + 1}{2}. \tag{4.37}$$

All terms in L are invariant under $\tau \to -\tau$. Therefore, a solution $y(x)$ of $B_\mathrm{n}\,y(x) = 0$, gives a solution $u_k(\tau) = \tau^n y(\pm k\tau)$ of $\mathrm{L}\,u_k(\tau) = 0$. As $k\tau < 0$, it is natural to choose the $-$ sign solution for u_k.

Solutions to equations of motion

The set of independent solutions to the Bessel's equation depends on the value of α; here,

$$\pm\alpha = \frac{3}{2} + \mathcal{O}(\epsilon), \tag{4.38}$$

so a preferred solution set is one that is independent around $\alpha = +3/2$. Furthermore, the domain of interest is $-\infty < \tau < 0$, so the solutions should be regular for this domain. One choice is to work with the two Bessel functions of the first kind, J_α and $J_{-\alpha}$, which are independent if α is not an integer, but $J_\alpha(x)$ does not diverge as $x \to 0$, so it cannot compensate for the factor τ^n in front. Another choice would be to use one or two of the Bessel functions of the second kind, Y_α and $Y_{-\alpha}$. Finally, a third choice, used here because it is most convenient to work with, are the two Hankel functions, $H_\alpha^{(1)}$ and $H_\alpha^{(2)}$, also known as the Bessel functions of the third kind. The Hankel functions have the property that they are convenient to work with theoretically and are always linearly independent, for any value of α. They diverge at $\tau = 0$ in such a way that the combination $\tau^n H_n^{(*)}(-k\tau)$ is regular.

4.1.5 Choice of vacuum

The solutions $u(\vec{k}, \tau)$ only depend on the norm, k, so we will write them as $u(k, \tau)$. Any linear combination of

$$\{\tau^n H_n^{(1)}(-k\tau),\ \tau^n H_n^{(2)}(-k\tau)\} \tag{4.39}$$

on the principal domain, $-\infty < \tau < 0$, gives a solution to the equations of motion, so the initial conditions must be used to select a particular combination. When the mode is well within the horizon, that is $k \gg aH$, and we consider a period much smaller than a Hubble time, the mode should effectively feel the Minkowski vacuum. The solution $\tau^n H_n^{(1)}(-k\tau)$ goes to the flat Minkowski solution in this limit, therefore this is the preferred solution; the operator that pairs to this mode is the annihilation operator that annihilates the vacuum, $|0\rangle$, so $a_{\vec{k}} |0\rangle = 0$. This choice of vacuum is called the Bunch-Davies vacuum.

The Hankel functions have a closed form for integer and half-integer α, so taking $n = 3/2$ gives a solution that is valid up to corrections of first order in slow-roll, which will be denoted by $u(k, \tau)$:[9]

$$u(k, \tau) := \frac{iH}{\sqrt{4\epsilon k^3}}(1 + ik\tau)e^{-ik\tau}. \tag{4.40}$$

The mode u only depends on the norm k, so the first argument is never negative. With this choice of mode and vacuum, the scalar perturbations are quantised as

$$\hat{\mathcal{R}}_{\mathrm{I}}(\vec{k}, \tau) = u(k, \tau)\hat{a}_{\vec{k}} + u^*(k, \tau)\hat{a}^{\dagger}_{-\vec{k}}. \tag{4.41}$$

The same quantisation applies to the field theory for $\mathcal{Q}_{\mathrm{I}}$, because, as mentioned, it has the same H_0. Finally, when going from the classical theory to the quantum theory it is typical to take all operators to be normal ordered. We shall remark on this more during the calculations, when situations are encountered that require normal ordering.

The choice of vacuum is correct for the non-interaction part of the system, but it fails to be the vacuum under the in-in formalism time evolution of the state, such that, as it stands, full vacuum expectation values do not vanish. The solution is to use the Hartle-Hawking prescription of the vacuum[20], which amounts to keeping the Bunch-Davies vacuum and deforming the contour of the integrals in the time-ordered exponentials so as to provide exponential damping for $t \to -\infty$. We will touch upon this again and with more detail in sections 4.2.1 and 4.4.4, where the calculations of expectation values in the in-in formalism are performed.

4.1.6 Wick's theorem

To work out vacuum expectation values, Wick's theorem is an invaluable tool. A contraction between two operators $\hat{A}$ and $\hat{B}$, for any ordering $: :$, is defined as

$$\hat{A}^\bullet \hat{B}^\bullet := \hat{A}\hat{B} - :\hat{A}\hat{B}:. \tag{4.42}$$

Using this with normal ordering, any product of fields can be rewritten in terms of normal ordered products and contractions. The advantage of using normal ordered products is that for products that involve creation and annihilation

operators, the ordering places annihilation operators to the right of creation operators and when sandwiched between vacuum states any such normal ordered product is directly seen to vanish. The contractions are often easy to work with as well, to the point of usually not being operators but (complex) numbers. For fields, like $\mathcal{R}$ and $\mathcal{Q}$, that are built up as

$$\hat{A} = \hat{A}^+ + \hat{A}^-, \tag{4.43}$$

where the positive part is associated with the creation operator and negative part with the annihilation operator, the contraction becomes

$$\hat{A}^\bullet \hat{B}^\bullet = \left[\hat{A}^-, \hat{B}^+\right]. \tag{4.44}$$

From this, a contraction between $\mathcal{Q}$ fields is a function (not an operator):

$$\bullet(k_1, \tau; k_2, \tau') := \mathcal{Q}(k_1, \tau)^\bullet \mathcal{Q}(k_2, \tau')^\bullet = (2\pi)^3 u(k_1, \tau) u^*(k_2, \tau') \delta^{(3)}(k_1 + k_2). \tag{4.45}$$

Wick's theorem will be used to work out expectation values of products of operators to products of normal ordered operators and contractions.

4.2 Calculation of (bi)spectrum

The goal of this section is to show how to calculate the (power) spectrum and bispectrum of the canonical scalar field. In the previous section the field theory was quantised and in the chapter before that the observables, such as the power spectrum and bispectrum, have been discussed. These results were all at leading order in slow-roll; in this chapter the calculation of a non-leading term for the bispectrum is exemplified, which we will refer to as the $\epsilon\dot{\eta}$-term (in S_3). This term has not been calculated before in literature, as far as we have been able to find.

The starting point is the solution for $\mathcal{R}_I$ and $\mathcal{Q}_I$ (eq. 4.41), the Wick contraction between two $\mathcal{Q}_I$'s (eq. (4.45)), and the S_3 action for $\mathcal{Q}$ (eq. 3.41 with field shift),

$$
S_3 = \int \mathrm{d}^4 x \bigg(a^3 \epsilon^2 \mathcal{Q}\dot{\mathcal{Q}}^2 + a\epsilon^2 \mathcal{Q}(\partial\mathcal{Q})^2 - 2a\epsilon^2 \dot{\mathcal{Q}}(\partial\mathcal{Q})(\partial\chi)
$$
$$
+ \frac{a^3\epsilon\dot{\eta}}{2}\mathcal{Q}^2\dot{\mathcal{Q}} + \frac{\epsilon^3}{2a}(\partial\mathcal{Q})(\partial\chi)\partial^2\chi + \frac{\epsilon^3}{4a}(\partial^2\mathcal{Q})(\partial\chi)^2 \bigg),
$$

along with the expressions for calculating the power spectrum (eq. 4.9) and the bispectrum (eq. 4.10):

$$
\langle \mathcal{R}(k_1)\mathcal{R}(k_2)\rangle(t) = \langle 0|_{\mathcal{R}} \, \mathcal{R}_I(k_1,t)\mathcal{R}_I(k_2,t)\,|0\rangle_{\mathcal{R}} + \mathcal{O}\big(\mathcal{R}_I^6\big),
$$
$$
\langle \mathcal{R}(k_1)\mathcal{R}(k_2)\mathcal{R}(k_3)\rangle(t) = \frac{\eta}{4}\langle 0| \, [\mathcal{Q}_I \star \mathcal{Q}_I](k_1,t)\mathcal{Q}_I(k_2,t)\mathcal{Q}_I(k_3,t)\,|0\rangle + 2 \text{ perm.}
$$
$$
+ i\int_{-\infty}^{t} \mathrm{d}t' \, \langle 0| \, [H_{I,3}(t'), \mathcal{Q}_I(k_1,t)\mathcal{Q}_I(k_2,t)\mathcal{Q}_I(k_3,t)]\,|0\rangle
$$
$$
+ \mathcal{O}\big(\mathcal{R}_I^8\big).
$$

For the power spectrum, we see that the interaction terms, which represent the non-Gaussianities, do not come in until much higher order, as was expected from the f_{NL}-toy model, eq. (3.50), where the second moment is not affected by f_{NL} when $f_{\mathrm{NL}}^2 P_{\mathcal{R}} \ll 1$ and from the results in that chapter we know that for the scalar field $f_{\mathrm{NL}}^2 P_{\mathcal{R}} \sim 10^{-9}\epsilon^2$, which is indeed much smaller than one. For the bispectrum calculation, we will focus on the $\epsilon\dot{\eta}$-term in S_3, $(1/2)a^3\epsilon\dot{\eta}\mathcal{Q}^2\dot{\mathcal{Q}}$, as a sufficiently rich example.

