Topics in Gravitation and Cosmology

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Sina Bahrami Taghanaki
August 2017

Topics in Gravitation and Cosmology

Sina Bahrami Taghanaki, Ph.D.

Cornell University 2017

This thesis is focused on two topics in which relativistic gravitational fields play an important role, namely early Universe cosmology and black hole physics. The theory of cosmic inflation has emerged as the most successful theory of the very early Universe with concrete and verifiable predictions for the properties of anisotropies of the cosmic microwave background radiation and large scale structure. Coalescences of black hole binaries have recently been detected by the Laser Interferometer Gravitational Wave Observatory (LIGO), opening a new arena for observationally testing the dynamics of gravity.

In part I of this thesis we explore some modifications to the standard theory of inflation. The main predictions of single field slow-roll inflation have been largely consistent with cosmological observations. However, there remain some aspects of the theory that are not presently well understood. Among these are the somewhat interrelated issues of the choice of initial state for perturbations and the potential imprints of pre-inflationary dynamics. It is well known that a key prediction of the standard theory of inflation, namely the Gaussianity of perturbations, is a consequence of choosing a natural vacuum initial state. In chapter 3, we study the generation and detectability of non-Gaussianities in inflationary scalar perturbations that originate from more general choices of initial state. After that, in chapter 4, we study a simple but predictive model of pre-inflationary dynamics in an attempt to test the robustness of inflationary predictions. We find that significant deviations from the standard predictions are unlikely to result from models in which the inflaton field decouples from the pre-inflationary degrees of freedom prior to freeze-out of the observable modes.

In part II we turn to a study of an aspect of the thermodynamics of black holes, a subject which has led to important advances in our understanding of quantum gravity. For objects which collapse to form black holes, we examine a conjectured relationship between the objects' entropy, the collapse

timescale, and the mass of the final black hole. This relationship is relevant for understanding the nature of generic quantum mechanical states of black hole interiors. In chapter 6 we construct a counter-example to a weak version of the conjectured relation.

Biographical Sketch

Sina spent most of his childhood and adolescence in Tehran. He is blessed to have an incredibly supportive family who prioritized Sina's education as the main objective of their lives. They taught him to be ambitious and to always expect the best of himself.

While living and studying in North Jersey, Sina decided to attend Stevens Institute of Technology where he spent most of his time studying mathematics. He found an interest in general relativity after reading a review article on mathematical relativity in an issue of Bulletin of the American Mathematical Society. Later on, he had the opportunity to learn and conduct research in general relativity and related fields as a doctoral student at Cornell University.

After graduation, Sina will be joining the Institute for Gravitation and the Cosmos at the Pennsylvania State University as a postdoctoral researcher where he will continue to conduct research on aspects of gravitation and theoretical cosmology.

To my beloved parents

Acknowledgements

My very modest accomplishments would have been impossible without the vast support of my parents and my lovely sister.

As a doctoral student at Cornell University, I was most fortunate to have Professor Eanna É. Flanagan as my doctoral advisor. He is an incredible teacher to whom I am indebted for my modest understanding of general relativity. I also benefited from discussions with several postdoctoral researchers, including Drs. David Nichols, Justin Vines, Jolyon Bloomfield, Kartik Prabhu, Leo Stein, and Shawn Henderson, as well as fellow graduate students Jordan Moxon and Alexander Grant.

I enjoyed learning physics and mathematics from many distinguished teachers at Cornell University. For this I would like to thank Professors Yuval Grossman, Csaba Csaki, Michael Niemack, Saul Teukolsky, Ira Wasserman, Leonard Gross, John Hubbard, Yuri Berest, Kenneth Brown, and William Thurston (of blessed memory). I would also like to thank the rest of my Ph.D. committee members, Professors Liam McAllister and Julia Thom-Levy.

Finally, I would like to thank Drs. Edward and Arline Joan Friedman and Dr. Norman Morgenstern Horing for more than a decade of continued support and encouragement.

The work of this thesis was supported by Cornell University, John and David Boochever prize fellowship in theoretical physics, National Science Foundation, and National Aeronautics and Space Administration.

Table of Contents

Preface

Citations to Previously Published Work

Chapters 3, 4, and 6 of this thesis have been adapted from the following three previously published articles,

> Sina Bahrami, Éanna Flanagan. Primordial non-Gaussianities in single field inflationary models with non-trivial initial states. JCAP 1410 (2014) 10, 010.
>
>
> Sina Bahrami, Éanna Flanagan. Sensitivity of inflationary predictions to pre-inflationary phases. JCAP 01 (2016) 027.
>
>
> Sina Bahrami. Saturating The Bekenstein-Hawking Entropy Bound With Initial Data Sets For Gravitational Collapse. Phys. Rev. D. 95.026006.
>

Permission from coauthors has been granted for these works to be included in this dissertation.

List of Abbreviations

The following abbreviations are commonly used in this document.

- BBN – *Big Bang Nucleosynthesis*.
- CDM – *Cold Dark Matter*.
- CMB – *Cosmic Microwave Background*.
- COBE – *Cosmic Background Explorer*.
- EFT – *Effective Field Theory*.
- EW – *Electro-Weak*, referring to the Electro-Weak unification scale.
- FRW – *Friedmann-Robertson-Walker*, referring to the Friedmann-Robertson-Walker metric.
- GUT – *Grand Unified Theory*.
- LSS – *Large Scale Structure*.
- pc – *parsec*, an astrophysical measure of distance that is approximately equal to 3.262 light years.
- UV – *Ultraviolet*.
- WMAP – *Wilkinson Microwave Anisotropy Probe*.

Notations and Conventions

Throughout this dissertation,

- For Part I, We work in the units $c = \hbar = 1$. Additionally, we shall use the reduced Planck mass given by $m_p = (8\pi G)^{-1/2}$ in lieu of the Newton's gravitational constant G. For Part II, we work in the units $c = \hbar = G = k_B = 1$.
- The four dimensional metric signature is $(- + ++)$.
- Dot denotes differentiation with respect to the coordinate t unless stated otherwise.
- Subscript 0 on the background functions indicates their present values unless stated otherwise.

Part I

Beyond The Standard Theory Of Inflation

Chapter 1

Introduction To Inflationary Theory

1.1 A brief review of the standard model of cosmology

Our best cosmological model describing the relatively recent history of the Universe's expansion is the standard model of cosmology. This model is founded on the premise that the Universe is homogeneous and isotropic on sufficiently large scales. The appropriate metric to describe the geometry of the spacetime on such scales is therefore the Friedmann-Robertson-Walker (FRW) metric

$$ds^2 = -dt^2 + a(t)^2\left[\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\right]. \tag{1.1.1}$$

Here $a(t)$ is called the scale factor, $d\Omega^2 \equiv d\theta^2 + \sin\theta^2 d\phi^2$ is the surface element on the unit 2-sphere, and $k \in \{0, \pm 1\}$ with $k = 0$ corresponding to flat and $k = +1$ (-1) corresponding to spherical (hyperbolic) constant time slices [1]. The metric given in Eq. (1.1.1) is the unique homogeneous and isotropic metric in four dimensions. [2]

Einstein's equations evaluated on the FRW background reduce to the following two Friedmann equations [3]

$$3m_p^2\left(H^2 + \frac{k}{a^2}\right) = \bar{\rho}_{\text{total}}, \tag{1.1.2}$$

$$m_p^2\frac{\ddot{a}}{a} = -\frac{1}{6}(\bar{\rho}_{\text{total}} + 3\bar{p}_{\text{total}}), \tag{1.1.3}$$

where $H \equiv \dot{a}/a$ is the Hubble constant and $\bar{\rho}_{\text{total}}$ and $\bar{p}_{\text{total}}$ are the total background energy density and pressure that are associated with the overall energy and matter content of the Universe. A large

[1] k is related to the Ricci scalar of the constant time slices via ${}^{(3)}R = 6k$.

[2] Proofs are straightforward to obtain. See e.g. [1].

[3] See App. B for a review of the derivation.

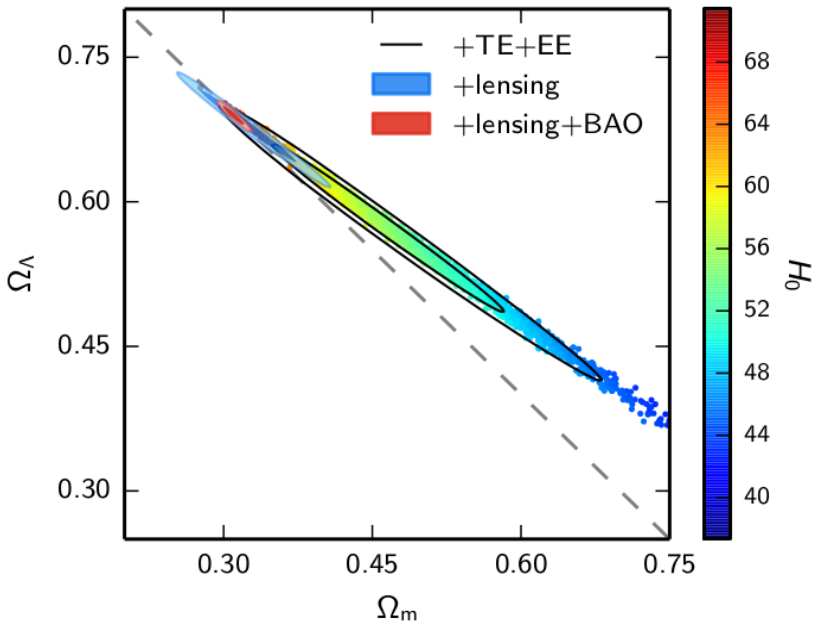


Figure 1: A plot of the density parameter Ω_m of non-relativistic matter versus the density parameter Ω_Λ for dark energy from a recent analysis by the Planck Collaborations [2]. It includes the most recent data associated with the temperature and polarization anisotropy maps of the CMB, as well as data coming from the gravitational lensing and Baryon Acoustic Oscillations. The solid black curves are one and two sigma probability contours. The color coding shows the most likely value of the Hubble constant H_0 in different areas of the parameter space. Reproduced by permission of the Planck Collaborations.

array of observations, including collective data from the Cosmic Microwave Background (CMB) and Type Ia Supernovae point to the following present composition of the Universe [2, 3, 4, 5]:

- Roughly 30% non-relativistic matter; about four fifth of which is thought to be cold dark matter, with the rest baryons.
- Roughly 70% dark energy.
- About 10^{-3}% radiation, consisting of photons and neutrinos [4].

The energy density for matter and radiation components dilute as a^{-3} and a^{-4} respectively. For dark energy, in absence of any observational evidence pointing to its potential dynamical nature,

[4]The photons (the CMB) were discovered in 1964 [6]. Previous calculations of Big Bang Nucleosynthesis using the Cold Big Bang hypothesis, in which the Universe is assumed to be matter dominated at arbitrarily high redshifts, had already exposed the need for a radiation dominated era in the early Universe [7, 8, 9].

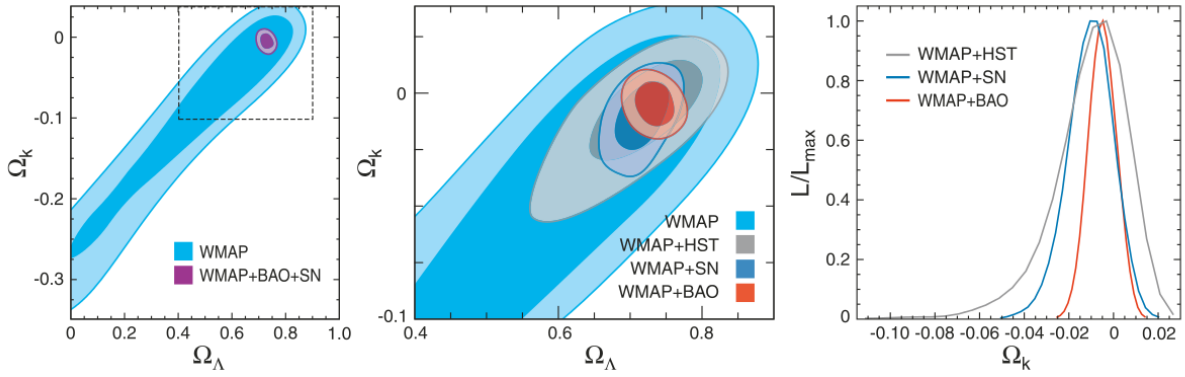


Figure 2: Constraints on Ω_k and Ω_Λ that were obtained from WMAP's CMB data as well as supernova data [11, 12]. Note that Ω_k is tightly constrained to a near zero value. Reproduced by permission of the authors.

the energy density is given by the so-called Cosmological Constant $\Lambda_0 \sim 10^{-120} m_p^4 \sim 10^{-48}\text{GeV}^4 = 10^{-12}\text{eV}^4$. [5]

Cosmologists find it convenient to re-express Eq. (1.1.2) as

$$H^2 = H_0^2 \Big[\Omega_{\Lambda_0} + \Omega_k (1+z)^2 + \Omega_m (1+z)^3 + \Omega_r (1+z)^4 \Big], \tag{1.1.4}$$

where $H_0 = 67.8 \pm 0.9$ km/(sec Mpc) is the present value of the Hubble constant, $z \equiv a_0/a - 1$ is the cosmological redshift, and the cosmological density parameters Ω_{Λ_0}, Ω_k, Ω_m, and Ω_r are defined as

$$\Omega_{\Lambda_0} \equiv \frac{\Lambda_0}{3m_p^2 H_0^2}, \qquad \Omega_k \equiv -\frac{k}{H_0^2}, \qquad \Omega_m \equiv \frac{\bar{\rho}_{m,0}}{3m_p^2 H_0^2}, \qquad \Omega_r \equiv \frac{\bar{\rho}_{r,0}}{3m_p^2 H_0^2}. \tag{1.1.5}$$

Here $\bar{\rho}_{m,0}$ is the present day density of non-relativistic matter and $\bar{\rho}_{r,0}$ that of radiation. Given the current composition of the Universe, $\Omega_{\Lambda_0} \approx 0.7$, $\Omega_m \approx 0.3$, and $\Omega_r \approx 10^{-5}$. The most recent set of observations have constrained the value of Ω_k to less than 10^{-2} [2]. As such, Ω_k is normally set to zero in the standard model of cosmology.

[5] The unnatural smallness of this universal constant is the well known cosmological constant problem [10]. Λ_0 should receive a contribution from the vacuum energy of the quantum fields present in the Universe, it is expected to receive a contribution $\sim 10^9\text{GeV}^4$ (assuming a cutoff on the order of the mass of the top quark) from the quantum fields present in the Standard Model of particle physics alone.

1.2 The puzzles of Hot Big Bang cosmology

As can be seen from Eq. (1.1.4), the radiation component dominates the overall energy density of the Universe at sufficiently high redshifts. However, that the radiation continues to dominate up to arbitrarily high redshifts was found to give rise to a number of cosmological puzzles, with some being severe fine tuning problems and some being inconsistencies with basic observations [6]. The Hot Big Bang theory was plagued with these problems for decades until the advent of the inflationary paradigm.

We now briefly review these problems that were dubbed the classical problems of cosmology [7] [8].

- **The Horizon Problem**

  The Hot Big Bang theory falls short of explaining the near isotropy of the CMB temperature map. Indeed, the temperature variations for the CMB photons are about one part in 10^5. This is despite the fact that the Hot Big Bang theory predicts a large number of causally disconnected patches on the CMB map. In fact, should we assume the Friedmann Eq. (1.1.4) to remain valid at arbitrarily high redshifts, the causal processes that originate on a spacelike slice in the far past can only influence regions that subtend an angular size on the order of a few degrees at the time the CMB photons began to freely stream towards us. This can be shown by calculating the ratio of the particle horizon distance d_p over the angular diameter distance d_A on the surface of last scattering [9]

$$\left.\frac{d_p}{d_A}\right|_{z\approx 1100} = \left.\frac{a_{ls}\int_0^{a_{ls}}\frac{da}{a^2H}}{a_{ls}\int_{a_{ls}}^{a_0}\frac{da}{a^2H}}\right|_{z\approx 1100} \sim \left.\frac{H_0^{-1}(1+z)^{-3/2}}{H_0^{-1}(1+z)^{-1}}\right|_{z\approx 1100} \sim 2 \text{ degrees},^{10} \tag{1.2.1}$$

  where a_{ls} is the scale factor at the surface of last scattering and we used Eq. (1.1.4) and the

[6] Roger Penrose classifies them as external versus internal problems, with internal problems being the results of the required self-consistency of a presumed UV theory regardless of its experimental success (e.g. the monopole problem in the context of most GUT theories), and external problems being the results of inconsistencies with observations [13].

[7] Almost all reviews of the theory of inflation include discussions of these problems. One can consult [14] or [11] for further information.

[8] Aside from the these problems, there have been a number of less-discussed problems associated with the Hot Big Bang cosmology that are motivated by theoretical hypotheses regarding the physics of the early Universe, such as the Polonyi fields problem [15].

[9] The surface of last scattering refers to the redshift at which the CMB photons stopped scattering off matter and began to freely propagate. This phenomenon is known to have occurred at the redshift $z \approx 1100$.

[10] Note that we are assuming that the spacetime geometry is approximated using the FRW metric at arbitrarily high redshifts. This may be challenged on the basis of our ignorance of the quantum gravity influence on the classical geometry.