4.2.1 Calculation in slow-roll

The exact solutions for the inflationary background (a, H, ϵ, η, etc.) for a general potential is not known. It is therefore impossible to perform the calculation of the (bi)spectrum exactly. There is, however, a method of calculation that can be used with generic background quantities, assuming only slow-roll, that minimises the error in the calculation from not knowing the exact evolution of these quantities. This method was first explained by Maldacena [20]. There are two related problems that come up here; the first was already mentioned when having to choose a vacuum when quantising the theory:

1. **Problem:** we want to calculate expectation values in the vacuum of the interacting theory, not of the free theory. **Solution:** for integrals, deform

the contour to provide exponential damping for early times, such that early contributions cancel out.

2. **Problem:** the background quantities are unknown, so we don't know how to take their limits or integrate them. **Solution:** the relevant quantities are taken to be slow-roll and together with freezing out of the modes, the background quantities can be evaluated at horizon crossing. The relative error this introduces is $\ll \mathcal{O}(1)$, which holds true both in limits and in performing integrals over time.

It is worth discussing these problems and their solutions separately for terms that involve integrals and that don't involve integrals. Throughout this chapter, a star will be used to denote quantities that are evaluated at horizon crossing, e.g. $\epsilon_* = \epsilon(t_k)$. This is related to, but not the same as the previously used star to denote quantities evaluated at the pivot, but if a quantity is scale-invariant then these coincide.

Terms not involving time integrals

For expectation value calculations (or terms thereof) that do not involve an integral over time, at the end of the calculation the result is a time-dependent expectation value. We want to evolve this expectation value to the infinite future, or at least until after freeze-out, as this gives the conserved expectation value outside the horizon. The issue is that we do not know how to take the limit, because the background quantities are unknown. As explained, a mode with wavenumber k freezes out as $k \ll aH$, or $-k\tau \ll 1$ in terms of conformal time. When these terms consist only of a product of $\mathcal{R}$ fields (or $\mathcal{Q}$ fields), such as when calculating expectation values of the fields, it follows that the expectation value freezes out as well. The closed form of the modes used is only valid when the slow-roll parameters (and H) can be taken to be constants and taking the limit for $\tau \uparrow 0$ is in principle not a valid operation, because the slow-roll parameters will change significantly over infinite time. Luckily, the freeze-out means that the limit only needs to be taken until the mode (or modes) are outside of the horizon, which happens exponentially fast, after which their value is constant up to corrections that are of fractional order $\mathcal{O}(-k\tau) \ll \mathcal{O}(1)$ for a single mode. For multiple modes, k_i, the order from the single mode is multiplied by the factor $H(t_\mathrm{S} - t_\mathrm{L}) \lesssim \mathcal{O}(10)$, where t_S (t_L) is the time that the shortest (longest) of the modes crosses the horizon and is for extreme squeezing[16] $k_\mathrm{L}/k_\mathrm{S} \lesssim 10^{-7}$ of order 7 at most, therefore this does not change the order estimate.
The procedure to get the final result is therefore to take the time-dependent result, evaluate all background quantities on the average time of horizon crossing for the mode, k, or modes, k_i, and then take the limit for $\tau \uparrow 0$. The relative error this introduces is $\ll \mathcal{O}(1)$, both for a single mode and multiple modes in all realistic momentum configurations.

Terms involving time integrals

For expectation value calculations (or terms thereof) that involve integrals over time, from the infinite past to now, there are the two issues mentioned. The first is that we want to calculate expectation values for the vacuum state of the interacting theory rather than the free theory. The second is similar to the issue

with expectation value calculations of non-integral terms, where the absence of an exact solution for the background quantities means that there are unknown functions in the integral, thus it cannot be evaluated. Both can be 'solved' by the same procedure, which allows the approximate calculation to be performed whilst minimising the error.

The time integral over perturbations is split into three logical time domains: evolution well inside the horizon, evolution near horizon crossing, and evolution well outside the horizon. For evolution well inside the horizon, that is at early times[1], the first issue needs to be tackled and this can be done using the Hartle-Hawking prescription for the vacuum, which translates to a choice of complex contour for the (normally) real contour time integral, such that early time fluctuations are exponentially suppressed. The exact contour used to achieve this suppression will be seen when performing the calculations. The contribution, therefore, from early times will be zero. For evolution near horizon crossing, the slow-roll parameters and H are approximately constant and can be taken as such, where again the relative error introduced by this is $\ll \mathcal{O}(1)$. For evolution well outside the horizon, the mode freezes out so ignoring this tail contribution contributes a relative error of $\ll \mathcal{O}(1)$.

Actually, this last part is not entirely true. The perturbations freeze out, so the expectation value of moments of the perturbations goes rapidly to a constant, and therefore the sum total of all terms at a given order in slow-roll go rapidly to a constant. However, when singling out any one term, it might be that it has a divergence, which is then necessarily cancelled by the divergence from a different term; these two terms must contain the exact same product of slow-roll parameters to be able to cancel out.

4.3 Power spectrum

The power spectrum is simply given by the expectation value of the operator $\mathcal{R}(k,\tau)\mathcal{R}(k',\tau)$ on the vacuum state. Using Wick's theorem, this is equal to the contraction between the two $\mathcal{R}$ fields (eq. (4.45)), such that

$$\langle \mathcal{R}(k_1)\mathcal{R}(k_2)\rangle(\tau) = (2\pi)^3 u(k,\tau)u^*(k',\tau)\delta^{(3)}(k+k'), \qquad (4.46)$$

and from this and eq. (3.67) the power spectrum is given by

$$P_{\mathcal{R}}(k;\tau) = (2\pi)^{-2}2k^3 u(k,\tau)u^*(k,\tau). \qquad (4.47)$$

The products of the mode functions are

$$u(k,\tau)u^*(k,\tau') = \frac{H^2}{4\epsilon k^3}(1+ik\tau)(1-ik\tau')e^{-ik(\tau-\tau')}. \qquad (4.48)$$

To take the limit for $\tau \uparrow 0$, the slow-roll parameters (and H) are taken at the time of horizon crossing for mode k, as discussed. If also $\tau = \tau'$ then this reduces to

$$u(k,\tau)u^*(k,\tau) = \frac{H_*^2}{4\epsilon_* k^3}(1-k^2\tau^2), \qquad (4.49)$$

---

[1]Early times, relative to the given mode: $-k\tau \gg 1$

which converges nicely in the limit $\tau \uparrow 0$. The result for the power spectrum of the canonical scalar field is therefore given by

$$P_{\mathcal{R}}^* = \frac{H_*^2}{8\pi^2 \epsilon_*}, \tag{4.50}$$

which agrees with the result from literature, eq. (3.70), when units of M_{Pl} are reintroduced.

4.4 Bispectrum

From this point onwards, the subscript 'I' will be dropped from the interaction picture fields; any field that is time dependent in this section is necessarily in the interaction picture, so there is no confusion caused by simplifying the notation. We will work in conformal time for fields for the rest of this section, but the slow-roll parameters are kept in terms of (normal) time, so as to apply the freeze-out reasoning that allows the approximate calculation, explained in section 4.2.1. The goal is to calculate the bispectrum up to second order in slow-roll from eq. (4.10), now in conformal time, with derivatives to conformal time denoted by primes, $\mathcal{Q}' := \partial_\tau \mathcal{Q}$,

$$\begin{aligned}
\langle \mathcal{R}(k_1)\mathcal{R}(k_2)\mathcal{R}(k_3) \rangle(\tau) &= \frac{\eta}{4} \langle 0| \left[\mathcal{Q} \star \mathcal{Q} \right](k_1, \tau)\mathcal{Q}(k_2, \tau)\mathcal{Q}(k_3, \tau) |0\rangle + 2 \text{ perm.} \\
&+ i \int_{-\infty}^{\tau} \mathrm{d}\tau' a(\tau') \langle 0| \left[H_{\text{I},3}(\tau'), \mathcal{Q}(k_1, \tau)\mathcal{Q}(k_2, \tau)\mathcal{Q}(k_3, \tau) \right] |0\rangle \\
&+ \mathcal{O}\!\left(\mathcal{R}^8, \epsilon^3\right),
\end{aligned} \tag{4.51}$$

with

$$\begin{aligned}
H_{\text{I},3} = -\int \mathrm{d}^3 x \bigg(& a\epsilon^2 \mathcal{Q}\mathcal{Q}'^2 + a\epsilon^2 \mathcal{Q}(\partial\mathcal{Q})^2 - 2\epsilon^2 \mathcal{Q}'(\partial\mathcal{Q})(\partial\chi) \\
& + \frac{a^2 \epsilon\dot{\eta}}{2} \mathcal{Q}^2\mathcal{Q}' + \frac{\epsilon^3}{2a}(\partial\mathcal{Q})(\partial\chi)\partial^2\chi + \frac{\epsilon^3}{4a}(\partial^2\mathcal{Q})(\partial\chi)^2 \bigg),
\end{aligned} \tag{4.52}$$

$$\chi = a\partial^{-2}\mathcal{Q}'. \tag{4.53}$$

Being explicit about all the arguments to the functions in the Hamiltonian gives the precise but also more cluttered expression, with t' being short for $t' = t(\tau')$,

$$\begin{aligned}
H_{\text{I},3}(\tau') = -\int \mathrm{d}^3\vec{x} \bigg(& a(\tau')\epsilon(t')^2 \mathcal{Q}(\vec{x}, \tau')\mathcal{Q}'(\vec{x}, \tau')^2 + a(\tau')\epsilon(t')^2 \mathcal{Q}(\vec{x}, \tau')(\vec{\partial}\mathcal{Q}(\vec{x}, \tau'))^2 \\
& - 2\epsilon(t')^2 \mathcal{Q}'(\vec{x}, \tau')(\vec{\partial}\mathcal{Q}(\vec{x}, \tau'))(\vec{\partial}\chi(\vec{x}, \tau')) + \frac{a(\tau')^2\epsilon(t')\dot{\eta}(t')}{2} \mathcal{Q}(\vec{x}, \tau')^2\mathcal{Q}'(\vec{x}, \tau') \\
& + \frac{\epsilon(t')^3}{2a(\tau')}(\vec{\partial}\mathcal{Q}(\vec{x}, \tau'))(\vec{\partial}\chi(\vec{x}, \tau'))\vec{\partial}^2\chi(\vec{x}, \tau') + \frac{\epsilon(t')^3}{4a(\tau')}(\vec{\partial}^2\mathcal{Q}(\vec{x}, \tau'))(\vec{\partial}\chi(\vec{x}, \tau'))^2 \bigg),
\end{aligned} \tag{4.54}$$

$$\chi(\vec{x}, \tau') = a(\tau')\vec{\partial}^{-2}\mathcal{Q}'(\vec{x}, \tau'). \tag{4.55}$$

The principal calculation can be split into two parts. The first part is to calculate the vacuum expectation value of a product of fields, after which the result of the

first line in eq. (4.51) is obtained. The second part is to calculate the integral over fields as in the second line in eq. (4.51); vacuum expectation values also need to be calculated for this, using the same technique as in the first part. The calculation will therefore proceed in this order, first calculating expectation values and then moving on to the integral.

Maldacena [20] was the first to have calculated the bispectrum of the canonical scalar field using the described method and does so at leading order, $\mathcal{O}(\epsilon^2)$, by performing another field shift to further simplify the ϵ^2-terms in the S_3 action and ignoring any terms that are higher-order in slow roll. The literature result quoted in eq. (3.90) is the result of this calculation. We will not use this additional field shift; we will start by showing the calculation for a general term in S_3, until we can go no further by keeping it general, after which the $\epsilon\dot{\eta}$-term will be considered as an example of how to perform the calculation for a term in S_3. This example is sufficiently rich that it can be followed to calculate any term in S_3 and also could be used to recover the result from literature (by using the additional field shift, or calculating each term on the first line of S_3). Furthermore, this term gives problems with convergence not found in any of the other terms in S_3, which turn out to be caused (or rather: compensated for) by neglecting boundary terms in time when deriving S_3. The conclusion will be that when calculating the bispectrum up to higher order in slow-roll, care should be taken to include boundary terms in time, which have been neglected in deriving S_3. The significance of boundary terms in time was also noted by Arroja and Tanaka [3].

4.4.1 Non-integral term

Applying Wick's theorem to the first expectation value (that is, the first line of eq. (4.51) and ignoring for a moment the two permutations) means all terms that are not fully contracted vanish between the vacuum and what remains is

$$\frac{\eta}{4(2\pi)^3} \int \mathrm{d}^3q \, \langle 0| \, \mathcal{Q}(q,\tau)\mathcal{Q}(k_1 - q,\tau)\mathcal{Q}(k_2,\tau)\mathcal{Q}(k_3,\tau) \, |0\rangle =$$
$$\frac{\eta}{4(2\pi)^3} \int \mathrm{d}^3q \Bigg[\bullet(q,\tau;k_1 - q,\tau) \bullet(k_2,\tau;k_3,\tau) + \bullet(q,\tau;k_2,\tau) \bullet(k_1 - q,\tau;k_3,\tau)$$
$$+ \bullet(q,\tau;k_3,\tau) \bullet(k_1 - q,\tau;k_2,\tau) \Bigg].$$
$$(4.56)$$

Filling in the contractions and then performing the integral over the internal momentum, q, which works on the delta functions gives for the three terms

$$\frac{\eta}{4(2\pi)^3} \int \mathrm{d}^3q \, \bullet(q,\tau;k_1 - q,\tau) \bullet(k_2,\tau;k_3,\tau) =$$
$$\frac{\eta}{4} \Big(\int \frac{\mathrm{d}^3q}{(2\pi)^3} \, u(q,\tau)u^*(q,\tau) \Big) u(k_2,\tau)u^*(k_3,\tau)(2\pi)^3\delta^{(3)}(k_1)(2\pi)^3\delta^{(3)}(k_2 + k_3),$$
$$(4.57)$$

$$\frac{\eta}{4(2\pi)^3} \int \mathrm{d}^3q \,\bullet(q,\tau;k_2,\tau)\,\bullet(k_1 - q,\tau;k_3,\tau) =$$

$$\frac{\eta}{4} u(k_2,\tau)u^*(k_2,\tau)u(k_1 + k_2,\tau)u^*(k_3,\tau)(2\pi)^3\delta^{(3)}(k_1 + k_2 + k_3)$$

$$= \frac{\eta}{4} u(k_2,\tau)u^*(k_2,\tau)u(k_1 + k_2,\tau)u^*(k_1 + k_2,\tau)(2\pi)^3\delta^{(3)}(k_1 + k_2 + k_3),$$
$$(4.58)$$

$$\frac{\eta}{4(2\pi)^3} \int \mathrm{d}^3q \,\bullet(q,\tau;k_3,\tau)\,\bullet(k_1 - q,\tau;k_2,\tau) =$$

$$\frac{\eta}{4} u(k_2,\tau)u^*(k_2,\tau)u(k_1 + k_2,\tau)u^*(k_1 + k_2,\tau)(2\pi)^3\delta^{(3)}(k_1 + k_2 + k_3).$$
$$(4.59)$$

Note how eqs. (4.58) and (4.59) are the same. This could have been seen immediately from eq. (4.56) by the symmetry $q \to k_1 - q$ of the internal momentum integral. The first term, eq. (4.57), has two delta-conditions on the momentum configuration, reducing this contribution from a two-dimensional to a one-dimensional 'shape'. In addition, this contribution only happens when one of the momenta is exactly 0. This is not detectable and arguably not physical. Throwing away these terms is equivalent to taking normal ordered operator products. In all three results the final momentum conservation $(2\pi)^3\delta^{(3)}(k_1 + k_2 + k_3)$ arises, as expected.