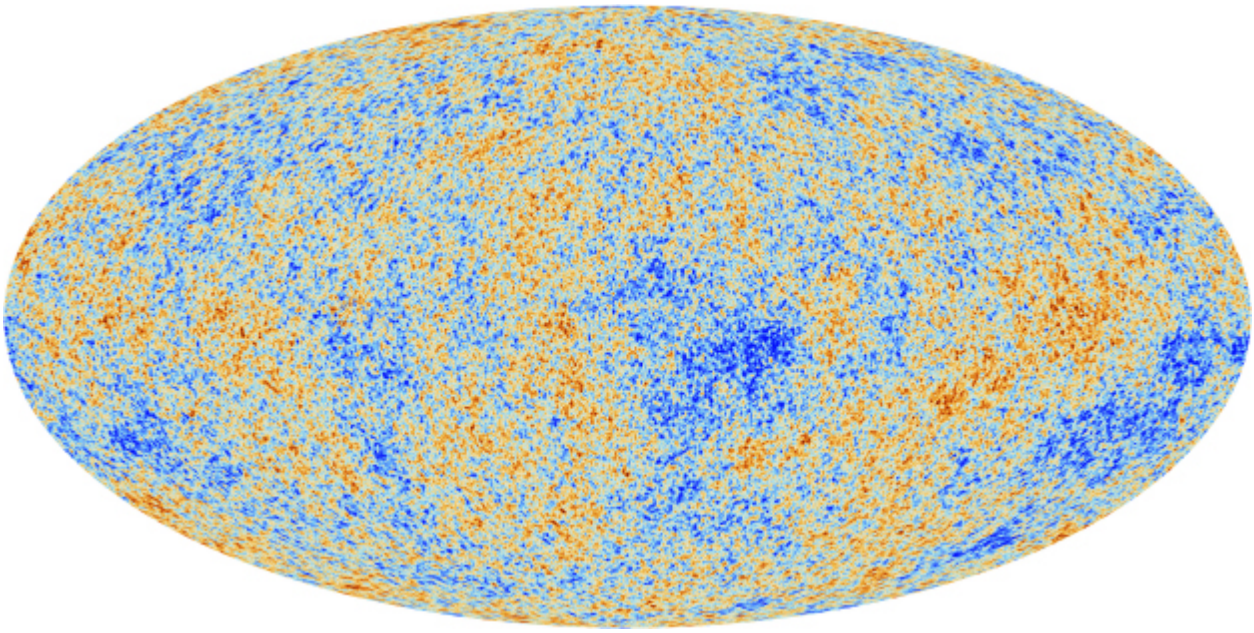


Figure 3: This is the most recent map of the CMB temperature anisotropy obtained by the Planck satellite (Image credit: European Space Agency & Planck Collaborations). The temperature anisotropy is roughly one part in 10^5, with the average temperature being $T \approx 2.8^\circ$ K.

definitions of particle horizon and angular diameter distance [11]. This implies that our sky must consist of about $2^2/(360)^2 \approx 3 \times 10^5$ causally disconnected regions.

- **The Flatness Problem**

  The present Universe is known to be spatially flat on the cosmological scales, as discussed above. The fractional contribution of spatial curvature to the overall energy density diminishes with increasing redshift. In fact, it follows from Eq. (1.1.4) that

$$\frac{\bar{\rho}_k}{\bar{\rho}_{\rm total}} \approx \frac{\Omega_k}{\Omega_r (1+z)^2} \qquad \text{as } z \to \infty. \tag{1.2.2}$$

  If presently $\Omega_k \lesssim 10^{-2}$ and Eq. (1.1.4) is to remain valid at arbitrarily high redshifts, then $\bar{\rho}_k/\bar{\rho}_{\rm total} \sim 10^{-57}$ at $z \sim 10^{30}$ when $\bar{\rho}_{\rm total} \sim \left(10^{18}{\rm GeV}\right)^4 \sim m_p^4$. This is to say that the spatial curvature is required to be finely tuned to unnaturally small values in the initial conditions for the Universe at the end of the quantum gravity era.

- **The Primordial Relic Problem**

[11] See App. A.

Grand Unified Theories (GUTs) typically involve spontaneous symmetry breaking mechanisms that involve degrees of freedom that have non-trivial topologies (e.g. magnetic monopoles). The number density of these relics estimated using the size of the particle horizon in the Hot Big Bang paradigm is roughly ten orders of magnitude below that of photons [16, 17] [12]. In absence of sufficient pair annihilation, they would be as abundant as baryons today. Similarly, as they are typically non-relativistic, their energy density scales with a^{-3}. This conflicts with observations.

1.3 The emergence of the standard inflationary paradigm

Inflation refers to a phase during the primordial era when the Universe underwent a nearly exponential expansion. Though the concept originally emerged in the context of the GUTs, it did not take long for cosmologists to realize that such an expansion phase could solve the classical puzzles of the Hot Big Bang cosmology [13].

To provide some historical background, we remind the reader that what is currently known as the old theory of inflation was proposed in the early 1980's based on the theory of first order cosmological phase transition [19, 20, 21, 22]. It was then proposed that the Universe underwent an exponential expansion due to the descent of a scalar field, called the inflaton, to the minimum of its effective potential following a period of thermal expansion [14]. Several inconsistencies were later pointed out, mainly with regards to how inflation ended and how the cosmological structures formed in these models.

The current theory of inflation emerged shortly after the old theory in 1983 following Andre

[12] Some monopole pair annihilation calculations decrease this figure by another ten orders of magnitude. See [18].

[13] Penrose is critical of this point of view. He regards the near uniformity of our Universe (an aspect of which is regarded as the horizon problem) as being consistent with the second law of thermodynamics. The entropy of our Universe as a whole should continue to increase in the future direction, which implies that our Universe began in a state of vanishingly small entropy, i.e. very finely tuned initial condition. In his view, the fundamental problem is the lack of a theoretical framework that explains why the Universe's entropy as a whole should decrease as we head towards the past. There is currently no compelling theory that explains this time asymmetry. He regards the basic inflationary premise, which purports to explain the observable structures from generic initial conditions, as being misconceived[13].

[14] This idea was partly based on the theory of false vacuum decay, which was completed a few years earlier [23, 24, 25].

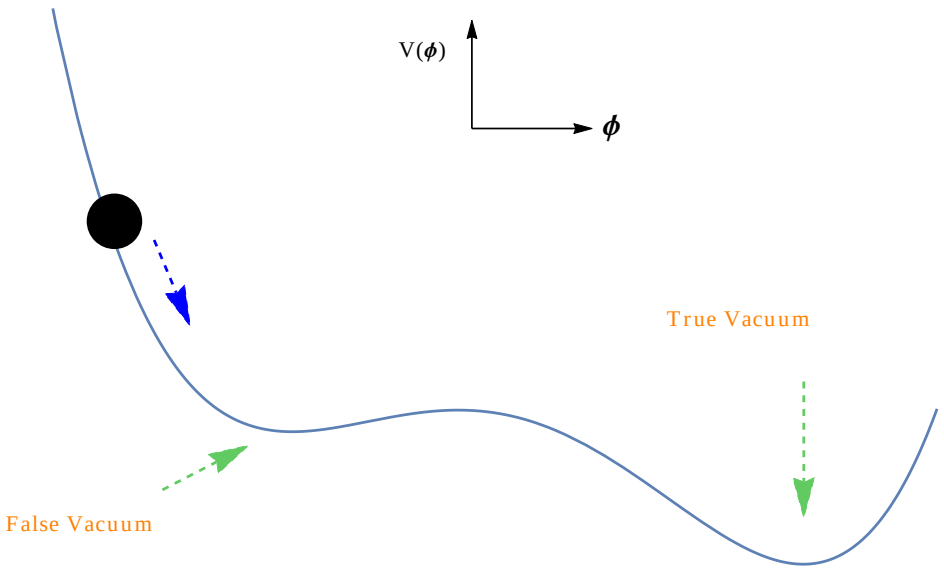


Figure 4: An illustration of the old inflationary scenario: The scalar field ϕ rolls down its effective potential $V(\phi)$, remains in the false vacuum for a while, and then reaches the true vacuum via quantum tunnelling. While stuck in the false vacuum, $V(\phi)$ is constant which results in the Universe's exponential expansion.

Linde's proposal of the chaotic inflation scenario [26][15]. Chaotic inflation dispelled the notion that inflation must have followed a period of thermal expansion, or that it was initiated as a result of the descent of the inflaton to the minimum of its effective potential. Instead, it brought attention to the fact that the main characteristic of inflation, namely a quasi-exponential expansion of the Universe, could be achieved by requiring the inflaton potential to be sufficiently flat. This paradigm later became known as the "slow-roll" inflation. Currently, the most successful models of inflation belong to this category.

Though we tend to regard inflation as an era of quasi-exponential expansion of the Universe, a more general definition has been proposed based on the mechanism that solves the classical cosmological puzzles. As it turns out, the puzzles can be circumvented with a sufficiently long period of shrinking comoving Hubble horizon, i.e. by requiring $d(aH)^{-1}/dt < 0$ during inflation [11]. It follows that $\ddot{a} > 0$, or that the Universe's expansion must accelerate during inflation. From Eq. (1.1.3) we then have $\bar{\rho}_{\rm inf} + 3\bar{p}_{\rm inf} < 0$ where $\bar{\rho}_{\rm inf.}$ and $\bar{p}_{\rm inf}$ are the energy density and pressure during inflation, implying that the equation of state parameter for the inflaton, $w_{\rm inf} \equiv \bar{p}_{\rm inf}/\bar{\rho}_{\rm inf}$, is

[15]Today, chaotic inflation refers to the single field models of inflation with monomial potentials.

less than $-1/3$ [16]. We will see in the review of the standard theory of inflation presented in chapter 2 that the properly normalized amplitude $\Delta_{\mathcal{R}}$ of the primordial scalar fluctuations generated during inflation is predicted to be roughly $H_{\rm inf}^2/\left(\sqrt{\dot{H}_{\rm inf}}m_p\right)$, where $H_{\rm inf}$ is the Hubble parameter during inflation. This quantity is measured to be nearly constant over the large range of scales probed using the CMB anisotropy maps. If one extrapolates this near-constancy to other scales, it follows that $a(t) \sim e^{H_{\rm inf}t}$, or that the Universe expands quasi-exponentially during inflation. This conclusion along with Eqs. (1.1.2) and (1.1.3) (with $k = 0$) implies that $m_p^2|\dot{H}_{\rm inf}| = |\bar{\rho}_{\rm inf} + \bar{p}_{\rm inf}|/2 \ll m_p^2 H_{\rm inf}^2$, or $w_{\rm inf} \equiv \bar{p}_{\rm inf}/\bar{\rho}_{\rm inf} \approx -1$.

For the particular case of $a \sim e^{H_{\rm inf}t}$, we can estimate the duration of inflation needed to eliminate the horizon and flatness problems [17]. To address the horizon problem, we need to have $d_p/d_A \sim 10^2$ degrees at the surface of last scattering. Denoting the scale factors at the onset and at the end of inflation by a_i and a_f respectively, we have

$$\left.\frac{d_p}{d_A}\right|_{z\approx 1100} = \left.\frac{a_{ls}\left(\int_0^{a_i} + \int_{a_i}^{a_f} + \int_{a_f}^{a_{ls}}\right)\frac{da}{a^2H}}{a_{ls}\int_{a_{ls}}^{a_0}\frac{da}{a^2H}}\right|_{z\approx 1100} \sim \frac{a_f}{a_i}10^6\sqrt{\frac{H_0}{H_{\rm inf}}}, \tag{1.3.1}$$

where we used Eq. (1.1.4) and assumed that d_p receives most of its contribution from the inflationary phase. If we assume that the energy scale of inflation is $\sqrt{m_p H_{\rm inf}} \sim 10^{16}$ GeV (a typical GUT scale), then to solve the horizon problem we need $a_f/a_i \sim 10^{24}$, which is more commonly expressed in terms of $N = \log(a_f/a_i)$, the number of e-folds, being larger than 55. On the other hand, if the energy scale of inflation is as low as 1 TeV, then roughly 24 e-folds would be sufficient to circumvent the horizon problem.

As for the flatness problem, we begin by assuming that $\bar{\rho}_k \sim \bar{\rho}_{\rm total}$ at the onset of inflation. In order to have $\Omega_k < 10^{-3}$ today, it is sufficient to have

$$\bar{\rho}_k\Big|_{a_f} = \bar{\rho}_k\Big|_{a_i}\left(\frac{a_i}{a_f}\right)^2 \sim m_p^2 H_{\rm inf}^2\left(\frac{a_i}{a_f}\right)^2 < 10^{-3}m_p^2 H_0^2\left(\frac{a_0}{a_f}\right)^2. \tag{1.3.2}$$

This implies that for GUT-scale inflation roughly 63 e-folds and for TeV-scale inflation roughly 33 e-folds is needed. All in all, most estimates show that roughly 55-60 e-folds of GUT-scale inflation is required to solve the classical cosmological problems.

[16] Equivalently, the stress energy tensor of the inflaton must violate the strong energy condition.

[17] The primordial relic problem turns out to be less restrictive, in the sense that it requires a significantly shorter period of inflation [17].

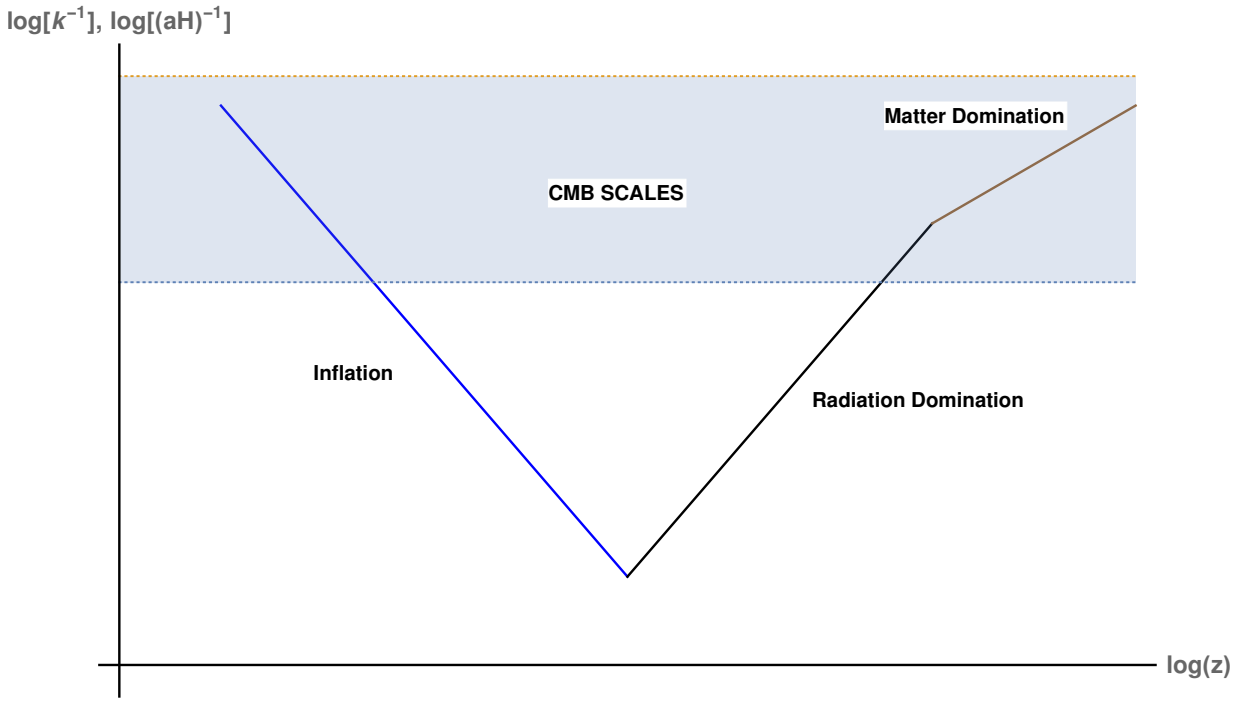


Figure 5: The comoving Hubble horizon, $(aH)^{-1}$, decreases during inflation. The fluctuations on the CMB sky were influenced by the causal processes within the Hubble horizon at time of inflation. Perturbations are generated when a given perturbation mode of comoving wavenumber k exits the horizon during inflation at $k \sim aH_{\rm inf}$.

In addition to solving the classical cosmological problems by design, inflation provides a natural mechanism for generation of primordial fluctuations via quantum mechanical vacuum fluctuations of the inflaton field. These fluctuations give rise to the classical macroscopic density perturbations that clump together at late times to form the currently observed cosmological structures. As we will review in chapter 2, the measured statistical features of cosmological perturbations collected over a wide range of scales, from the temperature and polarization anisotropy maps of the CMB as well as large scale structure (LSS), are well explained by the key properties of their primordial initial conditions; namely that they are nearly scale-invariant, adiabatic, and highly Gaussian. This level of observational success has granted a significant credence to the standard inflationary paradigm and has practically confirmed its occurrence.

1.4 Pushing the boundaries of the standard theory of inflation

The most successful realizations of the inflationary paradigm to date that are placed under the umbrella of "the standard theory of inflation" are single field slow-roll models of inflation. These field theoretic models consist of a single scalar field, the inflaton, which is typically endowed with a canonical kinetic term and a nearly flat potential [18]. The near flatness of its potential renders the so-called slow-roll regime an attractor for the background inflationary dynamics, which in turn results in a period of quasi-exponential expansion of the Universe. What renders the standard theory of inflation so appealing is the level of its success in predicting various features of cosmological perturbations, most importantly as it relates to properties of the CMB. Nonetheless, no precise knowledge of the energy scale at which inflation occurred exists. Despite the fact that most inflationary model builders prefer models of inflation with energy scale $\sqrt{m_p H_{\rm inf}} \sim 10^{16}$GeV, inflation could have theoretically occurred at energy scales as low as the Electroweak Unification scale ~ 1 TeV [19].

In absence of a "smoking gun" evidence of inflation pointing to its energy scale, it is difficult to gauge the importance of the hierarchy between the inflation scale and the Planck scale and the possibility of an interesting new physics that could have existed between the two scales. Meanwhile, as the precision of CMB and LSS measurements improve in the future cosmological observations, we move closer to identifying the detailed inflationary Lagrangian and the relevant degrees of freedom during inflation. As an example, the long standing discrepancy between the predictions of the standard theory and observations for the low-multipole modes of the temperature anisotropy map of the CMB serves as a motivations for such theoretical endeavours [20]. It is therefore imperative to examine the predictions of various extensions of the standard theory of inflation and distinguish viable directions in which inflation can be extended or modified.

The main directions in which the standard theory of inflation has been extended so far are models where inflation is driven by multiple degrees of freedom, and models where a drastically different pre-inflationary physics is assumed. The former direction typically involves various degrees

[18] There are exceptions, such as k-inflation, where the inflaton has a non-canonical kinetic term [27]. It should be added that these models have not been quite successful in explaining CMB observations.

[19] TeV-scale inflation models were considered at the time when the extra dimension models became popular. See e.g. [28] for one of the first studies of TeV-scale inflation.

[20] See the recent paper by Planck collaborations on the constraints on inflation [29]. The low-l discrepancy is statistically significant.

of fine tuning in order to curb the generation of non-adiabatic perturbation modes, as they are known to be highly suppressed [21][22]. The relevance of the latter is dependant on how long inflation lasted, with longer duration of inflation generally corresponding to less sensitivity to pre-inflationary phases. The main phenomena that belong to the latter category are the sub-Planckian degrees of freedom that decouple during inflation [23], modifications to the spacetime geometry [24], Ekpyrotic Universes [25], and modifications due to quantum gravity such as bouncing geometry as predicted by Loop Quantum Cosmology or quantum mechanical tunneling in a string theory landscape of vacua [26]. All these can leave detectable imprints on the primordial fluctuations that can be constrained using the present and future generation CMB and LSS observations.

1.5 Structure of the thesis for part I

In chapter 2, we will review the basic theory of single field inflation. We first discuss the fundamental aspects of the inflationary Lagrangian, its construction using symmetry principles, and its quantum mechanical robustness. We then study various examples of the background dynamics, as well as those of scalar and tensor perturbations during inflation. We will discuss the current observational constraints on models of single field inflation, distinguishing the currently favoured models from the ruled-out ones. We will close with a general discussion of how inflation ended.

In chapter 3, we will first discuss the motivations for considering non-vacuum initial states for inflationary perturbations. After reviewing our approach in parametrizing the non-vacuum initial states, we compute the inflationary 2-point and 3-point correlation functions for scalar perturbations. Lastly, we will discuss the observational consequences of non-vacuum initial states.

In chapters 4 we will discuss in details the predictions of a particular pre-inflationary model. We will examine and comment on the relevance of the modifications caused by the presence of a pre-

[21]Except perhaps for multipoles $l \lesssim 40$. See [29].

[22]There is a large literature on multifield models of inflation and their properties. See e.g. [30] for a discussion of their predictions.

[23]We will discuss an example of this phenomenon in chapter 4. It is also possible to model the influence of some pre-inflationary degrees of freedom on the inflationary correlation functions using the non-vacuum initial states formalism. See chapter 3 for an example.

[24]e.g. if the spacetime geometry is anisotropic at the onset of inflation.

[25]We will not review these models here. See [31] for a discussion of the foundations of the theory and its key predictions.

[26]See [32] for a review of Loop Quantum Cosmology, and [33, 34] for a discussion of inflation occurring in a string theory landscape.

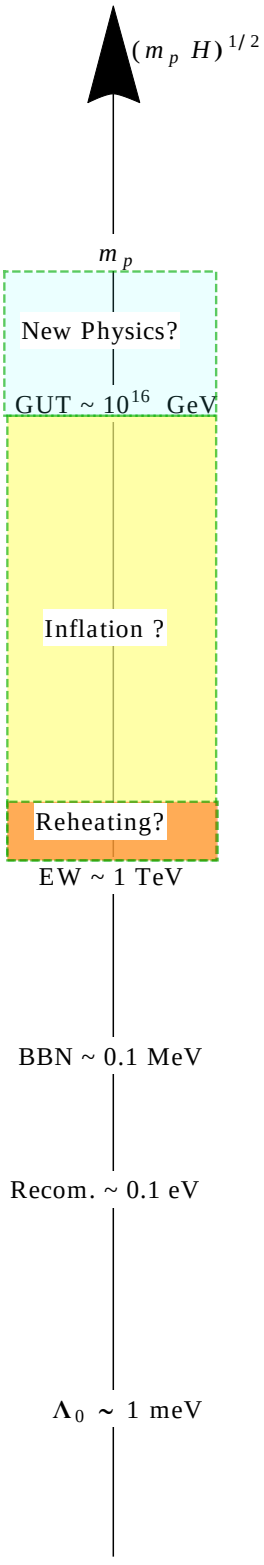


Figure 6: Energy Scale Axis.

inflationary degree of freedom on inflationary 2-point correlation functions.

Chapter 2

A Review Of The Theory Of Single Field Inflation

In this chapter we will review the rudiments of the theory of single field slow-roll inflation. We will discuss various aspects of the single field Lagrangian, including its sensitivity to the fundamental UV physics. We will discuss the general properties of the background spacetime and perturbations as they evolve during inflation. The observational constraints on inflationary predictions along with future observational expectations will also be reviewed. Finally, a brief discussion of the reheating and pre-heating processes will be given at the end of this chapter.

The material reviewed in this chapter are mainly based on [11, 35, 36, 37, 38, 39, 40, 41], which the reader can consult for more in depth discussions.

2.1 The single field action

In the standard inflationary theory, where inflation is driven by a single scalar field known as the inflaton, the overall action is the sum of the Einstein-Hilbert action and the standard scalar field action,

$$S[g_{ab},\phi] = S_{\rm EH}[g_{ab}] + S_{\rm inf}[g_{ab},\phi] = \int d^4x\ \sqrt{-g}\ \left[\frac{1}{2}m_p^2 R - \frac{1}{2}(\nabla\phi)^2 - V(\phi)\right], \tag{2.1.1}$$

where m_p is the reduced Planck mass, R is the Ricci scalar of the spacetime metric, and we define $(\nabla\phi)^2 \equiv g^{ab}\nabla_a\phi\nabla_b\phi$. As can be seen from Eq. (2.1.1), the inflaton field is minimally coupled to gravity and its action consists of a canonical kinetic term and a potential term.

We derive the dynamical equations for the metric and the inflaton field by varying the action

(2.1.1) with respect to g_{ab} and ϕ respectively. Doing so results in the following set of Einstein equations for the metric as well as the Klein-Gordon equation for the inflaton field

$$\frac{\delta(S_{\rm EH}+S_{\rm inf})}{\delta g^{ab}}=0\Rightarrow\frac{2\delta S_{\rm EH}}{m_p^2\delta g^{ab}}=-\frac{2\delta S_{\rm inf}}{m_p^2\delta g^{ab}}\Rightarrow G_{ab}=m_p^{-2}T_{ab},$$
$$\frac{\delta S_{\rm inf}}{\delta\phi}=0\Rightarrow\Box\phi-V_{,\phi}=0, \tag{2.1.2}$$

where $\Box\equiv g^{ab}\nabla_a\nabla_b$ and $V_{,\phi}\equiv\partial_\phi V$. Here the Einstein tensor G_{ab} and the inflaton stress energy tensor T_{ab} are given by

$$G_{ab}\equiv R_{ab}-\frac{1}{2}Rg_{ab},$$
$$T_{ab}=\nabla_a\phi\nabla_b\phi+g_{ab}\Big[-\frac{1}{2}(\nabla\phi)^2-V(\phi)\Big], \tag{2.1.3}$$

with R_{ab} being the Ricci tensor of the spacetime metric. It is straightforward to see that the stress energy tensor for the inflaton field can be cast in the perfect fluid format,

$$T_{ab}=(\rho_{\rm inf}+p_{\rm inf})u_au_b+p_{\rm inf}g_{ab}, \tag{2.1.4}$$

where the energy density $\rho_{\rm inf}$, the pressure $p_{\rm inf}$, and the unit four-velocity dual vector u_a for the inflaton field are given by

$$\rho_{\rm inf}=-\frac{1}{2}(\nabla\phi)^2+V(\phi),$$
$$p_{\rm inf}=-\frac{1}{2}(\nabla\phi)^2-V(\phi),$$
$$u_a\equiv\frac{\nabla_a\phi}{\sqrt{-g^{ab}\nabla_a\phi\nabla_b\phi}}. \tag{2.1.5}$$

As we briefly mentioned in the introduction, the observational evidence mainly coming from the CMB suggests that the spacetime geometry expanded quasi-exponentially during inflation. This is equivalent to saying that $w_{\rm inf}\equiv p_{\rm inf}/\rho_{\rm inf}\approx-1$, or that $(\nabla\phi)^2\ll V(\phi)$ and $V(\phi)\sim$ constant during inflation [1]. We therefore demand that the inflaton's potential remain nearly flat during the observationally relevant period of inflation. It is customary to quantify the near-flatness conditions

[1] Or during the "observable inflation", which is a period of inflation associated with the freeze-out of the currently observable scales.

on the potential $V(\phi)$ using the following quantities [42]

$$\epsilon_V \equiv \frac{1}{2} m_p^2 \left(\frac{V_{,\phi}}{V} \right)^2 \ll 1,$$
$$|\eta_V| \equiv \left| m_p^2 \frac{V_{,\phi\phi}}{V} \right| \ll 1,$$
$$\left| m_p^{-2n} (V_{,\phi})^{n-1} \frac{d^{n+1} V}{d\phi^{n+1}} \right| \ll 1, \qquad \text{for } n > 2. \tag{2.1.6}$$

The quantities defined in Eq. (2.1.6) are called the potential slow-roll parameters, with ϵ_V and η_V being of primary relevance to the inflationary dynamics. Constructing a sufficiently flat potential $V(\phi)$ from a known UV completion at high energies has been the primary challenge of inflationary model-building in the past few decades [35]. We shall recount some of the main approaches and their shortcomings below.

I A unifying theme: The effective field theory of single field inflation

The inflationary action (2.1.1) incorporates the vast majority of single field inflation models that have been studied in the past few decades. However, in lieu of any preferred internal symmetry for the inflaton Lagrangian, excluding other possible terms such as the non-canonical kinetic terms appears to be unnatural from the field theoretic perspective [2]. On a similar note, its regime of validity and the type of operators that become relevant in the ultraviolet (UV) regime cannot be determined from the action (2.1.1). This fact complicates the study of the UV sensitivities of inflation as a low energy limit of a Planckian (or perhaps a sub-Planckian) theory.

Fortunately, much progress has been made in formulating a generic inflationary theory in the past few years with the advent of the effective field theory (EFT) of single field inflation [43, 44] [3][4]. This formulation is based on the fact that the spacetime geometry during the inflationary era enjoys an enhanced symmetry in the limit of ϕ becoming non-dynamical. In this limit, $V(\phi) \to \Lambda$ and the scale factor for the FRW metric (1.1.1) is given by $a(t) \propto e^{\sqrt{\Lambda/(3m_p^2)}t}$ [5], for which the metric has a timelike Killing vector field [6]. For inflation to end, this symmetry must be broken. The key

[2] From the effective field theory viewpoint, one must include all possible operators that are consistent with the assumed symmetries of the Lagrangian.

[3] See [45] for an earlier approach to formulating an EFT for inflation.

[4] This construction can also be extended to multifield models of inflation where the field space trajectory is described by a single degree of freedom.

[5] This scale factor is for de Sitter spacetime.

[6] In the static de Sitter coordinates, this extra Killing vector field is just the time translation $\vec{\partial}_t$.

observation here is that the "gradual" breaking of this external symmetry singles out a preferred arrow of time, and along with the homogeneity and isotropy of the FRW metric, a preferred foliation for the spacetime. For the case of a single scalar field that slowly rolls down its potential, the leaves of this preferred foliation are the surfaces of constant background inflaton field. The effective inflationary Lagrangian remains valid in the regime where this approximate spacetime symmetry exists then consists of all geometric operators that transform properly under the smooth spacetime diffeomorphisms that preserve this preferred foliation.

To gain more insight into this construction, it is convenient to choose spacetime coordinates (t, x^i) so that

$$\phi = \mu^2[t + \mathcal{T}(t, x^i)], \tag{2.1.7}$$

where μ is some constant of dimension mass, t is the time coordinate, and $\mathcal{T}$ encodes the linear fluctuations. The EFT operators are combinations of the spacetime tensors and scalars, such as the Riemann curvature tensor R_{abcd} and its various contractions, as well as the tensors and scalars that transform properly only under the foliation preserving diffeomorphisms, such as the intrinsic metric h_{ab} and the extrinsic curvature K_{ab} of the constant ϕ surfaces[7]. The EFT action including the relevant (dim < 4) and marginally relevant (dim $= 4$) operators is then given by [8]

$$\begin{aligned}
S_{\rm EFT}[g_{ab}, \phi] = \int d^4x\sqrt{-g}\ \Big[&\frac{m_p^2}{2}R + m_1^2\alpha_1(\phi)n^a n^b R_{ab} + \mathcal{R}^{[2]} + ... \\
&+\alpha_2(\phi)g^{ab}\nabla_a\phi\nabla_b\phi + \alpha_3(\phi)h^{ab}\nabla_a\phi\nabla_b\phi + m_4^4\alpha_4(\phi) + ... \\
&+m_5^3\alpha_5(\phi)h^{ab}K_{ab} + m_6^2\alpha_6(\phi)K_{ab}K^{ab} + m_7^2\alpha_7(\phi)\Big(K_{ab}h^{ab}\Big)^2 + \mathcal{K}^{[3]} + \mathcal{K}^{[4]} + ... \\
&+\mathcal{K}\mathcal{R} + \mathcal{K}^{[2]}\mathcal{R} + ...\Big],
\end{aligned} \tag{2.1.8}$$

[7] The unit normal dual vector to the surfaces of constant ϕ is given by $n_a \equiv \phi_{,a}/\sqrt{-\phi_{,a}\phi^{,a}}$. The intrinsic metric to these surfaces is then given by $h_{ab} = g_{ab} + n_a n_b$, and the extrinsic curvature is given by $K_{ab} = h^c{}_a h^d{}_b \nabla_c n_d$. Note that all of these vectors and tensors are expressed as functions of ϕ.

[8] The EFT action that is given in this form is actually not very useful! Various terms give redundant contributions at each order in perturbations. A far more suitable format for studying the dynamics of fluctuations is gained by the perturbative expansion of various operators around some fixed (inflationary) background, e.g. [43]

$$S_{\rm EFT} = \int d^4x\ \sqrt{-g}\Big[\frac{m_p^2}{2}R + m_p^2\dot{H}g^{tt} - m_p^2(3H^2 + \dot{H}) + \frac{1}{2}m_2^4\alpha_2(\delta g^{tt})^2 - \frac{1}{2}m_3^3\alpha_3\delta g^{tt}\delta K^a{}_a + ...\Big],$$

given in the gauge where $\mathcal{T} = 0$.

where $n_a \equiv \phi_{,a}/\sqrt{-\phi_{,a}\phi^{,a}}$ is the unit normal to the surfaces of constant ϕ, $\mathcal{R}^{[m]}/\mathcal{K}^{[m]}$ includes all terms that are products of m different R_{abcd}/K_{ab} tensors and its various contractions accompanied by dimensionless functions of ϕ, m_is are constants of dimension mass, and α_is are dimensionless functions of ϕ. The ellipses in the expression above indicate the irrelevant (dim > 4) operators, such as $\left(\nabla_a\phi\nabla^a\phi\right)^2$, $R\nabla_a\phi\nabla^a\phi$, $\left(K_{ab}h^{ab}\right)^3 R_{ab}n^a n^b$, etc. Also note that due to the adiabatic background evolution (i.e. $|\dot{H}| \ll H^2$)[9], one generally assumes that α_is are slowly varying functions of their arguments. All models of inflation that involve a single effective degree of freedom and whose background evolves adiabatically can be cast in the form of the EFT action given in Eq. (2.1.8).

This EFT formalism is not only useful for discovering new inflationary phenomenology, but it also provides a platform for studying the UV sensitivity of inflation. The contribution of an irrelevant operator to the measured or expected values of the inflationary background parameters and the cosmological observables can be used to constrain the relevant combination of α_i and the UV suppression scale for the given operator. It should be mentioned here that the UV cutoff $\mathcal{C}$ for the EFT of inflation should be in the range $H_{\rm inf} \lesssim \mathcal{C} \lesssim m_p$ as it must remain sub-Planckian and yet larger than the Hubble constant $H_{\rm inf}$ during inflation, the latter being due to the fact that the observable modes exit the Hubble horizon with frequencies $\omega \sim H_{\rm inf}$. We provide more details on the issue of the UV sensitivity of inflation in the following subsection.

II Radiative stability of the inflationary Lagrangian

The inflationary Lagrangian has a somewhat surprising UV sensitivity that has rendered the construction of a satisfactory inflationary model from a known UV theory a persistent theoretical challenge. To appreciate the significance of this challenge, let us briefly review the celebrated "eta-problem" of inflation [10].

Consider the potential term that appears in the EFT action (2.1.8), which we choose to

[9] This adiabaticity is a consequence of the near de Sitter symmetry of the spacetime during inflation.

[10] This problem was noticed while attempting to construct supergravity models that incorporate inflation [46]. For a detailed discussion of this problem see [36]

parametrize as [11]

$$V(\phi) \equiv m_4^4 \alpha_4(\phi) = m_4^4 \left[\sum_{i=0}^{4} \lambda_i \left(\frac{\phi}{m_4} \right)^i + \sum_{i>4}^{\infty} \beta_i \left(\frac{\phi}{\mu} \right)^i \right]. \tag{2.1.9}$$

Here we have taken μ to be the UV cutoff for the EFT action, which we require it to be in the range $H_{\rm inf} \lesssim \mu \lesssim m_p$. We assume that $\phi \sim m_4 < \mu$ during the observable inflation, for which $V(\phi) \sim m_4^4$. Moreover, we fine tune the coefficients λ_i to values small compared to m_4^2/m_p^2 in order to arrange for a sufficiently flat potential. As $\phi \to \mu$, we can no longer ignore the irrelevant operators. This is the strong coupling limit of the EFT. We initially assume that the coefficients β_i are order unity, so that as $\phi \to \mu$ all irrelevant operators give an order m_4^4 contribution to the potential.