Result

The expression for the first expectation value has been reduced to

$$\frac{\eta}{2} u(k_2,\tau)u^*(k_2,\tau)u(k_1 + k_2,\tau)u^*(k_1 + k_2,\tau)(2\pi)^3\delta^{(3)}(k_1 + k_2 + k_3). \quad (4.60)$$

The products of two mode functions at the same time, as seen when calculating the power spectrum, is given by

$$u(k,\tau)u^*(k,\tau) = \frac{H_*^2}{4\epsilon_* k^3}(1 - k^2\tau^2). \quad (4.61)$$

The time-dependent result is therefore

$$\frac{\eta}{2} \frac{H_*^4}{16\epsilon_*^2 k_2^3 k_3^3}(1 - k_2^2\tau^2)(1 - k_3^2\tau^2)(2\pi)^3\delta^{(3)}(k_1 + k_2 + k_3) + 2 \text{ perm.} \quad (4.62)$$

As $\tau \uparrow 0$, this converges nicely to

$$\frac{\eta}{2} \frac{H_*^4}{16\epsilon_*^2 k_1^3 k_2^3 k_3^3}(k_1^3 + k_2^3 + k_3^3)(2\pi)^3\delta^{(3)}(k_1 + k_2 + k_3). \quad (4.63)$$

Relating this to the shape function via its definition, eq. (3.68), and using the result for $P_{\mathcal{R}}$, eq. (3.70) or (4.50), now evaluated at horizon crossing rather than at a pivot (so, we're cheating a little bit here in terms of full scale dependence), gives

$$S(k_1, k_2, k_3) = \frac{\eta}{8} \frac{k_1^3 + k_2^3 + k_3^3}{k_1 k_2 k_3}. \quad (4.64)$$

This is the local shape, with $f_{\mathrm{NL}}^{\mathrm{local}} = \frac{5}{18}\eta$, and the result matches that from literature, eq. (3.90).

4.4.2 Integral terms

The integral that needs to be calculated from eq. (4.51), now written in conformal time, is

$$i \int_{-\infty}^{\tau} \mathrm{d}\tau' a(\tau') \langle 0| [H_{\mathrm{I},3}(\tau'), \mathcal{Q}(k_1,\tau)\mathcal{Q}(k_2,\tau)\mathcal{Q}(k_3,\tau)] |0\rangle. \qquad (4.65)$$

Each term in $H_{\mathrm{I},3}(\tau')$ is of third order in $\mathcal{Q}$. Each $\mathcal{Q}$ can come with with zero or one time derivatives on it and with minus one, zero, one, or two spatial derivatives[2] on it. Schematically, the terms inside $H_{\mathrm{I},3}(\tau')$ can be denoted by

$$\beta(\tau') \int \mathrm{d}^3x \, D_1 \mathcal{Q}(x,\tau') D_2 \mathcal{Q}(x,\tau') D_3 \mathcal{Q}(x,\tau'), \qquad (4.66)$$

with D_i differential operators of powers of space and time, and $\beta(\tau')$ a time-dependent prefactor. Note that there are no x-dependent prefactors, because the only dependence on x comes directly from $\mathcal{Q}$. Going to Fourier space for spatial coordinates, the products of $\mathcal{Q}$ fields become convolutions:

$$\frac{\beta(\tau')}{(2\pi)^9} \int \mathrm{d}^3x \int \mathrm{d}^3k \int \mathrm{d}^3p \int \mathrm{d}^3q \, \hat{D}_1 \mathcal{Q}(k-q-p,\tau') \hat{D}_2 \mathcal{Q}(q,\tau') \hat{D}_3 \mathcal{Q}(p,\tau') e^{ikx},$$
$$(4.67)$$

where the Fourier transformed differential operators, $\hat{D}_i$, are related to D_i by replacing

$$\vec{\partial}_i^n \to \left(i\vec{k}_i\right)^n, \quad \forall n \in \mathbb{Z}$$
$$\partial_{\tau'} \to \partial_{\tau'}.$$

Note that for odd negative n the spatial substitution should be interpreted as $(ik)^n \hat{k}$. For convenience of notation, the vector arrows on momenta are often omitted when it's clear from the context what is meant. The integral over x gives

$$\int \mathrm{d}^3x \, e^{ikx} = (2\pi)^3 \delta^{(3)}(k). \qquad (4.68)$$

The contraction eq. (4.45) can be extended to the situation with $\hat{D}$ operators working on the fields as

$$\hat{D}_i \mathcal{Q}(k_i,\tau')^\bullet \hat{D}_j \mathcal{Q}(k_j,\tau')^\bullet = (2\pi)^3 \hat{D}_i u(k_i,\tau') \hat{D}_j u^*(k_j,\tau') \, \delta^{(3)}(k_i+k_j). \quad (4.69)$$

4.4.3 Momentum conservation

We are now in a position to work out the momentum conservation in a very general manner. Suppose that we have a term from H_{I} with n fields and we are looking at the expectation value of a product of m fields. First, an important observation is that the convolution of the term inside H_{I} can be chosen so as to have exactly one internal momentum on each field, except for on one field which carries all internal momenta, because of associativity of $\star$, giving:

$$\frac{\beta(\tau')}{(2\pi)^{3n}} \int \mathrm{d}^3x \int \mathrm{d}^3q_1 \, e^{iq_1 x} \int \mathrm{d}^3q_2 \cdots \mathrm{d}^3q_n \Big[$$
$$\hat{D}_1 \mathcal{Q}\Big(q_1 - \sum_{i=2}^{n} q_i, \tau'\Big) \hat{D}_2 \mathcal{Q}(q_2,\tau') \cdots \hat{D}_n \mathcal{Q}(q_n,\tau')\Big]. \qquad (4.70)$$

---

[2]The inverse laplacian, ∂^{-2}, is counted as minus two spatial derivatives.

For this part, we find it convenient to denote terms by the momentum they carry, as each $\mathcal{Q}$ comes with its own unique momentum.

A note on total derivatives

We can make one powerful observation now: if, instead of only having separate derivatives D_i on each field, there is a total spatial derivative,

$$\beta(\tau') \int \mathrm{d}^3 x \, \frac{\partial}{\partial x_i} \Big(D_1 \mathcal{Q}(x, \tau') D_2 \mathcal{Q}(x, \tau') D_3 \mathcal{Q}(x, \tau') \Big), \qquad (4.71)$$

then the total derivative of x becomes a total pre-factor of q_1 in:

$$\frac{\beta(\tau')}{(2\pi)^{3n}} \int \mathrm{d}^3 x \int \mathrm{d}^3 q_1 \, i q_1 e^{i q_1 x} \int \mathrm{d}^3 q_2 \cdots \mathrm{d}^3 q_n \Big[\\ \hat{D}_1 \mathcal{Q}\Big(q_1 - \sum_{i=2}^{n} q_i, \tau'\Big) \hat{D}_2 \mathcal{Q}(q_2, \tau') \cdots \hat{D}_n \mathcal{Q}(q_n, \tau') \Big]. \qquad (4.72)$$

and then performing the integral over $\mathrm{d}^3 x$ gives $\delta^{(3)}(q_1)$, therefore the entire term is seen to be 0 (assuming the $q_1 = 0$ Fourier coëfficient is finite). The conclusion is that any spatial boundary terms in the (full) action do not contribute to the (bi)spectrum at the one-integral level, which is a nice consistency check; spatial boundary terms should vanish anyway from requiring the fields to be square-integrable quantities such that Fourier theory can be applied with finite Fourier coëfficients (the exception being the Fourier delta-function).