The mass of the inflaton m_ϕ is on the order of $\sqrt{\lambda_2} m_4 < m_4^2/m_p$. The one loop correction from the ϕ^4 term to m_ϕ gives $\delta m_\phi \sim \sqrt{\lambda_4}\mu < m_4 \mu/m_p$. Therefore, $\delta m_\phi/m_\phi \sim \mu/m_4$, which can be large especially if $\mu \sim m_p$. The correction to the potential flatness parameter η_V defined in Eq. (2.1.6) becomes

$$\Delta\eta_V \sim \frac{\lambda_4 \mu^2 m_p^2}{m_4^4}. \tag{2.1.10}$$

If we take $\mu \sim m_p$ and assume $\lambda_4 \sim 0.01 m_4^2/m_p^2$, then $\Delta\eta_V \sim 0.01 m_p^2/m_4^2$ which can be large compared to unity if $m_4 \lesssim 10^{16}$GeV. Thus, the flatness of the classical potential can be easily spoiled by quantum loop corrections.

Aside from the issue of loop corrections, the monomials with $i > 4$ can contribute significantly to η_V. Even for $\phi \sim m_4$, the individual correction to $|\eta_V|$ for the ith monomial is

$$|^{(i)}\eta_V| = \left| i(i-1)\beta_i \frac{m_p^2}{\mu^2} \left(\frac{\phi}{\mu} \right)^{i-2} \right|, \qquad i > 4, \tag{2.1.11}$$

which can be order unity or larger if i is not too large and μ is not much larger than m_4.

The aforementioned problems admit two very different resolutions; either that the inflaton's Lagrangian is very finely tuned in order to keep all coefficients, including β_is, suppressed compared to unity under all radiative corrections, or that a physical mechanism exists that constrains the values of these coefficients and stabilizes them under generic loop corrections. The latter proposition has been explored by assuming a symmetry principle in the EFT of inflation that can keep loop corrections to the small EFT coefficients "technically natural". Whether such an EFT with the

[11] We do not claim that this particular form for $V(\phi)$ corresponds to a "generic" EFT potential. Our objective here is to demonstrate some aspects of the eta problem through an example.

assumed symmetry principle can result from a known UV theory, such as supergravity or string theory, is not currently understood [12].

2.2 Background dynamics

We now review the primary features of the background dynamics of the single field models of inflation. To begin, note that the background dynamics is governed by the Einstein equations (2.1.3) evaluated on the flat FRW background [13], which reduce to the Friedmann Eqs. (1.1.2) and (1.1.3)

$$\begin{aligned} 3m_p^2 H^2 &= \bar{\rho}_{\rm inf} = \frac{1}{2}\dot{\bar{\phi}}^2 + V, \\ m_p^2 \frac{\ddot{a}}{a} &= -\frac{1}{6}\Big(\bar{\rho}_{\rm inf} + 3\bar{p}_{\rm inf}\Big) = -\frac{1}{3}\Big[\dot{\bar{\phi}}^2 - V\Big], \end{aligned} \tag{2.2.1}$$

where the functions appearing with a bar are the background functions and we used Eq. (2.1.5). The two Friedmann equations can be combined to give a second order equation of motion for the background inflaton field

$$\ddot{\bar{\phi}} + 3H\dot{\bar{\phi}} + V_{,\bar{\phi}} = 0, \tag{2.2.2}$$

which is just the background Klein-Gordon equation given in the second line of (2.1.2). The middle term, $3H\dot{\bar{\phi}}$, is the well known the Hubble friction term. If the inflaton field starts with an appreciable kinetic energy, then the Hubble friction term dampens the kinetic energy, leading to the slow-roll regime. To see this, let us re-write the above Klein-Gordon equation as

$$\ddot{\bar{\phi}} + 3H\dot{\bar{\phi}} \pm \frac{\sqrt{2\epsilon_V}}{m_p} V = 0, \tag{2.2.3}$$

where we used ϵ_V given in Eq. (2.1.6). Should the potential be nearly flat, i.e. $\epsilon_V \ll 1$, and the kinetic energy dominate at the onset of inflation, we initially have

$$\ddot{\bar{\phi}} + 3H\dot{\bar{\phi}} \simeq 0 \Rightarrow \ddot{\bar{\phi}} \simeq \pm \frac{3}{\sqrt{6}m_p}\dot{\bar{\phi}}^2, \tag{2.2.4}$$

where the top (bottom) sign corresponds to $\dot{\bar{\phi}} < 0$ (> 0). In either case, it is implied that $|\dot{\bar{\phi}}|$ decreases with time. It can be shown that eventually $m_p H|\dot{\bar{\phi}}| \sim \sqrt{\epsilon_V} V$, which necessitates $\dot{\bar{\phi}}^2 \ll V$ for $\epsilon_V \ll 1$.

[12] See [35] for a recent review of the string theory models of inflation.

[13] We set the spatial curvature k of the FRW metric (1.1.1) to zero from now on.

As mentioned previously, the near flatness of the potential as formalised by the conditions (2.1.6) brings about the quasi-exponential expansion of the Universe during inflation. A different set of background slow-roll parameters can also be defined with respect to the changes of the Hubble parameter during inflation. The primary Hubble slow-roll parameters are given by

$$\epsilon \equiv -\frac{\dot{H}}{H^2} = \frac{\dot{\bar{\phi}}^2}{2m_p^2 H^2}, \qquad \eta \equiv \frac{\dot{\epsilon}}{H\epsilon} = 2\epsilon + 2\frac{\ddot{\bar{\phi}}}{\dot{\bar{\phi}}H},^{14} \tag{2.2.5}$$

where they are required to be

$$\epsilon \ll 1, \qquad |\eta| \ll 1. \tag{2.2.6}$$

We see in particular that $|\ddot{\bar{\phi}}/\dot{\bar{\phi}}H| \ll 1$, which is necessary to ensure that $\bar{\rho}_{\rm inf}$ remains dominated by the potential. Using Eqs. (2.2.1), (2.2.2), and (2.1.6), we can relate the two sets of slow-roll parameters by

$$\epsilon_V \approx \epsilon, \qquad \eta_V \approx 2\epsilon - \frac{\eta}{2}.^{15} \tag{2.2.7}$$

Note that the smallness of the potential (Hubble) slow-roll parameters implies the smallness of the Hubble (potential) slow-roll parameters. By convention, the end of inflation is the moment when $\epsilon_V \approx \epsilon = 1$ [16].

The other quantity that we will be interested in is the number of e-folds N_* from the end of inflation at $a_{\rm end}$ to the time a_* when the last of the currently observable modes left the Hubble horizon. This quantity can be computed using

$$N_* \equiv \int_{a_*}^{a_{\rm end}} \frac{da}{a} = \int_{t_*}^{t_{\rm end}} H dt = \int_{\bar{\phi}_*}^{\bar{\phi}_{\rm end}} \frac{H}{\dot{\bar{\phi}}} d\bar{\phi} \approx \int_{\bar{\phi}_*}^{\bar{\phi}_{\rm end}} \frac{d\bar{\phi}}{m_p\sqrt{2\epsilon_V}} = \int_{\bar{\phi}_*}^{\bar{\phi}_{\rm end}} \frac{V d\bar{\phi}}{m_p^2 V_{,\bar{\phi}}}. \tag{2.2.8}$$

The above equation is often used to determine the amount of scalar field excursion during inflation [17]. One usually assumes a value for N_* based on the minimum required amount of inflation depending on the energy scale at which it took place, and then executes the integral (2.2.8) from the value ϕ_* of the inflaton field when the observable modes are exiting the Hubble horizon to a final value $\bar{\phi}_{\rm end}$ that corresponds to the end of inflation.

[14] Some authors define the η parameter as $-\ddot{\bar{\phi}}/(\dot{\bar{\phi}}H)$ [11].

[15] If the definition in footnote 13 is used instead, then we have $\eta_V \approx \eta + \epsilon$.

[16] While this is true for most models of inflation, the end of inflation is more accurately marked by the moment when one of the slow-roll parameters becomes equal to one.

[17] That is to say, from the observable portion of inflation to the end of inflation.

I Examples of single field potentials

Let us review the basic properties of a number of inflationary potentials that have been used for constructing models of slow-roll inflation in the past [18]. As we will see in Sec. 2.6.III.a, there are regimes where these potentials are still observationally allowed.

- Monomial Potentials

  These potentials are expressed in the form

$$V(\bar{\phi}) = \lambda m_p^4 \left(\frac{\bar{\phi}}{m_p}\right)^n, \tag{2.2.9}$$

  where both λ and n are positive constants. The potential slow-roll parameters are

$$\epsilon_V = \frac{n}{2(n-1)}\eta_V = \frac{n^2}{2}\left(\frac{m_p}{\bar{\phi}}\right)^2, \tag{2.2.10}$$

  which implies that $\bar{\phi} > m_p$ during inflation [19]. The scalar field excursion between the observable inflation and the end of inflation is given by

$$\Delta\bar{\phi} = m_p\left[\sqrt{2nN_* + \frac{n^2}{2}} - \frac{n}{\sqrt{2}}\right]. \tag{2.2.11}$$

  It follows that the scalar field traverses a super-Planckian distance during inflation as long as $n \ll N_*$, even though $V^{1/4} < m_p$ at all times. Furthermore, as we will discuss in Sec. 2.5.I.a, the now well measured dimensionless scalar power spectrum can be used to constrain the value of the potential coupling λ. As we will later discuss, the CMB measurements suggest that

$$\frac{V^3}{V_{,\bar{\phi}}^2 m_p^6} \sim 10^{-7}, \tag{2.2.12}$$

  which for the case of monomial potentials gives $\lambda \sim 10^{-7} n^2 (m_p/\phi_*)^{n+2} < 10^{-7}$. [20]

- Hilltop Potentials

  In some models, inflation begins as the inflaton field rolls away from an unstable equilibrium in the potential V. In vicinity of the equilibrium point, the potential is approximated as

$$V(\bar{\phi}) = \Lambda_0\left[1 - \lambda\left(\frac{\bar{\phi}}{m_p}\right)^n + \mathcal{O}\left(\left[\frac{\bar{\phi}}{m_p}\right]^{n+1}\right)\right] \tag{2.2.13}$$

[18] For a more comprehensive discussion of the inflationary potentials see [35] and [36].

[19] Unless $n \ll 1$. Usually one assume $n \gtrsim \mathcal{O}(1)$.

[20] assuming that the value of the background inflaton field remains approximately constant for the range of scales probed by the CMB.

with Λ_0 and λ being some positive constants. While generically $n = 2$, for the following analysis we take $n > 2$. Clearly, $\lambda(\bar{\phi}/m_p)^n \ll 1$ at the onset of inflation. The potential slow-roll parameters for this potential are

$$\epsilon_V = \frac{\lambda^2 n^2}{2} \frac{\left(\frac{\bar{\phi}}{m_p}\right)^{2n-2}}{\left[1 - \lambda\left(\frac{\bar{\phi}}{m_p}\right)^n\right]^2}, \qquad \eta_V = -\lambda n(n-1) \frac{\left(\frac{\bar{\phi}}{m_p}\right)^{n-2}}{1 - \lambda\left(\frac{\bar{\phi}}{m_p}\right)^n}. \tag{2.2.14}$$

A class of inflationary models, called the small field models of inflation, correspond to $\bar{\phi} \lesssim m_p$ and $\lambda \lesssim 1$ throughout inflation. For these models $\epsilon_V \ll |\eta_V|$ and inflation ends when $|\eta_V| = 1$, from which it follows that $\lambda \gtrsim [n(n-1)]^{-1}$.

- Starobinsky Potential

  This particular potential, which emerged in the context of modifications to gravity in the early Universe [19], is given by

$$V(\bar{\phi}) = \frac{m_p^4}{4\lambda}\left(1 - e^{-\sqrt{\frac{2}{3}}\frac{\bar{\phi}}{m_p}}\right)^2, \tag{2.2.15}$$

  where $\lambda > 0$. The potential slow-roll parameters are given by

$$\epsilon_V = \frac{4}{3}\left(\frac{e^{-\sqrt{\frac{2}{3}}\frac{\bar{\phi}}{m_p}}}{1 - e^{-\sqrt{\frac{2}{3}}\frac{\bar{\phi}}{m_p}}}\right)^2, \qquad \eta_V = \frac{4}{3} e^{-\sqrt{\frac{2}{3}}\frac{\bar{\phi}}{m_p}} \frac{2e^{-\sqrt{\frac{2}{3}}\frac{\bar{\phi}}{m_p}} - 1}{\left[1 - e^{-\sqrt{\frac{2}{3}}\frac{\bar{\phi}}{m_p}}\right]^2}. \tag{2.2.16}$$

  Inflationary solutions then require $\bar{\phi} \gtrsim m_p$ for the potential to remain nearly flat. Inflation ends when $\epsilon_V = 1$, which occurs for $\bar{\phi}/m_p \approx 0.94$. It follows from the integral (2.2.8) that $\Delta\bar{\phi} \approx 4.5\ m_p$ for $N_* \sim 60$. Therefore, much like the monomial potentials, the inflationary model based on the Starobinsky potential requires a super-Planckian field excursion. Moreover, it follows from Eq. (2.2.12) that $\lambda \sim 10^{10}$ for $N_* \sim 60$.

- Axion Monodromy Potential

  Some string theory motivated models of inflation are based on the inflaton field being a pseudoscalar axion [47]. The instanton-induced corrections to the inflaton's potential take the form

$$V(\bar{\phi}) = \frac{V_0}{2}\left[1 - \cos\left(\frac{\bar{\phi}}{f}\right)\right], \tag{2.2.17}$$

for some positive constant f which is typically required to be larger than m_p. The potential slow-roll parameters in this case are given by

$$\epsilon_V = \frac{m_p^2}{2f^2}\frac{\sin\left(\frac{\bar{\phi}}{f}\right)^2}{\left[1-\cos\left(\frac{\bar{\phi}}{f}\right)\right]^2}, \qquad \eta_V = \frac{m_p^2}{f^2}\frac{\cos\left(\frac{\bar{\phi}}{f}\right)}{1-\cos\left(\frac{\bar{\phi}}{f}\right)}. \tag{2.2.18}$$

A short computation using the integral (2.2.8) reveals that a field excursion $\Delta\bar{\phi} \sim \mathcal{O}(1)m_p$ is expected for $N_* \sim 60$ and $f \sim \mathcal{O}(1)m_p$. Note that for $f \gg \bar{\phi}$ we simply recover the $\bar{\phi}^2$ monomial potential at the leading order in expansions.

2.3 Inflation and the generation of primordial perturbations

The main success of the standard theory of inflation is not solving the classical cosmological problems, which it does by design, but rather providing a natural framework for the generation of the primordial fluctuations with properties that are in excellent agreement with CMB and LSS observations. In this section, we will review the mechanism by which such perturbations are generated during inflation.

I Basics of the inflationary perturbation theory

To study the dynamics of perturbations around some fixed inflationary background, we expand the action (2.1.1) in powers of the linear metric and inflaton perturbation functions defined by

$$g_{ab}(t,x^i) = \bar{g}_{ab}(t) + \delta g_{ab}(t,x^i), \qquad \phi(t,x^i) = \bar{\phi}(t) + \delta\phi(t,x^i). \tag{2.3.1}$$

There are a total of ten metric perturbation functions that are classified based on their transformation properties under rotation: four scalar functions, four vector functions, and two tensor functions. We ignore the vector perturbations in the inflationary setting as they only give rise to decaying modes. The four scalar functions perturb the flat FRW metric as

$$ds^2 = -(1+2\Phi)dt^2 + 2aB_{,i}dtdx^i + a^2\Big[(1-2\Psi)\delta_{ij} + 2E_{,ij}\Big]dx^i dx^j. \tag{2.3.2}$$

Note that the scalar perturbation functions Φ, B, Ψ, E do not transform as spacetime scalars [21]. The tensor perturbations are given by

$$ds^2 = -dt^2 + a^2\Big[\delta_{ij} + h_{ij}\Big]dx^i dx^j, \tag{2.3.3}$$

[21] See App.B for a review of their transformation properties.

where the spatial metric h_{ij} is traceless and transverse, i.e. [22]

$$h^i{}_i = 0, \qquad \partial_i h^i{}_j = 0, \tag{2.3.4}$$

giving it two independent degrees of freedom. Unlike the scalar perturbation functions, h_{ij} is independent of the coordinate gauges.

II Generation of scalar perturbations

In the case of single field inflation, the perturbations to the inflaton couple to the scalar perturbations of the metric. Due to the rotational invariance of the background geometry, the scalar perturbations do not couple to the tensor perturbations at linear order. Additionally, not all scalar perturbation functions are propagating degrees of freedom; some can be eliminated by a judicious choice of coordinate gauge and some can be integrated out using the linearly-perturbed Einstein constraint equations.

To understand the dynamics of the scalar perturbations at the linear order, it is convenient to expand the action (2.1.1) to second order in the scalar perturbations [23]. The second order action can be shown to be a functional of one gauge invariant scalar field [48, 49, 50]

$$\delta_2 S_{\text{scalar}} = \frac{1}{2}\int d\tau\ d^3x\ \left[\nu_{,\tau}^2 - \delta^{ij}\partial_i\nu\partial_j\nu + \frac{\zeta''}{\zeta}\nu^2\right], \tag{2.3.5}$$

where the Mukhanov variable ν is defined by

$$\nu \equiv a\delta\phi + \zeta\Psi, \qquad \zeta \equiv \frac{a\dot{\phi}}{H}, \tag{2.3.6}$$

and we have switched to using the conformal time $\tau \equiv \int dt/a$ in the action (2.3.5) [24]. Varying Eq. (2.3.5) with respect to ν, we arrive at the Mukhanov-Sasaki equation

$$\nu'' - \partial_i\partial^i\nu - \frac{\zeta''}{\zeta}\nu = 0, \tag{2.3.7}$$ [25]

[22] Indices on h_{ij} are uppered using δ^{ij}.

[23] The first order action vanishes by virtue of the background equations.

[24] Note that $a \sim e^{Ht} \sim -(H\tau)^{-1}$ during slow-roll inflation. The limit $\tau \to -\infty$ is the big bang singularity and $\tau \to 0$ is the limit of infinite inflation. In reality, inflation must end for a value τ_{end} of τ with $\tau_{\text{end}} < 0$.

[25] More generally, this is

$$\nu'' - c_s^2\partial_i\partial^i\nu - \frac{\zeta''}{\zeta}\nu = 0$$

because the EFT of inflation allows for a sound speed c_s to be different from 1. See [37].

where prime denotes differentiation with respect to τ and $\partial_i\partial^i \equiv \delta^{ij}\partial_i\partial_j$. After decomposing ν in terms of its Fourier modes by $\nu(\tau,\mathbf{k}) = \int d^3x/(2\pi)^{3/2} e^{-i\mathbf{k}.\mathbf{x}}\nu(\tau,\mathbf{x})$[26], Eq. (2.3.7) becomes

$$\nu'' + \left[k^2 - \frac{\zeta''}{\zeta}\right]\nu = 0, \tag{2.3.8}$$

with $k^2 = \delta^{ij}k_ik_j$. Equation (2.3.8) can be solved analytically for slow-roll inflation by approximating frequencies $\omega(\tau)$ given by $\omega^2 = k^2 - \zeta''/\zeta$ using

$$\frac{\zeta''}{\zeta} = \frac{2 + 3\epsilon + \frac{3}{2}\eta}{\tau^2} + \mathcal{O}(2), \tag{2.3.9}$$

where $\mathcal{O}(2)$ denotes corrections that are second order or higher in the slow-roll parameters. Using this approximation, the solution to Eq. (2.3.8) becomes

$$\nu = c_1\sqrt{-k\tau}H_l^{(1)}(-k\tau) + c_2\sqrt{-k\tau}H_l^{(2)}(-k\tau), \tag{2.3.10}$$

where $H^{(1,2)}$ are the Hankel functions of first and second kind and $l = \sqrt{9 + 12\epsilon + 6\eta}/2$.

The Mukhanov variable ν can be re-scaled to define another gauge invariant quantity $\mathcal{R} \equiv \nu/\zeta$ which is known as the comoving curvature perturbation. This quantity can be proven to be conserved on the superhorizon scales [27] in some models of inflation, including single field models [51, 52]. Therefore, inflation can still be predictive despite the absence of a preferred model for reheating as long as we are interested in probing the dynamics of the perturbation modes that re-enter the Hubble horizon during and after the radiation dominated era.

III Generation of primordial gravitational waves

The tensor perturbations defined in (2.3.3) contain propagating degrees of freedom, which are inflationary gravitational waves. The action (2.1.1) can be expanded to second order in h_{ij} giving the leading order action for the inflationary gravitational waves

$$\delta_2 S_{\text{tensor}} = \frac{m_p^2}{8}\int d\tau\ d^3x\ a^2\left[h_{ij,\tau}h^{ij}_{,\tau} - \delta^{kl}\partial_l h_{ij}\partial_k h^{ij}\right]. \tag{2.3.11}$$

The equation of motion for the tensor fields h_{ij} are then given by

$$\Box h_{ij} = 0. \tag{2.3.12}$$

[26] Boldface notation is used to denote spatial vectors.

[27] A super(sub)horizon mode has physical wavelength λ larger (smaller) than H_{inf}^{-1}.

Just as in the case of scalar perturbations, we express the above equation in momentum space by writing $h_{ij}(\tau, \mathbf{x}) = \sum_{A=\{+,\times\}} \int d^3k/(2\pi)^{3/2} e^{i\mathbf{k}.\mathbf{x}} 2h^A(\tau, \mathbf{k}) e^A_{ij}(\hat{\mathbf{k}})/(m_p a)$ [28], where the polarization tensors satisfy $e^A_{ij} e^{A'\ ij} = 2\delta^{A'A}$. The result is

$$h^{A\prime\prime} + \left[k^2 - \frac{a''}{a}\right] h^A = 0 \tag{2.3.13}$$

for each polarization. Just as in the case of scalar perturbations, we can find approximate analytic solutions for Eq. (2.3.13) by writing

$$\frac{a''}{a} = \frac{2 + 3\epsilon}{\tau^2} + \mathcal{O}(2), \tag{2.3.14}$$

which yields solution of the form (2.3.10) but with $l = \sqrt{9 + 12\epsilon}/2$.

2.4 Subhorizon adiabaticity of the perturbations

The inflationary scalar and tensor perturbation modes exhibit adiabaticity in the limit of becoming deeply subhorizon. Formally, the scalar and tensor modes are said to become adiabatic when the frequencies

$$\omega_{\rm s}(\tau, k) \equiv \sqrt{k^2 - \frac{\zeta''(\tau)}{\zeta(\tau)}}, \qquad \omega_{\rm t}(\tau, k) \equiv \sqrt{k^2 - \frac{a''(\tau)}{a(\tau)}} \tag{2.4.1}$$

of the scalar and tensor perturbation modes satisfy

$$\left|\frac{\omega'_{\rm s}(\tau, k)}{\omega_{\rm s}(\tau, k)^2}\right| \ll 1, \qquad \left|\frac{\omega'_{\rm t}(\tau, k)}{\omega_{\rm t}(\tau, k)^2}\right| \ll 1, \tag{2.4.2}$$

for a given k [29]. It follows from Eqs. (2.3.14) and (2.3.9) that if slow-roll inflation continues in the limit of $\tau \to -\infty$, or equivalently as $a \to 0$, we have

$$\left|\frac{\omega'_{\rm s}(\tau, k))}{\omega_{\rm s}(\tau, k)^2}\right| \sim \left|\frac{\omega'_{\rm t}(\tau, k))}{\omega_{\rm t}(\tau, k)^2}\right| \sim \left|\frac{1}{\tau^3 k^3}\right| \to 0. \tag{2.4.3}$$

Therefore, any given inflationary scalar or tensor perturbation mode is in an adiabatic regime should we go sufficiently deep into the past. Intuitively, their dynamics effectively decouples from the background evolution and they propagate as they do in the flat spacetime as the hierarchy between their physical wavelength and the Hubble radius $H^{-1}_{\rm inf}$ increases.

We will see in the next section that this property of the inflationary perturbations underlies the existence of a natural choice of vacuum, which render the inflationary predictions unique.

[28] The numerical factor $2/m_p$ is added to render the overall pre-factor of action (2.3.11) 1/2. This convention is used for convenience when we quantize the inflationary gravitational waves.

[29] Some also require $|\omega''/\omega^3| \ll 1$. For our purposes, the conditions given in Eq. (2.4.2) are sufficient.

I Natural choice of vacuum and the quantizations scheme

We now review how the adiabaticity of scalar and tensor perturbation modes during inflation leads to the construction of a natural choice of vacuum. Though this procedure is only reviewed for scalar perturbation modes, the same exact analysis can be replicated for tensor perturbations. For more discussions and generalization of this formalism, we recommend consulting [53, 54].

We begin by expressing the Mukhanov field ν as

$$\nu(\tau, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^{3/2}} \nu(\tau, \mathbf{k}) e^{i\mathbf{k}.\mathbf{x}}, \tag{2.4.4}$$

where $\nu(\tau, \mathbf{k}) \equiv a_{\mathbf{k}} \nu(\tau, k) + a^*_{-\mathbf{k}} \nu^*(\tau, k)$ with $a_{\mathbf{k}}$ being a $\mathbf{k}$ dependent constant and $\nu(\tau, k)$ being a solution to Eq. (2.3.8)[30] [31]. The conjugate momentum $\pi(\tau, \mathbf{x})$ to the Mukhanov field is given by

$$\pi(\tau, \mathbf{x}) \equiv \frac{\delta L_{\text{scalar}}(\tau)}{\delta \nu'(\tau, \mathbf{x})} = \nu'(\tau, \mathbf{x}), \tag{2.4.5}$$

where the scalar Lagrangian L_{scalar} is defined using $\delta_2 S_{\text{scalar}} = \int d\tau L_{\text{scalar}}$ in Eq.(2.3.5).

Upon quantization, we promote the classical fields ν and π to quantum operators $\hat{\nu}$ and $\hat{\pi}$ that satisfy the following equal time commutation relations

$$[\hat{\nu}(\tau, \mathbf{x}), \hat{\pi}(\tau, \mathbf{y})] = i\delta(\mathbf{x} - \mathbf{y}), \qquad [\hat{\nu}(\tau, \mathbf{x}), \hat{\nu}(\tau, \mathbf{y})] = [\hat{\pi}(\tau, \mathbf{x}), \hat{\pi}(\tau, \mathbf{y})] = 0. \tag{2.4.6}$$

The quantization procedure also applies to the previously regarded constants $a_{\mathbf{k}}$ and their complex conjugates $a^*_{\mathbf{k}}$ which now become the quantum creation and annihilation operators $\hat{a}_{\mathbf{k}}$ and $\hat{a}^\dagger_{\mathbf{k}}$ satisfying

$$[\hat{a}_{\mathbf{k}}, \hat{a}^\dagger_{\mathbf{k}'}] = \frac{1}{\langle \nu, \nu \rangle} \delta(\mathbf{k} - \mathbf{k}'), \qquad [\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}] = [\hat{a}^\dagger_{\mathbf{k}}, \hat{a}^\dagger_{\mathbf{k}'}] = 0, \tag{2.4.7}$$

where $\langle \nu, w \rangle \equiv i(\nu' w^* - \nu w^{*\prime})$ for the solutions $\nu(\tau, k)$ and $w(\tau, k)$ to Eq. (2.3.8). The quantity $\langle \nu, \nu \rangle$ is conserved by virtue of Eq. (2.3.8) and is conventionally set to one, after which ν is called a mode function. The quantum vacuum state $|0_{\mathbf{k}}\rangle$ for the mode $\mathbf{k}$ is then defined by

$$\hat{a}_{\mathbf{k}} |0_{\mathbf{k}}\rangle = 0. \tag{2.4.8}$$

Note that this definition of the vacuum state $|0_{\mathbf{k}}\rangle$ depends on the choice of solution $\nu(\tau, \mathbf{k})$, which can be arbitrary except for the condition $\langle \nu, \nu \rangle = 1$. In fact, the operators $\hat{a}_{\mathbf{k}}$ and $\hat{a}^\dagger_{\mathbf{k}}$ are formally

[30] Given that Eq.(2.3.8) is second order, ν and ν^* span the space of solutions.

[31] We require $\langle \nu, \nu \rangle > 0$. See under Eq. (2.4.7) for definition of the inner product $\langle ... \rangle$.

defined by

$$\hat{a}_{\mathbf{k}} \equiv \big\langle \nu(\tau,k), \hat{\nu}(\tau,\mathbf{k}) \big\rangle, \qquad \hat{a}^{\dagger}_{\mathbf{k}} \equiv -\big\langle \nu^{*}(\tau,k), \hat{\nu}(\tau,-\mathbf{k}) \big\rangle, \tag{2.4.9}$$

which implies that their definitions are dependent on the choice of mode function $\nu(\tau,k)$.

In certain circumstances, however, symmetries of the spacetime single out preferred vacua. In flat spacetime, for instance, the preferred vacua for a Klein-Gordon field are invariant under the action of the unitary representation of the time translation operator, i.e.

$$\hat{U}(\tau,\tau_i)|0_{\mathbf{k}}\rangle \equiv e^{-i\int_{\tau_i}^{\tau} d\tau \hat{H}}|0_{\mathbf{k}}\rangle = e^{-i\omega(k)\Delta\tau}|0_{\mathbf{k}}\rangle^{32}, \tag{2.4.10}$$

where $\hat{H}$ is the quantum Hamiltonian operator and $\omega(k)$ is some positive constant. Note that in particular, the vacuum state $|0_{\mathbf{k}}\rangle$ must be an eigenstate of $\hat{H}$ with a positive eigenvalue and that $\langle 0_{\mathbf{k}}|\hat{H}|0_{\mathbf{k}}\rangle$ is the minimum of all non-vanishing matrix elements of $\hat{H}(\mathbf{k})$.

In the context of inflation, the existence of a preferred [33] vacuum state for the perturbations is guaranteed in the limit of exact adiabaticity, i.e. as $\omega(k) \to k$. To see this, recall that the Hamiltonian operator for the inflationary scalar perturbations is given by

$$\begin{aligned}\hat{H}_{\text{scalar}}(\tau) &\equiv \int d^3x\ \hat{\pi}(\tau,\mathbf{x})\hat{\nu}'(\tau,\mathbf{x}) - \hat{L}_{\text{scalar}}(\tau) = \frac{1}{2}\int d^3x\ \left[\hat{\pi}^2 + \partial_i\hat{\nu}\partial^i\hat{\nu} - \frac{\zeta''}{\zeta}\hat{\nu}^2\right] \\ &= \frac{1}{2}\int d^3k \left[\hat{\pi}(\tau,\mathbf{k})\hat{\pi}(\tau,-\mathbf{k}) + \Big(k^2 - \frac{\zeta''}{\zeta}\Big)\hat{\nu}(\tau,\mathbf{k})\hat{\nu}(\tau,-\mathbf{k})\right],\end{aligned} \tag{2.4.11}$$

which in the limit of $|\zeta''/\zeta| \ll k^2$ reduces to that of a free massless Klein-Gordon field propagating in the flat spacetime. It follows without much difficulty that the vacuum state $|0_{\mathbf{k}}\rangle$ is an eigenstate of $\hat{H}_{\text{scalar}}$ in this limit if and only if $\nu \to \sqrt{1/(2k)}e^{-ik\tau}$ [34]. This particular initial condition fully determines the mode function at all times during inflation,

$$\nu = \sqrt{-\frac{\pi\tau}{4}}H_l^{(1)}(-k\tau) \tag{2.4.12}$$

which follows from matching the solution provided in Eq. (2.3.10) to the expected initial condition at $k|\tau| \to \infty$. Therefore, a preferred set of mode functions and vacua can be selected in the context of inflation thanks to the adiabaticity property of perturbation modes. These vaccua are known as the Bunch-Davies vacua and we denote them by $|0\rangle_{\text{BD}}$ [55].

[32] We ignored the $\delta(0)$ factor that emerges as a result of taking the continuum limit of space.
[33] Preferred in the sense of invariance under the action of time-translation operator.
[34] Recall that we required $\langle \nu,\nu\rangle = 1$, which implies that $e^{+ik\tau}$ modes do not contribute.

II The inflationary trans-Planckian problem

In defining the Bunch-Davies vacua we relied on choosing the initial conditions for a perturbation mode $\mathbf{k}$ in the limit of $k|\tau| \to \infty$. Even if we are content with defining the Bunch-Davies vacua at an initial time τ_i for which $k|\tau_i| \gg 1$ but finite, the physical frequency $\omega_{\rm phys}(k)$ of the given perturbation mode might be blueshifted to values comparable to or larger than m_p, depending on the energy scale of inflation and the value of τ_i. This poses a trans-Planckian problem for inflation, since our formalism for evolving the inflationary perturbations are only valid as long as $\omega_{\rm phys} \lesssim m_p$. It is interesting to note that the observationally relevant modes today could have had sub-Planckian wavelengths at the onset of inflation. Assuming for simplicity that inflation was immediately succeeded by the radiation dominated era, for a mode with a wavelength λ_0 today to have a wavelength larger than the Planck length at the onset of inflation we must have

$$(\lambda_0 H_0)\frac{\Lambda_0^{1/4}\Omega_r^{1/4}}{H_0\Lambda_{\rm inf}^{1/4}e^N} \gtrsim m_p^{-1} \Rightarrow \lambda_0 \gtrsim 10^{-58}\Big(\frac{\Lambda_{\rm inf}}{\Lambda_0}\Big)^{1/4} e^N H_0^{-1}, \tag{2.4.13}$$

where $\Lambda_0^{1/4} \sim 10^{-3}$eV is the current energy scale of the Universe, $\Lambda_{\rm inf}^{1/4}$ is the energy scale of inflation, and N is the number of e-folds of inflation. For a GUT-scale inflation model and $N \approx 60$ e-folds, the above inequality gives $\lambda_0 \gtrsim 10^{-4}H_0^{-1} \sim 100$ kpc. Longer wavelength modes would have been Planckian at the onset of inflation for a longer duration of inflation.

There have been several attempts to estimate the deviations from the standard inflationary predictions that result from trans-Planckian modification to dynamics [35]. One either assumes a specific Planck theory within which the pertinent corrections are computed [36], or takes a phenomenological approach where the corrections are codified into the deviations of the initial state for the inflationary perturbations from the natural Bunch-Davis vacua [37]. We will return to this latter topic in chapter 3.

2.5 Inflationary correlation functions

As mentioned previously, the most attractive aspect of inflation is that it provides a natural mechanism for generating the primordial perturbations that seed the observable structures today.

[35] See [56] for a review of various approaches to this problem and a summary of their conclusions.
[36] See [57] for an analysis using the Horava-Lifshitz gravity.
[37] See [58] for the first analysis of the trans-Planckian problem using this method.

By virtue of being stochastic, these perturbations are fully characterized by their correlation functions. We will review the fundamental properties of these correlation functions in this section.