Contractions

A contraction between any two terms, p_i and p_j, causes one momentum to be substituted for the other, $p_i = -p_j$, because of the resulting delta-function $\delta^{(3)}(p_i + p_j)$. If a contraction is made between any external momenta, k_i and k_j, then these momenta will be constrained to be exactly opposite each other, by $\delta^{(3)}(k_i + k_j)$, resulting in a momentum configuration of dimension one less, which does not give a physical contribution. Another way to look at is is that the operators should be taken to be normal ordered, as has been mentioned, in which case the contraction between two external momenta is a contraction within a normal ordered group, thus this gives exactly zero. Therefore, each external momentum should be contracted with an internal momentum. This means that each contraction contains at least one internal momentum; if p_i is the internal momentum then the contraction becomes

$$\frac{1}{(2\pi)^3} \int \mathrm{d}^3 p_i \, (\ldots) \hat{D}_i \mathcal{Q}(p_i)^\bullet \hat{D}_j \mathcal{Q}(p_j)^\bullet = (\ldots) \hat{D}_i u(-p_j) \hat{D}_j u^*(p_j) \qquad (4.73)$$

where each p_i is replaced by $-p_j$, also inside the terms in the $(\ldots)$, which represents all other terms not displayed. We have also suppressed writing the time argument, τ or τ'. If p_i is not the internal momentum, then p_j is, and the same formula holds, but then with momenta swapped.

Suppose that $n = m$, such as for the bispectrum calculation being performed, then there is an internal momentum for each external momentum and thus there are only contractions between an internal and an external momentum, and not

between two internal momenta. Then each internal momentum is replaced by the negative of an external momentum, and q_1 is replaced by $\sum k_i$, after which the delta function from eq. (4.68) becomes the momentum conservation,

$$(2\pi)^3 \delta^{(3)}\Big(\sum_{i=1}^m k_i\Big), \tag{4.74}$$

all the momentum integrals have been performed, and each term like eq. (4.66) is reduced to a product of mode functions. The result when internal momenta are to the left of external momenta is:

$$(2\pi)^3 \delta^{(3)}\Big(\sum_{i=1}^m k_i\Big)$$
$$\times \beta(\tau') \sum_{\sigma \in \mathrm{Sym}(\{1,\dots,m\})} \tilde{D}_1 u(k_{\sigma(1)}, \tau') \cdots \tilde{D}_m u(k_{\sigma(m)}, \tau') \cdot u^*(k_1, \tau) \cdots u^*(k_m, \tau), \tag{4.75}$$

and the result when internal momenta are to the right of external momenta is:

$$(2\pi)^3 \delta^{(3)}\Big(\sum_{i=1}^m k_i\Big)$$
$$\times \beta(\tau') u(k_1, \tau) \cdots u(k_m, \tau) \sum_{\sigma \in \mathrm{Sym}(\{1,\dots,m\})} \tilde{D}_1 u^*(k_{\sigma(1)}, \tau') \cdots \tilde{D}_m u^*(k_{\sigma(m)}, \tau'), \tag{4.76}$$

where each occurrence of the internal momentum q_i inside the operator $\hat{D}_i$ has been replaced by $-k_{\sigma(i)}$ to yield $\tilde{D}_i$; note that for q_1, which was taken to be 'special', the replacement should either happen before replacing $q_1 \to q_1 - \sum k_i$, or the full expression appears inside $\hat{D}_1$ needs to be replaced ($q_1 - \sum k_i \to -k_{\sigma(1)}$). Now that the momentum conservation has been established, it's useful to consider stripped correlators for the rest, such that the $(2\pi)^3 \delta^{(3)}(\sum k_i)$ factor is dropped from the remaining calculations.

For each term in $H_{\mathrm{I},m}$ we have the expression that gives its contribution to the m-point function:

$$u(k_1, \tau) \cdots u(k_m, \tau) i \int_{-\infty}^{\tau} \mathrm{d}\tau' a(\tau') \beta(\tau') \sum_{\sigma} \tilde{D}_1 u^*(k_{\sigma(1)}, \tau') \cdots \tilde{D}_m u^*(k_{\sigma(m)}, \tau'),$$

$$- u^*(k_1, \tau) \cdots u^*(k_m, \tau) i \int_{-\infty}^{\tau} \mathrm{d}\tau' a(\tau') \beta(\tau') \sum_{\sigma} \tilde{D}_1 u(k_{\sigma(1)}, \tau') \cdots \tilde{D}_m u(k_{\sigma(m)}, \tau'). \tag{4.77}$$

4.4.4 Example: $\epsilon\dot{\eta}$-contribution

The calculation will now proceed for the stripped correlator for the bispectrum, $m = 3$, and for the $\epsilon\dot{\eta}$-term in $H_{\mathrm{I},3}$, with $\beta = -\frac{1}{2}a^2\epsilon\dot{\eta}$, $D_1 = D_2 = \mathbb{1}$, $D_3 = \partial_{\tau'}$,

such that in this case $\tilde{D}_i = D_i$, and the expression is

$$
\begin{aligned}
-\frac{i\epsilon_*\dot{\eta}_*}{2} & u(k_1,\tau)u(k_2,\tau)u(k_3,\tau)\int_{-\infty}^{\tau}\mathrm{d}\tau' a(\tau')^3 \\
& \times \sum_{\sigma} u^*(k_{\sigma(1)},\tau')u^*(k_{\sigma(2)},\tau')\partial_{\tau'}u^*(k_{\sigma(3)},\tau') \\
+\frac{i\epsilon_*\dot{\eta}_*}{2} & u^*(k_1,\tau)u^*(k_2,\tau)u^*(k_3,\tau)\int_{-\infty}^{\tau}\mathrm{d}\tau' a(\tau')^3 \\
& \times \sum_{\sigma} u(k_{\sigma(1)},\tau')u(k_{\sigma(2)},\tau')\partial_{\tau'}u(k_{\sigma(3)},\tau').
\end{aligned}
\tag{4.78}
$$

The slow-roll parameters have been taken outside of the integral by evaluating them at horizon crossing. For this, we need to add the assumption that $\dot{\eta}$ is a slowly varying function of time; when this cannot be assumed, then the contribution must be calculated by specifying $\dot{\eta}$ as a function of time and calculating the integral with $\dot{\eta}$ kept inside.

Filling in the mode functions

For this term, the symmetry of k reduces the sum over σ: it's only significant which k_i is associated to the mode with the derivative acting on it (D_3). We can calculate the expression for the time derivative associated to k_3 and write '+2 perm.' to denote the three different permutations that arise, for instance, (k_1, k_2, k_3), (k_3, k_1, k_2), and (k_2, k_3, k_1); the σ-permutations between the momenta in the first two slots can be taken into account by an overall factor of 2. Using $a(\tau')H(\tau') = -1/\tau' + \mathcal{O}(\epsilon)$ to write a in terms of τ', eq. (4.40) for the mode function,

$$
u(k,\tau) = \frac{iH}{\sqrt{4\epsilon k^3}}(1+ik\tau)e^{-ik\tau},
$$

and its derivative,

$$
\partial_{\tau'}u(k,\tau') = \frac{iH}{\sqrt{4\epsilon k^3}}k^2\tau' e^{-ik\tau'},
\tag{4.79}
$$

where H and ϵ are taken to be constant for this solution of the mode function[3], evaluating all slow-roll quantities on the horizon (both from $u(\tau)$ and $u(\tau')$) and finally taking the limit for $\tau \uparrow 0$ for the piece before the integral[4] gives for eq. (4.78):

$$
\frac{i\dot{\eta}_* H_*^3 k_3^2}{64\epsilon_*^2 k_1^3 k_2^3 k_3^3}\left[\int_{-\infty}^{\tau}\mathrm{d}\tau'\frac{1}{\tau'^2}(1-ik_1\tau')(1-ik_2\tau')e^{iK\tau'}\right.
$$
$$
\left.-\int_{-\infty}^{\tau}\mathrm{d}\tau'\frac{1}{\tau'^2}(1+ik_1\tau')(1+ik_2\tau')e^{-iK\tau'}\right]
\tag{4.80}
$$

$$
+ \text{2 perm.}
$$

---

[3]It would also be inaccurate to add the derivatives of H and ϵ, because the mode functions used here are themselves approximations which require ϵ and η to be constant to be correct, so any higher order in slow-roll terms would need to be derived directly from the equations of motion.

[4]The limit for $\tau \uparrow 0$ can be done separately for the piece before the integral and for the integral in this case; in general, $\lim a(t)b(t) = \lim a(t) \cdot \lim b(t)$ if both exists.

Hartle-Hawking vacuum

The integrals should be regularised to give strong dampening when $K\tau' \ll -1$, or equivalently when $|K\tau'| \gg 1$ (because $\tau' < 0$), in order to have the expectation value be for the vacuum of the interacting theory, rather than the free theory. This is the Hartle-Hawking prescription for the vacuum, as mentioned, which also validates the procedure of taking the slow-roll parameters outside of the integral. The exponential inside the integral should give exponential damping in this limit, which can be achieved by skewing the integration domain into the imaginary plane via

$$\tau' \to \tau'(1 \pm i\varepsilon), \tag{4.81}$$

where $-$ is for the integral coming from the anti-time-ordered exponential, written in the first line in eq. (4.80), and $+$ for the integral coming from the time-ordered exponential, written in the second line. Taking the limit $\varepsilon \downarrow 0$ after the integral gives for the $+$-contour

$$\lim_{\varepsilon \downarrow 0} \frac{1}{1 + i\varepsilon} \int_{-\infty}^{\tau} \mathrm{d}\tau' \, (1 - \varepsilon k_1 \tau' + i k_1 \tau')(1 - \varepsilon k_2 \tau' + i k_2 \tau') \frac{k_3^2}{\tau'^2} e^{-iK\tau'} e^{-\varepsilon|K\tau'|}, \tag{4.82}$$

which has the desired exponential damping, and the prefactor comes from

$$\frac{\mathrm{d}\tau'}{\tau'^2} \to \frac{\mathrm{d}\tau'}{(1 + i\varepsilon)\tau'^2}. \tag{4.83}$$

Applying this regularisation does not give additional terms and it does not modify existing terms, other than sending any boundary terms on $\tau' = -\infty$ to zero (which was the goal of the regularisation). Therefore, calculating the unregularised expression with neglected boundary terms is equivalent to calculating the regularised integral.

The rest of the calculation will proceed using the unregularised form and boundary terms at $-\infty$ will be discarded. At this point, it can also be seen that the two integrals in eq. (4.80) are complex conjugates, so in fact only one needs to be calculated and regularised, after which the other one directly follows by complex conjugation. Because the first ($-$-contour) is subtracted from the other ($+$-contour), the complex conjugate is not even directly needed, as it is equivalent to taking the imaginary part: $z - z^* = 2i\Im(z)$.

Solving the time integrals

There are three different types of integrals in eq. (4.80) that need to be solved:

$$I_1(\tau) = \int_{-\infty}^{\tau} \mathrm{d}\tau' \, e^{-iK\tau'}, \tag{4.84}$$

$$I_2(\tau) = \int_{-\infty}^{\tau} \mathrm{d}\tau' \, \frac{1}{\tau'} e^{-iK\tau'}, \tag{4.85}$$

$$I_3(\tau) = \int_{-\infty}^{\tau} \mathrm{d}\tau' \, \frac{1}{\tau'^2} e^{-iK\tau'}. \tag{4.86}$$

The first can be solved directly and with regularisation applied becomes:

$$I_1(\tau) = \left[\frac{ie^{-iK\tau'}}{K} \right]_{-\infty}^{\tau} \longrightarrow \frac{ie^{-iK\tau}}{K}. \tag{4.87}$$

The second is proportional to a named integral called the 'Exponential Integral', Ei,

$$\mathrm{Ei}(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t} \, \mathrm{d}t, \tag{4.88}$$

or its complex equivalent,

$$\mathrm{E}_1(z) = \int_{z}^{\infty} \frac{e^{-t}}{t} \, \mathrm{d}t. \tag{4.89}$$

This function has a branch cut, typically chosen to be $(-\infty, 0)$ over which it jumps by $i\pi$ when going from the upper-half plane to the lower-half plane. The integral can be rewritten to match the desired form:

$$iK\tau' = t \iff \tau' = -it/K, \tag{4.90}$$

$$\mathrm{E}_1(z) = -\int_{-i\infty/K}^{-iz/K} \frac{e^{-iK\tau'}}{\tau'} \, \mathrm{d}\tau'. \tag{4.91}$$

The function E_1 is holomorphic on its domain ($\mathbb{C}$ without the branch cut) and therefore the complex equivalent of the fundamental theorem of calculus applies, with E_1 an antiderivative for the integrand, such that

$$I_2(\tau) = \int_{-\infty}^{\tau} \mathrm{d}\tau' \frac{1}{\tau'} e^{-iK\tau'} = -\left(\mathrm{E}_1(iK\tau) - \mathrm{E}_1(-iK\infty) \right). \tag{4.92}$$

The exponential integral has the limit

$$\lim_{z \to \pm i\infty} \mathrm{E}_1(z) = \pm i\pi, \tag{4.93}$$

but this is not even needed here because of the applied regularisation, giving the result

$$I_2(\tau) = -\mathrm{E}_1(iK\tau) + i\pi \longrightarrow -\mathrm{E}_1(iK\tau). \tag{4.94}$$

The third integral, I_3, can be reduced to the second case:

$$
\begin{aligned}
I_3(\tau) = \int_{-\infty}^{\tau} \mathrm{d}\tau' \frac{1}{\tau'^2} e^{-iK\tau'} &= \left[\frac{-e^{-iK\tau'}}{\tau'} \right]_{-\infty}^{\tau} - iK \int_{-\infty}^{\tau} \mathrm{d}\tau' \frac{e^{-iK\tau'}}{\tau'} \\
&= \left[\frac{-e^{-iK\tau'}}{\tau'} \right]_{-\infty}^{\tau} - iK I_2(\tau) \\
&\longrightarrow \frac{-e^{-iK\tau}}{\tau} + iK \, \mathrm{E}_1(iK\tau).
\end{aligned} \tag{4.95}
$$

Expressing eq. (4.80) in terms of these integrals, using the $+$ contour as the base (which gives an overall minus sign), gives:

$$\frac{\dot{\eta}_* H_*^3 k_3^2}{32\epsilon_*^2 k_1^3 k_2^3 k_3^3} \Im \left[-k_1 k_2 \, I_1(\tau) + i(k_1 + k_2) I_2(\tau) + I_3(\tau) \right] + 2 \text{ perm.} \tag{4.96}$$

Filling in the solutions to the integrals then gives

$$\frac{\dot{\eta}_* H_*^3 k_3^2}{32\epsilon_*^2 k_1^3 k_2^3 k_3^3} \Im \left[-k_1 k_2 \frac{i e^{-iK\tau}}{K} + \frac{-e^{-iK\tau}}{\tau} + i k_3 \, \mathrm{E}_1(iK\tau) \right] + 2 \text{ perm.} \tag{4.97}$$

To take the imaginary part, the analytic series expansion for E_1 can be used; it is given by

$$\mathrm{E}_1(z) = -\gamma - \log z - \sum_{j=1}^{\infty} \frac{(-z)^j}{j j!}, \tag{4.98}$$

where γ is the Euler-Mascheroni constant, valid for all z not on the negative real axis. Using this, the limit for τ approaching 0 is

$$\lim_{\tau \uparrow 0} \mathrm{E}_1(iK\tau) = -\gamma - \log |K\tau| + i\frac{\pi}{2}. \tag{4.99}$$

Taking the imaginary part of the E_1-term,

$$\Im\{i\,\mathrm{E}_1(iK\tau)\} \to -\gamma - \log |K\tau|, \tag{4.100}$$

so the final time-dependent expression for the bispectrum contribution of the $\epsilon\dot{\eta}$-term is

$$\frac{\dot{\eta}_* H_*^3 k_3^2}{32\epsilon_*^2 k_1^3 k_2^3 k_3^3} \left(-k_1 k_2 \frac{\cos(K\tau)}{K} + \frac{\sin(K\tau)}{\tau} - k_3\gamma - k_3 \log |K\tau| \right) + 2 \text{ perm.} \tag{4.101}$$