I Correlation functions of the scalar perturbations

The scalar perturbation function of interest is the gauge-invariant comoving curvature perturbations given by [38]

$$\mathcal{R} \equiv \Psi + \frac{a}{\zeta}\delta\phi = \frac{\nu}{\zeta}, \tag{2.5.1}$$

where ν and ζ are defined in Eq. (2.3.6). In addition to being gauge-invariant, any given Fourier mode of $\mathcal{R}$ is proven to be conserved on the superhorizon scales in all single field models of inflation. In fact, it can be seen from Eq. (2.4.12) that

$$\mathcal{R}(\tau, k) \to i\left[\frac{1}{4k^3\epsilon(\tau_k)}\right]^{1/2}\frac{H(\tau_k)}{m_p}\,^{39}, \tag{2.5.2}$$

in the superhorizon limit, i.e. when $|k\tau| \ll 1$. Note that we have computed the background functions at a time τ_k, when $k = aH$. It follows that $\mathcal{R}$ is a natural candidate for the initial conditions of the metric and matter perturbation modes that re-enter the Hubble horizon and become dynamical later on in the course of the Universe's expansion [40].

We are interested in computing the equal-time n-point quantum expectation values of the quantized modes of $\mathcal{R}$, i.e.

$$\left\langle\hat{\mathcal{R}}(\tau_0, \mathbf{k}_1)...\hat{\mathcal{R}}(\tau_0, \mathbf{k}_n)\right\rangle \equiv \frac{\langle\Omega|\hat{\mathcal{R}}(\tau_0, \mathbf{k}_1)...\hat{\mathcal{R}}(\tau_0, \mathbf{k}_n)|\Omega\rangle}{\langle\Omega|\Omega\rangle}, \tag{2.5.3}$$

where τ_0 is a conformal time at which all modes are superhorizon, which is usually taken to be $\tau_{\rm end}$. The state $|\Omega\rangle$ is typically the vacuum state of the full theory, which differs from the Bunch-Davies vacuum if one considers interactions. The above quantum correlation function is most conveniently computed in the interaction picture,

$$\left\langle\hat{\mathcal{R}}(\tau_0, \mathbf{k}_1)...\hat{\mathcal{R}}(\tau_0, \mathbf{k}_n)\right\rangle = \frac{\langle\Omega|\hat{U}^\dagger(\tau_0, \tau_i)\hat{\mathcal{R}}^I(\tau_0, \mathbf{k}_1)...\hat{\mathcal{R}}^I(\tau_0, \mathbf{k}_n)\hat{U}(\tau_0, \tau_i)|\Omega\rangle}{\langle\Omega|\Omega\rangle}, \tag{2.5.4}$$

[38] See App. B for the general definition of $\mathcal{R}$.

[39] We are assuming that $|\epsilon|, |\eta| \ll 1$ as the mode k exits the Hubble horizon.

[40] The exact mechanism through which the initial conditions for the metric and matter perturbations are derived from $\mathcal{R}$ can be found in many standard references, such as [14].

where

$$\hat{U}(\tau_0,\tau_i) \equiv T\ \exp\left\{ -i\int_{\tau_i}^{\tau_0} d\tau' \hat{H}^I_{\rm int}(\tau')\right\},$$
$$\hat{\mathcal{R}}^I(\tau,\mathbf{k}) \equiv \frac{1}{\zeta(\tau)}\hat{\nu}^I(\tau,\mathbf{k}) = \hat{a}^I_{\mathbf{k}}\frac{\nu(\tau,k)}{\zeta(\tau)} + \hat{a}^{\dagger I}_{-\mathbf{k}}\frac{\nu^*(\tau,k)}{\zeta(\tau)}. \tag{2.5.5}$$

The quantum time evolution operator $\hat{U}$ involves a time-ordered product of the interactions Hamiltonian $\hat{H}^I_{\rm int}$ as well as some conformal time τ_i when we start evolving the perturbations modes [41]. The mode functions of the interaction picture $\hat{\mathcal{R}}$ are specified by the solution to the linear Mukhanov-Sasaki equation (2.3.8), and their associated operators $\hat{a}_{\mathbf{k}}$ annihilate the Bunch-Davies vacua.

Finally, note that $|\Omega\rangle$ can be expressed in terms of the Bunch Davies vacuum $|0\rangle_{\rm BD}$ for a judicious choice of integration limits on $\hat{U}$. Indeed, it can be shown without much difficulty that [42]

$$\lim_{\tau\to-\infty[1-i\epsilon]}\langle\Omega|0\rangle_{\rm BD}\ e^{-i\int_\tau^{\tau_i}\hat{H}}|\Omega\rangle = \lim_{\tau\to-\infty[1-i\epsilon]} e^{-i\int_\tau^{\tau_i}\hat{H}}|0\rangle_{\rm BD}$$
$$\Rightarrow |\Omega\rangle = \lim_{\tau\to-\infty[1-i\epsilon]}\Big(\langle\Omega|0\rangle_{\rm BD}\Big)^{-1} e^{-i(\tau_i-\tau)[\omega_0-\hat{\omega}]}\ \hat{U}(\tau_i,\tau)|0\rangle_{\rm BD}, \tag{2.5.6}$$

where $\hat{H}$ is the Hamiltonian of the full theory and ω_0 and $\hat{\omega}$ are the vacuum eigenvalues of the free and full Hamiltonians respectively. Note that we also made use of the near time-independence of the Hamiltonian for the modes in the deep subhorizon limit. Taking this simplification into account, the n-point correlation function of $\hat{\mathcal{R}}$ in the interaction picture becomes

$$\big\langle\hat{\mathcal{R}}(\tau_0,\mathbf{k}_1)...\hat{\mathcal{R}}(\tau_0,\mathbf{k}_n)\big\rangle \equiv \frac{{}_{\rm BD}\big\langle 0\big|\hat{U}^\dagger\big(\tau_0,-\infty[1-i\epsilon]\big)\hat{\mathcal{R}}^I(\tau_0,\mathbf{k}_1)...\hat{\mathcal{R}}^I(\tau_0,\mathbf{k}_n)\hat{U}\big(\tau_0,-\infty[1-i\epsilon]\big)\big|0\big\rangle_{\rm BD}}{{}_{\rm BD}\big\langle 0\big|\hat{U}^\dagger\big(\tau_i,-\infty[1-i\epsilon]\big)\hat{U}\big(\tau_i,-\infty[1-i\epsilon]\big)\big|0\big\rangle_{\rm BD}}, \tag{2.5.7}$$

which is straightforward to compute once $\hat{H}^I_{\rm int}$ is found. We shall drop the superscript I, the subscript BD, and the normalization constant in the denominator from here on.

Below we will discuss the 2-point and 3-point correlation functions of $\hat{\mathcal{R}}$, ignoring the higher order ones as they are highly suppressed and unlikely to be observationally relevant in the near future.

[41] All modes are deeply subhorizon at τ_i.

[42] This method of expressing the interacting vacuum in terms of the free vacuum by deforming the contour of integration is a standard practice in QFT. See e.g. [59].

I.a Scalar power spectrum

The first non-trivial correlation function that does not vanish in the linear regime is the 2-point correlation function [43]. If $\hat{\mathcal{R}}$ has a Gaussian statistics, then all of its correlation functions are given by products of its 2-point correlation function [44]. The leading 2-point correlation function is given by

$$\langle 0|\hat{\mathcal{R}}(\tau_0, \mathbf{k_1})\hat{\mathcal{R}}(\tau_0, \mathbf{k_2})|0\rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2)\mathcal{P}_{\mathcal{R}}(k_1), \tag{2.5.8}$$

where $\mathcal{P}_{\mathcal{R}}$ is called the power spectrum of the comoving curvature perturbations, or simply the scalar power spectrum. It follows from the definition of $\hat{\mathcal{R}}$ and Eq. (2.5.2) that

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{H(\tau_k)^2}{4k^3\epsilon(\tau_k)m_p^2}. \tag{2.5.9}$$

Conventionally, the dimensionless scalar power spectrum is defined via [45]

$$\Delta^2_{\mathcal{R}}(k) \equiv \frac{k^3}{2\pi^2}\mathcal{P}_{\mathcal{R}}(k) = \frac{H(\tau_k)^2}{8\pi^2\epsilon(\tau_k)m_p^2}. \tag{2.5.10}$$

$\Delta^2_{\mathcal{R}}$ has a slight scale dependence that vanishes in the limit of exact de Sitter symmetry for the background. In fact, $\Delta^2_{\mathcal{R}}$ is typically expressed as

$$\Delta^2_{\mathcal{R}}(k) = A_s\left(\frac{k}{k_*}\right)^{n_s - 1 + \frac{1}{2}\alpha_s \ln(k/k_*)}, \tag{2.5.11}$$

where k_* is some fiducial scale, A_s is the value of $\Delta^2_{\mathcal{R}}$ at k_*, and n_s and α_s are the spectral tilt and the running of the spectral tilt of $\Delta^2_{\mathcal{R}}$. To leading order in the slow-roll parameters, n_s and α_s are given by [35]

$$n_s = 1 - 2\epsilon - \eta, \qquad \alpha_s = 16\epsilon^2 - 6\epsilon\eta + \eta\chi, \tag{2.5.12}$$

where $\chi \equiv \dot{\eta}/(\eta H)$. Note that $\alpha_s \sim (n_s - 1)^2 \ll 1$. In Sec. 2.6.III we will quote the most recent measurements and bounds on A_s, n_s, and α_s.

I.b Scalar bispectrum and non-Gaussianities

The leading non-Gaussianities of $\hat{\mathcal{R}}$ are provided by its non-vanishing 3-point correlation function. The 3-point correlation function is given by

$$\left\langle \hat{\mathcal{R}}(\tau_0, \mathbf{k}_1)\hat{\mathcal{R}}(\tau_0, \mathbf{k}_2)\hat{\mathcal{R}}(\tau_0, \mathbf{k}_3)\right\rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)\mathcal{B}_{\mathcal{R}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3), \tag{2.5.13}$$

[43] Ignoring the contributions due to $\hat{H}_{\rm int}$, the 1-point function $\langle 0|\hat{\mathcal{R}}|0\rangle$ vanishes.

[44] The odd-point correlation functions vanish in this case. Including interaction terms results in non-Gaussian statistics, in which case the odd-point correlation functions do not generally vanish.

[45] This definition comes from the real space power spectrum of $\hat{\mathcal{R}}$, which is given by $\int d\ln k\ \Delta^2_{\mathcal{R}}(k)$.

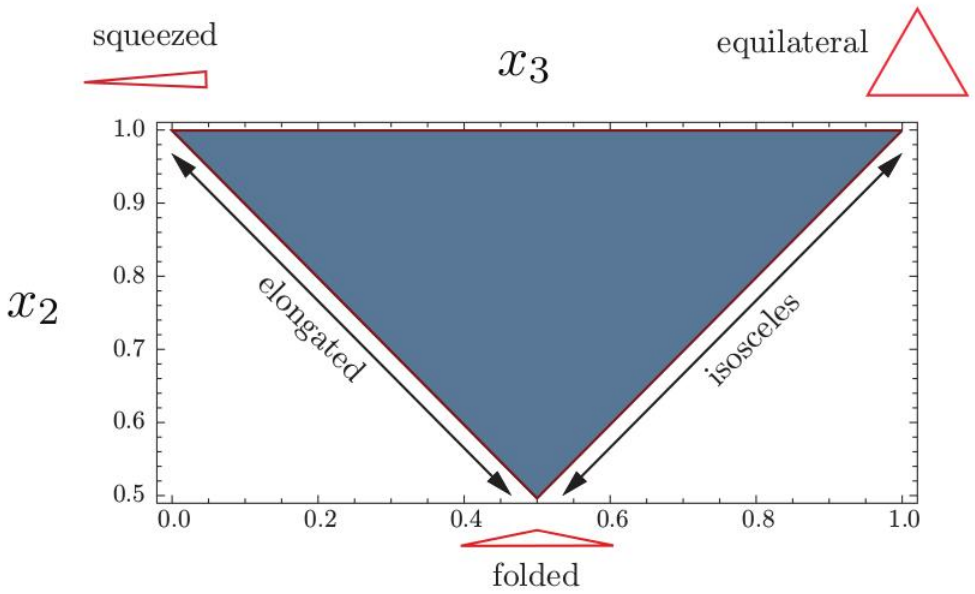


Figure 7: Different points on this triangle correspond to different momentum configurations for the bispectrum. This picture is taken from [11]. Reproduced by permission of the author.

where $\mathcal{B}_\mathcal{R}$ is called the bispectrum of the comoving curvature perturbation, or simply the scalar bispectrum. Translational and rotational invariance of the background renders the bispectrum a function of only three variables, which are conventionally taken to be the largest mode, say k_1, as well as $x_2 \equiv k_2/k_1$ and $x_3 \equiv k_3/k_1$ [46] [47]. Without any loss of generality, we require $x_3 < x_2 < 1$. Also, since the three modes form a triangle, i.e. $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$, we have $x_2 + x_3 > 1$. [see Fig. 7]. The three corners of the triangle are the squeezed limit $x_3 \to 0$, $x_2 \to 1$, the equilateral limit $x_3 \to 1$, $x_2 \to 1$, and the folded limit $x_3 \to 1/2$, $x_2 \to 1/2$. Two of the edges of the triangle are the elongated limit $x_2 + x_3 \to 1$ with $0 \leq x_3 \leq 1/2$, and the isosceles limit $x_2 - x_3 \to 0$ with $1/2 \leq x_3 \leq 1$.

It is conventional to express $\mathcal{B}_\mathcal{R}$ as

$$\mathcal{B}(k, x_2, x_3) = \frac{18}{5} f_{\rm NL} \mathcal{P}_\mathcal{R}(k)^2 \mathcal{S}(k, x_2, x_3), \tag{2.5.14}$$

where $k \equiv (k_1 k_2 k_3)^{1/3}$, $f_{\rm NL}$ is the so-called amplitude of non-Gaussianities, and $\mathcal{S}$ is known as the shape function for the bispectrum with $\mathcal{S}(k, 1, 1) = 1$[11, 60]. Various models of inflation result in

[46] It is also common to use $k \equiv (k_1 k_2 k_3)^{1/3}$ instead of k_1.

[47] If the scalar perturbations are scale-invariant, then the bispectrum is a homogeneous function of degree -6. In that case the mathematical information in the bispectrum only depends on x_2 and x_3.

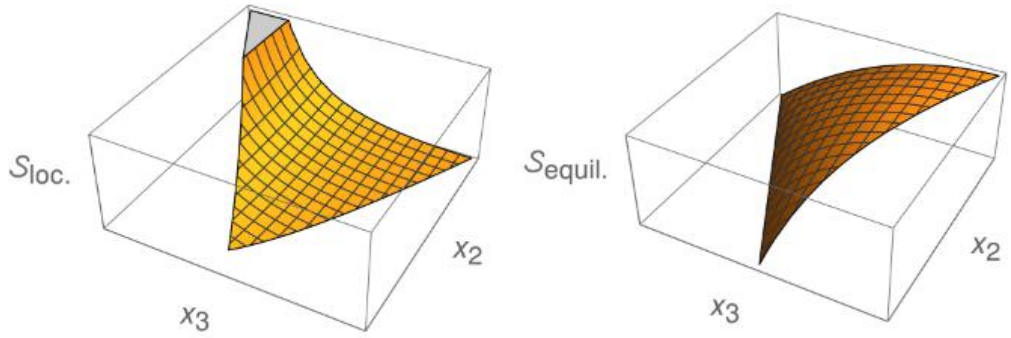


Figure 8: The plots of $\mathcal{S}_{\rm loc.}$ and $\mathcal{S}_{\rm equil.}$. The local shape peaks for $x_2, x_3 \ll 1$, while the equilateral shape peaks for $x_2 = x_3 = 1$.

different shape functions. Two of the well studied shape functions are the local shape given by

$$\mathcal{S}_{\rm loc.}(k, x_2, x_3) \equiv \frac{1}{3}\left(\frac{x_3^2}{x_2} + \frac{x_2^2}{x_3} + \frac{1}{x_3 x_2}\right), \tag{2.5.15}$$

which peaks in the squeezed limit, and the equilateral shape

$$\mathcal{S}_{\rm equil.}(k, x_2, x_3) \equiv \left(x_2 + x_3 + \frac{1}{x_2} + \frac{1}{x_3} + \frac{x_2}{x_3} + \frac{x_3}{x_2}\right) - \left(\frac{x_3^2}{x_2} + \frac{x_2^2}{x_3} + \frac{1}{x_3 x_2}\right) - 2, \tag{2.5.16}$$

which peaks in the equilateral limit [48] [see Fig. 8].

To determine the shape function and the amplitude of non-Gaussianities in a given model of inflation, one is first required to know $\hat{H}_{\rm int}$, as the bispectrum vanishes in the linear regime. In fact [49],

$$\left\langle \hat{\mathcal{R}}(\tau_0, \mathbf{k}_1)\hat{\mathcal{R}}(\tau_0, \mathbf{k}_2)\hat{\mathcal{R}}(\tau_0, \mathbf{k}_3)\right\rangle = -i\left\langle 0\left|\left[\hat{\mathcal{R}}(\tau_0, \mathbf{k}_1)\hat{\mathcal{R}}(\tau_0, \mathbf{k}_2)\hat{\mathcal{R}}(\tau_0, \mathbf{k}_3), \int_{-\infty(1\mp i\epsilon)}^{\tau_0} d\tau' \hat{H}^I_{\rm int}(\tau')\right]\right|0\right\rangle$$
$$+\mathcal{O}(\hat{H}^2_{\rm int}). \tag{2.5.17}$$

This calculation was first done for the standard single field models of inflation in [62]. It was found that $f_{\rm NL} \sim \mathcal{O}(\epsilon, \eta)$ and that its shape function is approximately a superposition of the local and the equilateral shapes. The smallness of $f_{\rm NL}$ for the standard single field models is an indication

[48] These definitions are used in [61].

[49] Beware of the sign in the lower limit of the time integral ! When computing $\hat{\mathcal{R}}^3\hat{H}_{\rm int}$ we use the top sign, and the bottom sign when computing $\hat{H}_{\rm int}\hat{\mathcal{R}}^3$.

of the small self-interactions of the inflaton field. This predicted value is well within the current observational bounds on non-Gaussianities, which we will review in Sec. 2.6.III. As we will see in chapter 3, single field models of inflation with non-vacuum initial states are capable of producing large non-Gaussianities in various momentum configurations [50].