4.4.5   Issue with convergence

Taking the limit of eq. (4.101) for $\tau \uparrow 0$ gives

$$\frac{\dot{\eta}_* H_*^3 k_3^2}{32\epsilon_*^2 k_1^3 k_2^3 k_3^3} \left(\frac{-k_1 k_2}{K} + K - k_3\gamma - k_3 \log |K\tau| \right) + 2 \text{ perm.} \tag{4.102}$$

We see that this term has a problem with convergence, as $\lim_{\tau \uparrow 0} \log |K\tau| = -\infty$. There should be another $\epsilon\dot{\eta}$-term in the action that can be taken together with this one to give a convergent result. There is a boundary term in time that was neglected by Maldacena [20] when deriving the S_3 action (via integration by parts) that also gives an $\epsilon\dot{\eta}$ term. This term was pointed out by Arroja and Tanaka [3], who also note that boundary terms in time can offer non-negligible contributions, as seen in this explicit example. When adding this boundary term to the action,

$$S_{\text{boundary}} = \int \mathrm{d}^4 x \, \partial_t \left(-\frac{\epsilon\eta a^3}{2} \mathcal{R}^2 \dot{\mathcal{R}} \right), \tag{4.103}$$

this gives an exactly equal but opposite contribution for $\epsilon\dot{\eta}$, such that the 'correct' S_3 action does not contain an $\epsilon\dot{\eta}$-term.

Slow-roll corrections to the mode functions

We also investigated if slow-roll corrections to the mode functions, $u(k,\tau)$, can produce $\epsilon\dot{\eta}$-terms in the bispectrum, but this cannot be the case. The slow-roll parameters appear in the corrections to u (so in the corrections to $\mathcal{R}$) only via the combinations given by eq. (A.10). These would then by multiplied by the slow-roll parameters in front of terms in the S_3 action (eq. 3.41), which would give at the lowest order corrections of $\mathcal{O}(\epsilon^4)$, whereas $\epsilon\dot{\eta}$ is $\mathcal{O}(\epsilon^2)$. We conclude that the slow-roll combination $\epsilon\dot{\eta}$ does not contribute to the bispectrum.

Chapter 5

Conclusion

5.1 Summary

In this thesis, we started with a quick introduction to the CMB, after which inflation was introduced. The (arguably) simplest model for inflation, the canonical scalar field with slow-roll potential, was discussed in detail: we showed how the inflaton coupled to gravity reproduces a period of (isotropic) inflation and how quantum fluctuations of the inflaton, also called perturbations, can reproduce the CMB anisotropies. The primary observables, the power spectrum, bispectrum, and quantities derived from these, of the CMB were defined for the scalar field and theoretical results from literature were given, along with their values from measurement data. The definitions of the observables can be straight-forwardly extended to different single-field models of inflation. One of the most important theoretical results of inflation in general is that there exist quantities (modes) that are conserved when outside the Hubble horizon, which allows perturbations from inflation to be directly related to observed anisotropies in the CMB. The CMB is statistically homogeneous and isotropic and has a nearly scale-invariant power spectrum, which directly reduces the physical degrees of freedom in the power spectrum and bispectrum. The slow-roll scalar field critically also reproduces a near scale-invariant power spectrum.

After this review of literature results, we set out to calculate the contribution of self-interactions of the scalar field to the bispectrum. These contributions are subleading, as the leading-order bispectrum comes directly from coupling to gravity, and have not been the subject of literature review in a general setting; specific examples, such as for resonating potentials with a fixed form, have been studied in detail. We calculated the leading-order power spectrum and for the bispectrum the contribution of the $\epsilon\dot{\eta}$-term from the cubic action (S_3), using general slow-roll assumptions, in the formalism laid out by Maldacena [20]. The calculation was performed in a generic way, such that it doubles as a general review of n-point function calculations in this formalism, and can be directly applied to any term in the S_3 action.

We find a $\log|K\tau|$ divergence, as $\tau \uparrow 0$, for the $\epsilon\dot{\eta}$-term which cannot be the final result, as this would mean that the mode does not freeze out. There is a boundary term in time that exactly cancels the $\epsilon\dot{\eta}$-term[3]. After investigating slow-roll corrections to the mode function $u(k, \tau)$, we find that these cannot produce $\epsilon\dot{\eta}$-corrections to the bispecturm. We conclude that the slow-roll combination $\epsilon\dot{\eta}$ does not contribute to the bispectrum.

5.2 Beyond the canonical scalar field

The canonical scalar field is the simplest model of inflation, but there are many other models, both single-field and multi-field, that can offer inflation compatible with current data. Furthermore, there are two 'issues' with the simplest model that may prove it to be false:

1. The non-Guassianities produced are very small, of order $\mathcal{O}(\epsilon)$, which are not detectable from measurement. Any positive detection of non-Gaussianity in the bispectrum would falsify the simplest model of inflation.

2. From the Lyth bound[19] and its generalisations (e.g. [17][7]), any positive

detection of tensor modes (gravitational waves) would strongly disfavour most single scalar field models. Detection of gravitational waves in the CMB would force ϕ to become super-planckian ($\phi > M_{\rm Pl}$), which makes it problematic to have the single scalar field model as the effective field theory of some quantum gravity model.

These are at least two good motivations to look at more general models of single-field inflation, possibly allowing for multiple fields to be present with one field determining the background evolution (called 'single-clock'). The *effective field theory of inflation*, formulated by Cheung et al. [12], generalises the canonical scalar field to single-clock models and can express all such models in one formalism. Scalar perturbations arise naturally as Goldstone bosons from the broken time diffeomorphisms, because the background evolution spontaneously breaks Lorentz invariance. Using the effective field theory of inflation can give very general results about classes of single-clock models. For instance, a reduced speed of sound, $c_{\rm s} < 1$, directly gives enhanced non-Gaussianity[12] (larger $f_{\rm NL}$). Another result is that it allows a generalisation of the Lyth bound to models beyond the canonical scalar field[7]. A third result is that, by reasoning about the various energy scales during inflation using this theory, predictions may be made what kind of effects 'new' physics near the Hubble scale may have on observations[6].

The generality of the method of bispectrum calculation that we used here means that it can be used also in the effective field theory of inflation to calculate bispectrum contributions. It would be interesting to investigate in this general setting if we can derive new relations between observable quantities, such that restrictions from measurements on the quantities can be combined to give tighter bounds and possible even rule out classes of inflation models. The extension of the Lyth bound[17] is an example of such an endeavour, which uses both the spectral index, $n_{\rm s}$, and the tensor-to-scalar ratio, r, to tighten the Lyth bound.

Acknowledgements

Appendix A

Conformal time and slow-roll

In this appendix the expression of the conformal time, τ, in terms of aH and the slow-roll parameters is derived. The starting point is the definition of conformal time

$$d\tau = \frac{dt}{a}, \tag{A.1}$$

which can be integrated to give

$$\tau - \tau_0 = \int_{t_0}^{t} \frac{dt}{a} = \int_{a_0}^{a} \frac{da}{a^2 H} = \left[\frac{-1}{a} \cdot \frac{1}{H} \right]_{a_0}^{a} - \int_{a_0}^{a} da \frac{\dot{H}}{aH^2} \frac{dt}{da}, \tag{A.2}$$

using first $da = aH \, dt$ and then partially integrating a^{-2}. Choose (τ_0, a_0) such that the a_0-boundary term on the right cancels against τ_0, giving

$$\tau = \frac{-1}{aH} + \int_{a_0}^{a} da \frac{\epsilon}{a^2 H}. \tag{A.3}$$

This is a recursive expression for τ, where at each step the integral

$$I_{\tau,n} = \int_{a_0}^{a} da \frac{1}{a^2} \frac{f_n}{H}, \tag{A.4}$$

needs to be solved for the function f_n that is given by the previous iteration, with $f_0 = 1$ and $f_1 = \epsilon$. Following the same steps as for τ gives a recurisve relation for I_τ,

$$I_{\tau,n} = \left[\frac{-1}{a} \cdot \frac{f_n}{H} \right]_{a_0}^{a} + \int_{a_0}^{a} da \frac{\epsilon f_n + \dot{f}_n/H}{a^2 H}. \tag{A.5}$$

Therefore, the problem can be written as the solution of a recursive relation:

$$\tau - \tau_0 = I_{\tau,0}, \tag{A.6}$$

$$I_{\tau,n} = \left[\frac{-1}{aH} \cdot f_n \right]_{a_0}^{a} + I_{\tau,n+1}, \tag{A.7}$$

$$f_{n+1} = \epsilon f_n + \dot{f}_n/H, \tag{A.8}$$

$$f_0 = 1. \tag{A.9}$$

The function f_n is a polynomial in slow-roll parameters of exactly order n. This can be seen by how the tower of slow-roll parameters $\{p_n\}$ is defined when using Hubble slow-roll parameters:

$$p_{n+1} = \frac{\dot{p}_n}{p_n H} \;\Leftrightarrow\; p_n p_{n+1} = \dot{p}_n/H, \qquad p_0 = \epsilon.$$

Therefore, each $I_{\tau,n}$ starts at order n in slow-roll and the series expression for τ can be truncated after the desired precision in slow-roll is reached. The freedom to choose (τ_0, a_0) can again be used to cancel all a_0-boundary terms in the truncation. The expression for τ, expressed up to third order, is:

$$\tau = \frac{-1}{aH} \left(1 + \epsilon + \epsilon^2 + \epsilon\eta + \epsilon^3 + 2\epsilon^2\eta + \epsilon\eta^2 + \epsilon\eta\xi + \mathcal{O}(\epsilon^4) \right). \tag{A.10}$$

Bibliography

[1] N. Arkani-Hamed and J. Maldacena. 'Cosmological Collider Physics'. In: *ArXiv e-prints* (Mar. 2015). arXiv: 1503.08043 [hep-th].

[2] K. Array et al. 'BICEP2 / Keck Array VI: Improved Constraints On Cosmology and Foregrounds When Adding 95 GHz Data From Keck Array'. In: *ArXiv e-prints* (Oct. 2015). arXiv: 1510.09217.

[3] F. Arroja and T. Tanaka. 'A note on the role of the boundary terms for the non-Gaussianity in general k-inflation'. In: *Journal of Cosmology and Astroparticle Physics* 5, 005 (May 2011), p. 5. DOI: 10.1088/1475-7516/2011/05/005. arXiv: 1103.1102 [astro-ph.CO].

[4] D. Babich, P. Creminelli and M. Zaldarriaga. 'The shape of non-Gaussianities'. In: *Journal of Cosmology and Astroparticle Physics* 8, 009 (Aug. 2004), p. 9. DOI: 10.1088/1475-7516/2004/08/009. eprint: astro-ph/0405356.

[5] D. Baumann. 'TASI Lectures on Inflation'. In: *ArXiv e-prints* (July 2009). arXiv: 0907.5424 [hep-th].

[6] D. Baumann and D. Green. 'Equilateral non-gaussianity and new physics on the horizon'. In: *Journal of Cosmology and Astroparticle Physics* 9, 014 (Sept. 2011), p. 14. DOI: 10.1088/1475-7516/2011/09/014. arXiv: 1102.5343 [hep-th].

[7] D. Baumann and D. Green. 'A field range bound for general single-field inflation'. In: *Journal of Cosmology and Astroparticle Physics* 5, 017 (May 2012), p. 17. DOI: 10.1088/1475-7516/2012/05/017. arXiv: 1111.3040 [hep-th].

[8] C. L. Bennett et al. 'Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results'. In: *Astrophysical Journal, Supplement* 208, 20 (Oct. 2013), p. 20. DOI: 10.1088/0067-0049/208/2/20. arXiv: 1212.5225.

[9] X. Chen. 'Primordial Non-Gaussianities from Inflation Models'. In: *Advances in Astronomy* 2010, 638979 (2010), p. 72. DOI: 10.1155/2010/638979. arXiv: 1002.1416 [astro-ph.CO].

[10] X. Chen, R. Easther and E. A. Lim. 'Large non-Gaussianities in single-field inflation'. In: *Journal of Cosmology and Astroparticle Physics* 6, 023 (June 2007), p. 23. DOI: 10.1088/1475-7516/2007/06/023. eprint: astro-ph/0611645.

[11]   X. Chen, R. Easther and E. A. Lim. 'Generation and characterization of large non-Gaussianities in single field inflation'. In: *Journal of Cosmology and Astroparticle Physics* 4, 010 (Apr. 2008), p. 10. DOI: 10.1088/1475-7516/2008/04/010. arXiv: 0801.3295.

[12]   C. Cheung et al. 'The effective field theory of inflation'. In: *Journal of High Energy Physics* 3, 014 (Mar. 2008), pp. 14–014. DOI: 10.1088/1126-6708/2008/03/014. arXiv: 0709.0293 [hep-th].

[13]   Scott Dodelson. *Modern Cosmology*. San Diego, California: Academic Press, 2003. ISBN: 0-12-219141-2.

[14]   D. J. Fixsen. 'The Temperature of the Cosmic Microwave Background'. In: *Astrophysical Journal* 707 (Dec. 2009), pp. 916–920. DOI: 10.1088/0004-637X/707/2/916. arXiv: 0911.1955.

[15]   R. Flauger and E. Pajer. 'Resonant non-gaussianity'. In: *Journal of Cosmology and Astroparticle Physics* 1, 017 (Jan. 2011), p. 17. DOI: 10.1088/1475-7516/2011/01/017. arXiv: 1002.0833 [hep-th].

[16]   J. Ganc and E. Komatsu. 'Scale-dependent bias of galaxies and μ-type distortion of the cosmic microwave background spectrum from single-field inflation with a modified initial state'. In: *Physical Review D* 86.2, 023518 (July 2012), p. 023518. DOI: 10.1103/PhysRevD.86.023518. arXiv: 1204.4241 [astro-ph.CO].

[17]   J. Garcia-Bellido et al. 'Lyth bound of inflation with a tilt'. In: *Physical Review D* 90.12, 123539 (Dec. 2014), p. 123539. DOI: 10.1103/PhysRevD.90.123539. arXiv: 1408.6839 [hep-th].

[18]   E. Komatsu et al. 'Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation'. In: *Astrophysical Journal, Supplement* 192, 18 (Feb. 2011), p. 18. DOI: 10.1088/0067-0049/192/2/18. arXiv: 1001.4538 [astro-ph.CO].

[19]   D. H. D. H. Lyth and A. A. Riotto. 'Particle physics models of inflation and the cosmological density perturbation'. In: *Physics Reports* 314 (June 1999), pp. 1–146. DOI: 10.1016/S0370-1573(98)00128-8. eprint: hep-ph/9807278.

[20]   J. Maldacena. 'Non-gaussian features of primordial fluctuations in single field inflationary models'. In: *Journal of High Energy Physics* 5, 013 (May 2003), p. 13. DOI: 10.1088/1126-6708/2003/05/013. eprint: astro-ph/0210603.

[21]   NASA / WMAP Science Team. *7 Year CMB Angular Spectrum*. [Image; online]. 2011. URL: http://map.gsfc.nasa.gov/media/121238/index.html (visited on 08/12/2015).

[22]   NASA / WMAP Science Team. *Nine Year Microwave Sky*. [Image; online]. 2014. URL: http://map.gsfc.nasa.gov/media/121238/index.html (visited on 08/12/2015).

[23]   Planck Collaboration et al. 'Planck 2015 results. XIII. Cosmological parameters'. In: *ArXiv e-prints* (Feb. 2015). arXiv: 1502.01589.

[24]   Planck Collaboration et al. 'Planck 2015 results. XVII. Constraints on primordial non-Gaussianity'. In: *ArXiv e-prints* (Feb. 2015). arXiv: 1502.01592.

[25]   Steven Weinberg. *Cosmology.* New York: Oxford UP, 2008. ISBN: 978-0-19-852682-7.

[26]   WikiMedia Foundation. *Cosmic Microwave Background Spectrum from COBE.* [Image; online] Created by user 'Quantum Doughnut'. 2011. URL: `https://en.wikipedia.org/wiki/File:Cmbr.svg` (visited on 08/12/2015).

Index