II Correlation functions of the tensor perturbations

Much like the comoving curvature perturbation, the tensor perturbation h_{ij} is both gauge invariant and conserved on superhorizon scales. In fact, it follows from Eq. (2.3.13) that

$$\frac{2h^A(\tau,k)}{a(\tau)m_p} \to i\sqrt{2}\frac{H(\tau_k)}{k^{3/2}m_p}, \tag{2.5.18}$$

in the limit $|k\tau| \ll 1$, ignoring the subleading corrections due to the slow-roll parameters. The tensor perturbations not only give rise to primordial gravitational waves, but they also seed some of the CMB anisotropies, as we will review in Sec. 2.6.

The correlation functions of the quantized modes of h_{ij} are computed similarly to their scalar counterparts. Defining $\hat{h}^A_{ij} \equiv 2h^A e^A_{ij}/(m_p a)$, the n-point correlation function of the tensor modes in the interaction picture is given by

$$\left\langle \hat{h}^{A_1}_{i_1 j_1} ... \hat{h}^{A_n}_{i_n j_n} \right\rangle \equiv \frac{{}_{\rm BD}\left\langle 0 \middle| \hat{U}^\dagger\big(\tau_0, -\infty[1-i\epsilon]\big) \hat{h}^{A_1 I}_{i_1 j_1} ... \hat{h}^{A_n I}_{i_n j_n} \hat{U}\big(\tau_0, -\infty[1-i\epsilon]\big) \middle| 0 \right\rangle_{\rm BD}}{{}_{\rm BD}\left\langle 0 \middle| \hat{U}^\dagger\big(\tau_0, -\infty[1-i\epsilon]\big) \hat{U}\big(\tau_0, -\infty[1-i\epsilon]\big) \middle| 0 \right\rangle_{\rm BD}}, \tag{2.5.19}$$

where $|0\rangle_{\rm BD}$ is the Bunch-Davies vacuum defined for the tensor modes and $\hat{U}$ is the time evolution operator given in Eq. (2.5.5) with $\hat{H}^I_{\rm int}$ being the interaction picture Hamiltonian for the tensor perturbations. The n-point correlation functions are then usually summed over all polarizations. As we did for the scalar correlation functions, we shall drop the superscript I and the subscript BD, as well as the normalization factor in the denominator from now on.

Here we restrict attention merely to the 2-point tensor correlation function. For the 3-point tensor correlation function, as well as mixed scalar-tensor 3-point correlation functions see [62].

[50] Large non-Gaussianities can also be produced in other models of inflation, such as some multi-field models [63].

II.a Tensor power spectrum

Similar to the scalar perturbations, the leading 2-point tensor correlation function is given by

$$\sum_{A,A'} \langle 0|\hat{h}^A_{ij}(\tau_0, \mathbf{k_1})\hat{h}^{A'ij}(\tau_0, \mathbf{k_2})|0\rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2)\mathcal{P}_h(k_1), \tag{2.5.20}$$

where $\mathcal{P}_h$ is called the tensor power spectrum. It follows from the definition of $\hat{h}$ and Eq. (2.5.18) that

$$\mathcal{P}_h(k) = 4\frac{H(\tau_k)^2}{k^3 m_p^2}, \tag{2.5.21}$$

with its dimensionless counterpart being

$$\Delta^2_h(k) = \frac{k^3}{2\pi^2}\mathcal{P}_h(k) = \frac{2H(\tau_k)^2}{\pi^2 m_p^2}. \tag{2.5.22}$$

Due to its small scale dependence through $H(\tau_k)$, Δ^2_h is conventionally expressed as

$$\Delta^2_h = A_t\left(\frac{k}{k_*}\right)^{n_t}, \tag{2.5.23}$$

where k_* is some fiducial scale, A_t is the value of Δ^2_h at k_*, and n_t is the spectral tilt of Δ^2_h. To leading order in the slow-roll parameters, it is given by [11]

$$n_t = -2\epsilon. \tag{2.5.24}$$

So far there has been no detection of primordial gravitational waves. Note that such a detection will determine the energy scale of inflation, as Δ^2_h has no dependence on the slow-roll parameters.

II.b Tensor to scalar ratio and the Lyth bound

It has become customary to classify models of inflation based on their values of tensor-to-scalar ratio r defined by

$$r \equiv \frac{\Delta^2_h}{\Delta^2_{\mathcal{R}}} = 16\epsilon. \tag{2.5.25}$$

Given the fact that $\Delta^2_{\mathcal{R}}$ has been measured over a range of scales, the value of r for a given model determines the energy scale of inflation for that particular model via

$$\Lambda^{1/4}_{\rm inf} \equiv 3^{1/4}\sqrt{m_p H_{\rm inf}} = \left(\frac{3}{2}\pi^2 r \Delta^2_{\mathcal{R}}\right)^{1/4} m_p, \tag{2.5.26}$$

where we used Eq. (2.5.22). For $\Delta^2_{\mathcal{R}} \approx 2 \times 10^{-9}$ (the measured value), $\Lambda^{1/4}_{\rm inf}$ is given by

$$\Lambda^{1/4}_{\rm inf} \approx 1.3 \times 10^{-2}\ r^{1/4} m_p. \tag{2.5.27}$$

A GUT-scale inflation requires $\Lambda^{1/4}_{\rm inf} \approx 10^{-2} m_p$, which in turn requires $r \sim \mathcal{O}(0.1)$.

The tensor-to-scalar ratio can also be related to the amount of the inflaton field excursion during the observable inflation through the so-called Lyth bound [64]. To derive this bound, note that from Eqs. (2.5.10) and (2.5.22) we have

$$\begin{aligned} r = 16\epsilon &= \frac{8}{m_p^2}\left(\frac{d\bar{\phi}}{dN}\right)^2 \\ \Rightarrow \frac{\Delta\bar{\phi}}{m_p} &= \int_{N_{\rm end}}^{N_*} dN\ \sqrt{\frac{r}{8}}, \end{aligned} \tag{2.5.28}$$

where $N_{\rm end}$ denotes the number fold corresponding to the end of inflation. A lower bound for the integral given above is provided by assuming that r is only non-zero for the CMB scales which exit the horizon over an e-fold interval $\Delta N \sim 7$. Taking this, we have

$$\frac{\Delta\bar{\phi}}{m_p} \gtrsim 0.25 \times \sqrt{\frac{r}{0.01}}. \tag{2.5.29}$$

It follows from Eq. (2.5.29) that a lack of detection for the inflationary gravitational waves rules out the large field models, while a detection of $r \sim \mathcal{O}(0.01)$ implies a significant Planck sensitivity for inflation.

2.6 Observational tests of the predictions of inflation

In this section we briefly review the observational predictions of inflation. We will see that the main properties of the primordial perturbations as predicted by the standard inflationary theory, namely their near scale invariance, adiabaticity, and Gaussianity, are consistent with the observed features of the CMB and LSS. We quote the most recent measurements and bounds on the inflationary parameters, as well as a list of currently favoured single field models. We end with a brief discussion of how the upcoming missions can enhance our understanding of the physics of inflation.

I Cosmic Microwave Background radiation

The CMB was detected in 1965 by Wilson and Penzias [65, 66]. Shortly after its discovery, its frequency spectrum was shown to be that of a nearly perfect black body with a currently measured

temperature of 2.72548 ± 0.00057 °K [67]. Not long after its discovery it was realized that the CMB is most likely left over radiation from the early Universe [68]. Today, CMB observations are instrumental in probing the physics of the early Universe.

The first successful detection of CMB fluctuations on large angular scales was performed by the Cosmic Background Explorer (COBE) in 1992. The measurements of COBE supported the notion of CMB temperature fluctuations having a nearly scale invariant spectrum on the angular scales larger than 7 degrees [69]. Subsequent measurements, including the prominent nine-year study of the CMB temperature and polarization anisotropies by the Wilkinson Microwave Anisotropy Probe (WMAP) [12] and the recent comprehensive measurements of anisotropies by the Planck collaborations [2, 70], have continued to advance our understanding of the CMB physics [51].

I.a Temperature anisotropies

As mentioned above, the CMB has a nearly perfect black body spectrum, with a temperature of about 2.725°K. There are, however, small anisotropies of about one part in 10^5 that can be described by a perturbed Boltzmann distribution for the photons,

$$\mathcal{N}(p) \propto \frac{1}{e^{p/\{T_0[1+\Theta(\hat{n})]\}} - 1}, \tag{2.6.1}$$

where $\mathcal{N}(p)$ is the distribution function of photons with energy p and $\Theta(\hat{n}) \equiv \Delta T(\hat{n})/T$ is the temperature anisotropy with $\hat{n}$ being a direction in the sky at which the measurement is conducted. The properties of $\Theta(\hat{n})$ are described by its correlation functions, mainly through its power spectrum $\left\langle \Theta(\hat{n})\Theta(\hat{n}') \right\rangle$, which contains all of the information if $\Theta(\hat{n})$ is a Gaussian stochastic function.

To compute this correlation function, $\Theta(\hat{n})$ is first expanded in terms of its multipole moments,

$$\Theta(\hat{n}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm}^T Y_{lm}(\hat{n}). \tag{2.6.2}$$

Given the reality of the temperature field, $a_{lm}^* = (-1)^m a_{l-m}$. The anisotropy information is contained in the coefficients a_{lm}, whose 2-point correlation function is given by

$$\langle a_{lm}^T a_{l'm'}^{T*} \rangle = \delta_{ll'}\delta_{mm'} C_l^{TT}, \tag{2.6.3}$$

[51] A detailed treatment of the theoretical basics of the CMB physics can be found in [14, 71].

where C_l^{TT} is the TT angular power spectrum of the temperature anisotropies. This correlation function is related to its primordial counterpart via [52]

$$C_l^{TT} = \frac{2}{\pi} \int k^2 dk \; \mathcal{P}_{\mathcal{R}}(k) \big[\Delta_{Tl}(k)\big]^2, \tag{2.6.4}$$

where Δ_{Tl} is the temperature transfer function that takes into account the evolution of the once superhorizon modes upon their horizon entry. Its exact form is determined by solving the Boltzmann transport equations for the coupled photon-matter-gravitation system [53]. Our measurements of C_l^{TT} can then be used to reconstruct $\mathcal{P}_{\mathcal{R}}$, which turns out to be nearly scale invariant as predicted by inflation. Moreover, a key prediction of inflation is that all primordial perturbation modes became superhorizon while having the same phase [54]. It is currently believed that this property of the primordial perturbations is responsible for giving rise to the acoustic peaks seen in the temperature anisotropy map of the CMB. Should the perturbation modes be oscillating with different phases upon horizon entry, they would have interfered destructively at the epoch of recombination. See Fig. 9 for the most recent temperature anisotropy map by the Planck collaborations.

I.b Polarization anisotropies

In addition to temperature anisotropies, the CMB photons are expected to carry linear polarization patterns that are induced by Thomson scattering processes with free electrons during the epoch of recombination. The polarization patterns actually result from the free electrons seeing a quadrupole moment in the incident radiation field.

The polarizations of the CMB photons at any given direction on the sky are described by two Stokes parameters Q and U which transform as spin-2 fields under rotations about the given direction on the sky. Therefore, they can be expanded in terms of the spin-2 spherical harmonics,

$$(Q \pm iU) = \sum_{l,m} a_{\pm 2,lm} \; {}_{\pm 2}Y_{lm}(\hat{n}). \tag{2.6.5}$$

It is , however, much simpler to consider the spin-0 fields

$$E = \sum_{l,m} a_{lm}^E Y_{lm}(\hat{n}), \qquad B = \sum_{l,m} a_{lm}^B Y_{lm}(\hat{n}), \tag{2.6.6}$$

[52] There is also a suppressed contribution due to primordial tensor perturbations [71].
[53] See [14] for a detailed discussion.
[54] See Eq. (2.5.2).

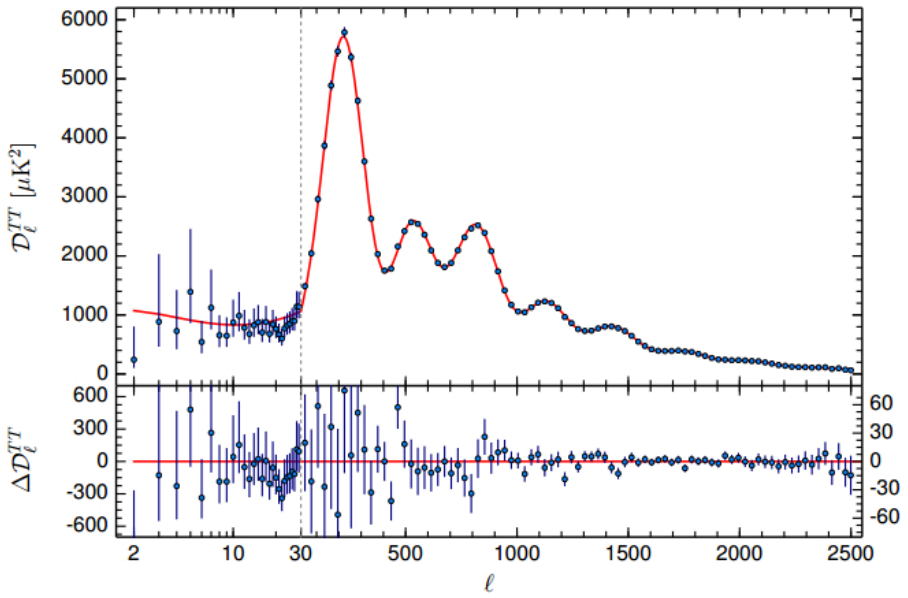


Figure 9: Planck 2015 temperature power spectrum obtained from [2]. The measurements become cluttered by cosmic variance for $l \lesssim 5$ and various other foreground effects for $l \gtrsim 2000$. The information contained in $l \sim \mathcal{O}(1)$ is particularly relevant for the physics of the early stages of a 60 e-fold GUT scale inflation. Reproduced by permission of the Planck Collaborations.

with $a^E_{lm} = -(a_{+2,lm} + a_{-2,lm})/2$ and $a^B_{lm} = i(a_{+2,lm} - a_{-2,lm})/2$. The E-mode polarization pattern is curl free and parity-even, while the B-mode polarization pattern is divergence free and parity-odd. See Fig. 10

The E-mode and B-mode patterns cannot be correlated with one another because of their parities [55]. Moreover, unlike the E-mode pattern, the B-mode pattern cannot be correlated with Θ. Therefore, we expect to have BB, EE, and TE angular power spectra that are denoted by C_l^{BB}, C_l^{EE}, and C_l^{TE} respectively[56]. These angular power spectra are defined by similar relations to the one given in Eq. (2.6.3) and are related to their primordial counterparts by similar integrals to the one provided in Eq. (2.6.4). What is different for each angular power spectrum is not just its respective transfer function, but also its corresponding primordial power spectra. In fact, it can be shown that primordial scalar perturbations do not give rise to the B-mode polarization patterns [57]. Hence, their primordial origin can only be attributed to primordial gravitational waves.

[55] Provided that we assume a parity-invariant theory.

[56] See Fig. 11.

[57] Scalar perturbation cannot seed B-modes because of their even parity. Tensors perturbations can induce this pattern as they are not invariant under rotations on the plane transverse to their momenta.

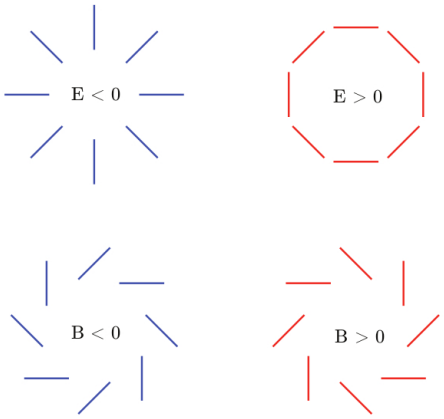


Figure 10: The E-mode and B-mode CMB polarization patterns taken from [11]. Reproduced by permission of the author.

A detection of the B-mode polarization pattern not only determines the energy scale of inflation and the inflaton field excursion [58], but it also serves as the "smoking gun" evidence of inflation's occurrence. Indeed, inflation appears to be the only compelling early Universe theory that produces non-vanishing tensor modes that are superhorizon at the epoch of recombination.

II Large Scale Structure

LSS observations indirectly probe the distribution of the cold dark matter (CDM) through the distribution of galaxies and their clusters [59]. Unlike CMB observations that become cluttered at multipoles $l \gtrsim 2000$ or $k \lesssim 0.1\text{Mpc}^{-1}$ due to Silk damping [60], LSS observations can in principle be pushed down to smaller scales [61]. Furthermore, LSS observations contain more information than CMB observations for a given range of scales since they provide three dimensional maps of the sky. Much information regarding the dynamics of inflation is expected to be learned from the current and upcoming LSS observations that will complement information from CMB observations.

The basic theory needed to explain LSS observations can be described in two stages. The first stage requires tracking the evolution of CDM perturbations that are seeded by the same primordial perturbations that gave rise to CMB temperature anisotropies. This is done by relating the correlation functions of the CDM density parameter $\delta_{\text{cdm}} \equiv \delta\rho_{\text{cdm}}/\bar{\rho}_{\text{cdm}}$ to those of the primordial

[58] See the Lyth bound (2.5.29).
[59] See [73] for a comprehensive review of its theoretical foundation.
[60] See [71] for an explanation.
[61] This may not be useful in practice as modes below about 10 Mpc become nonlinear.

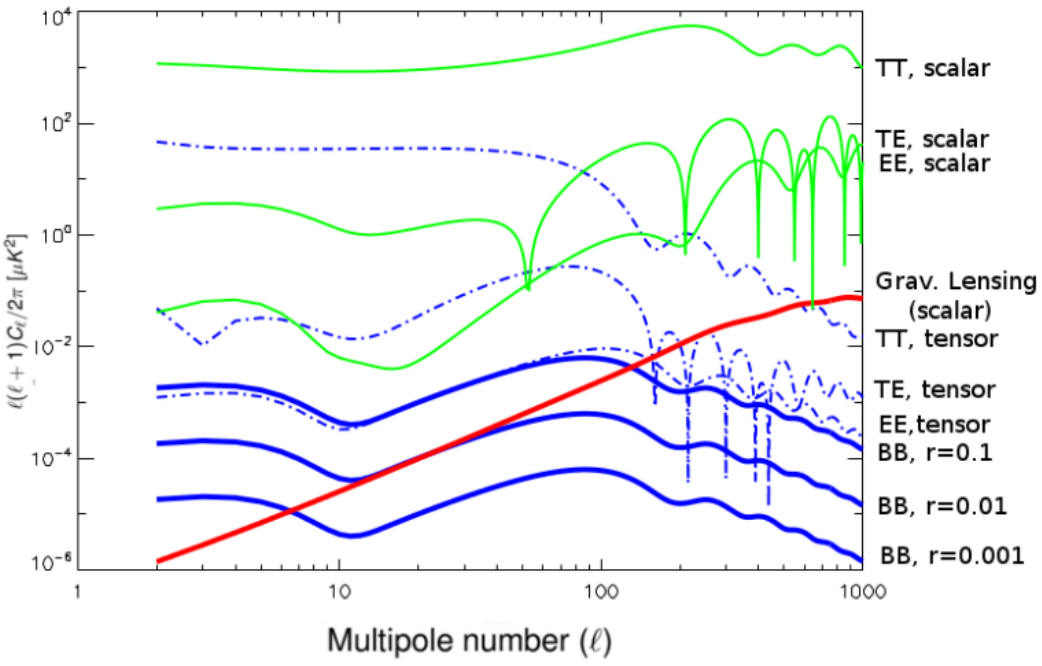


Figure 11: This is a plot of various CMB angular power spectra. The contributions due to the scalar (green curves) and the tensor (blue curves) perturbations to each power spectrum are separately shown. The dashed blue curves assume a tensor-to-scalar ratio $r = 0.1$, while the solid blue curves correspond to a few different values of r. The solid red curve is the BB power spectrum due to the non-linear scalar perturbations that arise as a result of gravitational lensing. Note that the acoustic peaks seen in the top three plots are consistent with a key prediction of inflation that the primordial perturbations are adiabatic. The plot is taken from [72]. Reproduced by permission of the author.

scalar perturbations. The power spectrum of the CDM for linear perturbation modes is related to that of the comoving curvature perturbations by

$$\mathcal{P}_{\delta_{\rm cdm}}(z,k) = T_{\delta_{\rm cdm}}(z,k)^2 \mathcal{P}_{\mathcal{R}}(k), \tag{2.6.7}$$

where z is redshift and $T_{\delta_{\rm cdm}}$ is the CDM transfer function for linear modes. After this, the second stage is relating $\mathcal{P}_{\delta_{\rm cdm}}$ to the power spectrum of its baryonic tracers, such as the galaxy power spectrum $\mathcal{P}_g$. This is done through the galaxy bias parameter. Determining the galaxy bias parameter on small scales is challenging due to our lack of complete understanding of the theory of galaxy formation. See Fig. 12 for a recent reconstruction of the CDM density contrast power spectrum using a variety of LSS surveys and CMB data.

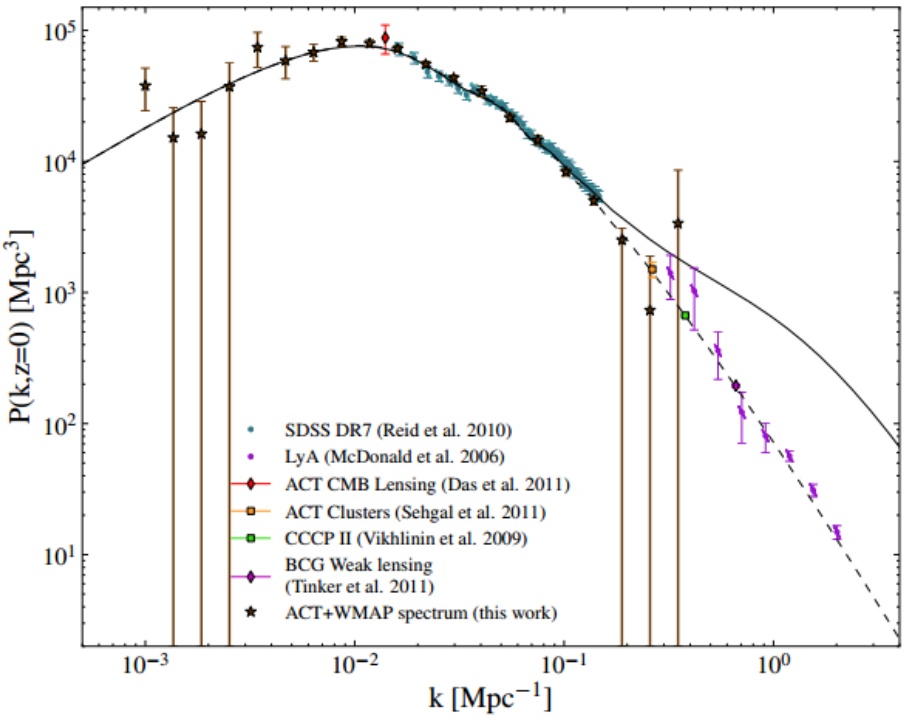


Figure 12: The $\delta_{\rm cdm}$ power spectrum reconstructed using a variety of LSS surveys and CMB data, taken from [74]. Reproduced by permission of the ACT Collaborations.

III Current constraints on the inflationary parameters

The inflationary parameters of primary interest are the amplitude A_s, the spectral tilt n_s and the running α_s of the dimensionless scalar power spectrum, the amplitude $f_{\rm NL}$ of the scalar non-Gaussianities, as well as the tensor-to-scalar ratio r. Of these five parameters, measurements have been made for A_s and n_s, while the rest are merely constrained. We quote some of the most recent results from CMB and LSS observations in Table 1. As can be seen from the table, the proximity of n_s to 1 is an indication of the near-scale invariance of primordial scalar perturbations, as correctly predicted by the standard theory of inflation. Also, the absence of any detection of non-Gaussianities shows that primordial scalar perturbations are nearly Gaussian. The prediction of the standard inflationary theory that $f_{\rm NL} \sim \mathcal{O}(\epsilon, \eta)$ is consistent with the quoted bounds, though the bounds are not yet precise enough to confirm the predicted value of $f_{\rm NL}$.

Finally, note that the results of Table 1 can be used together with Eqs. (2.5.12) and (2.5.25) to

		Planck	Planck+BAO	Planck+BICEP2+Keck
1	$\ln(10^{10}A_s)$	3.094 ± 0.034	3.093 ± 0.034	
2	n_s	0.9644 ± 0.0049	0.9673 ± 0.0043	
3	α_s	-0.0085 ± 0.0076	-0.0125 ± 0.0091	
4	r	< 0.149	< 0.166	< 0.12
5	$f_{\rm NL}^{\rm loc.}$	0.8 ± 5.0		
6	$f_{\rm NL}^{\rm equil.}$	-4 ± 43		

Table 1: Some of the recent constraints on the inflationary parameters of primary interest [29, 61, 75]. The quoted bound on r from Planck+BICEP2+Keck has 2σ errors, with all the rest having 1σ errors. The pivot scale is $k_* = 0.05$ Mpc^{-1}.

constrain the Hubble slow-roll parameters during inflation. The most stringent bound on ϵ is

$$\epsilon = \frac{r}{16} < \frac{0.12}{16} = 0.0075, \tag{2.6.8}$$

which results in

$$\eta < 0.0177, \tag{2.6.9}$$

for the pivot scale $k_* = 0.05$ Mpc^{-1} and ignoring the error bars.

III.a Currently favoured models of inflation

Single field models of inflation are typically classified based on the their predicted r and n_s values. The current bounds on r and the measurements of n_s rule out some of these models while favouring others. In Table 2 we list the values of r and n_s for the potentials discussed in Sec. 2.2.I.

Of the potentials listed in Table 2, all contain regimes that are compatible with the values listed in Table 1. The monomial potentials with $n \gtrsim 2$ and the hilltop potentials with $n \lesssim 4$ appear to be ruled out. A careful analysis of various inflationary potentials was performed in [29], with a summary presented in Fig. 13. As indicated in the figure, the Starobinsky potential is particularly favoured.

Potential	r	$r\|_{N_*=60}$	n_s	$n_s\|_{N_*=60}$
Monomial	$\frac{4n}{N_*}$	$\frac{n}{15}$	$1-\frac{2+n}{2N_*}$	$1-\frac{2+n}{120}$
Hilltop	≈ 0	≈ 0	$1-\frac{2}{N_*}\frac{n-1}{n-2}$	$1-\frac{n-1}{30(n-2)}$
Starobinsky	$\frac{12}{N_*^2}$	0.0033	$1-\frac{2}{N_*}$	0.9666
Axion Monodromy	$\frac{8m_p^2/f^2}{\exp(m_p^2N_*/f^2)-1}$	< 0.09	$1-\frac{m_p^2}{f^2}\frac{\exp(m_p^2N_*/f^2)+1}{\exp(m_p^2N_*/f^2)-1}$	< 0.965

Table 2: Here the r and n_s values for the potentials considered in Sec. 2.2.I are listed. For the derivations leading to these see [35, 38]. Recall that $n > 2$ for the Hilltop potentials considered in Sec. 2.2.I. For the Axion Monodromy potential, $r|_{N_*=60}$ and $n_s|_{N_*=60}$ values correspond to $m_p/f < 1/9$.

IV Upcoming observations and projections

Much information has been obtained from the CMB anisotropy measurements regarding the physics of inflation in the past few decades. We expect to learn more from the current and upcoming CMB polarization measurements. Nonetheless, LSS observations will be the primary source of advancement in observational cosmology for the next few decades [62]. In Table 3 we summarize the forecasts for a number of current and future observational missions as it relates to the physics of inflation.

		$\sigma(r)$	$\sigma(n_s)$	$\sigma(\alpha_s)$	$\sigma(f_{\rm NL}^{\rm loc.})$	$\sigma(f_{\rm NL}^{\rm equil.})$
1	SPIDER	1.3×10^{-2}				
2	AdvACT	2.1×10^{-3}				
3	SPT-3G	$*\circ\ 1.05\times10^{-2}$	$*\circ\ 3.91\times10^{-3}$			
4	CMB-S4	10^{-3}			*1.8	*21.2
5	DESI		$*1.9\times10^{-3}$	$*1.9\times10^{-3}$	*2.5	
6	EUCLID	$*1.9\times10^{-2}$	$*2\times10^{-3}$	$*10^{-3}$	4.7	16

Table 3: 1 σ errors on various inflationary parameters from a number of currently running balloon borne (SPIDER [77]) and ground based (AdvACT [78] and SPT-3G [79]) CMB observations, as well as a few upcoming LSS missions (EUCLID [80] and DESI [81]) and a proposed CMB mission (CMB-S4 [82]). The asterisks and circles denote conjunction with the latest PLANCK and BOSS [83] results respectively. Some of the $\sigma(r)$ forecasts are taken from [84].

[62] There are also efforts to use 21 cm observations [76].

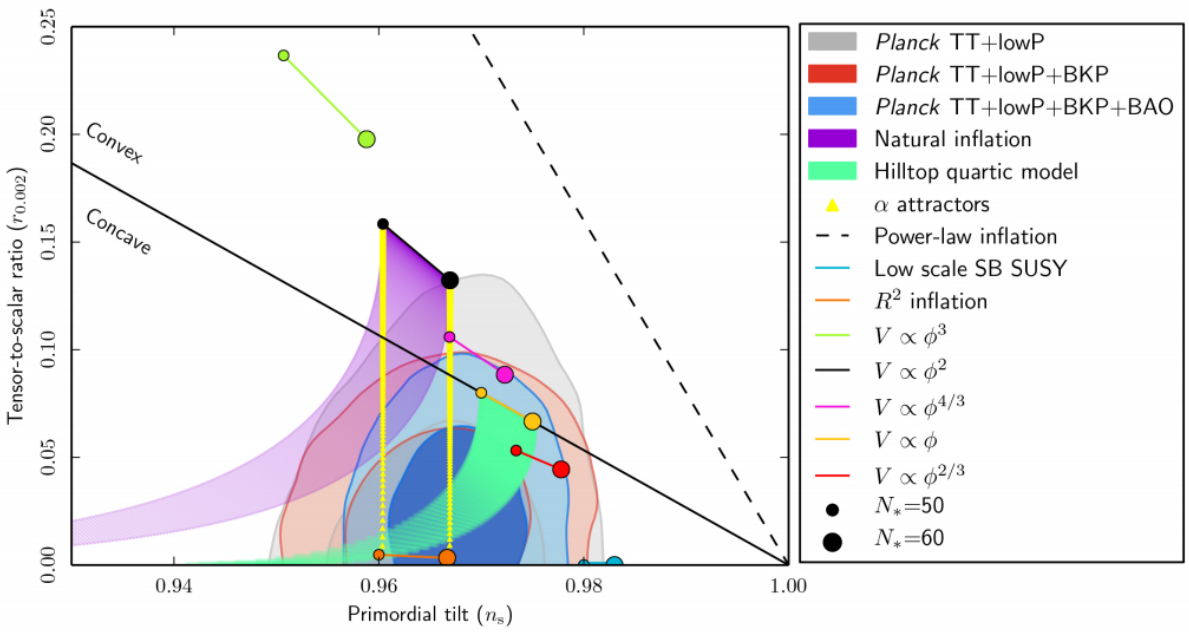


Figure 13: Here is a plot of r (at pivot scale $k_*0.002$ Mpc^{-1}) versus n_s (at pivot scale $k_*0.05$ Mpc^{-1}) compiled from various CMB and LSS observations [29]. The plot contains 1σ and 2σ contours for different sources of constraints. A number of inflationary potentials, including various monomial potentials, the hilltop quartic potential, the Starobinsky potential (R^2 inflation) and the axion monodromy potential (natural inflation). Note that the Staronisky potential is comfortably within the most stringent bounds. Reproduced by permission of the Planck Collaborations.

2.7 Reheating and pre-heating: The end of inflation

Conventionally, inflation ends when either one of the Hubble slow-roll parameters becomes equal to unity. At this time, the Universe no longer expands quasi-exponentially since the kinetic energy and the potential fluctuations of the inflaton field are large enough to completely break the background de Sitter symmetry. Once this occurs, the Universe enters a new phase and eventually makes a transition to the radiation dominated era. Microscopically, the inflaton field decays during this period to intermediary particles, which then thermalize and decay further to the Standard Model particles. This transient period is generally referred to as reheating. The exact dynamics of this period remains model-dependent and is largely unconstrained.

The simplest approach to reheating is to assume that the background inflaton field $\bar{\phi}$ decays with a rate Γ while oscillating at the bottom of its potential at the onset of reheating [63]. The dynamics of

[63]The brief summary provided here closely follows the discussion given in [41].

the inflaton during this stage is dependent on the form of its potential, with $V = m^2\phi^2/2$ resulting in a reheating phase that effectively expands as a dust dominated Universe. This phase ends when $H \sim \Gamma$, at which time the Universe is presumably radiation dominated with a temperature $T \sim \sqrt{m_p \Gamma}$.

However, the microscopic details of reheating are more complicated. It is now understood that the above simplistic picture was likely preceded by an era of pre-heating, when the inflaton decay resulted in an exponentially amplified production of intermediary particles. A text-book toy model to demonstrate this phenomenon is given by considering the decay of the inflaton to a scalar field χ through a term $g^2\phi^2\chi^2/2$ that can be added to their interaction Lagrangian. The equation of motion for the χ modes is then given by[64]

$$\chi''(z,k) + \big[A(k) - 2q\cos(2z)\big]\chi(z,k) = 0, \tag{2.7.1}$$

where prime denotes differentiation with respect to z and

$$z \equiv mt, \qquad q \equiv \frac{g^2\Phi^2}{4m^2}, \qquad A(k) \equiv \frac{k^2}{m^2} + 2q, \tag{2.7.2}$$

with Φ being the amplitude of the inflaton oscillations at the bottom of its potential. Depending on the value of q, Eq. (2.7.1) implies exponentially growing solutions for a certain range of χ modes. If $q \ll 1$, one has the so-called narrow resonance phenomenon where the number density of χ particles with $k \sim m$ exponentially increases with time. On the other hand, $q \gg 1$ results in the broad resonance phenomenon where the number density of χ particles grows exponentially for $k \lesssim \sqrt{gm\Phi}$. As it turns out, the latter reheating process is likely to be more efficient since the background expansion redshifts previously large modes to values $k \lesssim \sqrt{gm\Phi}$, resulting in the number densities of a larger range of modes to be amplified over time. The pre-heating process eventually ends when the backreaction due to $g^2\langle\chi^2\rangle$ to the effective mass of the inflaton field becomes comparable to m^2. This typically occurs within a time period much shorter than one Hubble time. After that, the simplistic reheating process described previously takes place, though predominately through the newly generated χ particles.

[64] χ field is a test scalar field with no bare mass in the Lagrangian.

Chapter 3

Single Field Models Of Inflation With Non Bunch-Davies Vacuum Initial States

3.1 Overview

When the epoch of inflation that we can probe observationally occurs shortly after the start of inflation, there are a number of different non-standard effects that can arise. One type of effect is the possibility of non Bunch-Davies vacuum initial states for the perturbations. These can result from effects at earlier times that are outside the domain of validity of a single field model.

Recall that non Bunch-Davies initial states can modify both the spectrum and bispectrum of the perturbations, and in particular can enhance the size of the bispectrum. Non-Gaussianities arising from specific classes of non trivial initial states have been studied by many authors, including particle number eigenstates [85], Gaussian states [86], general multimode squeezed (i.e. vacuum) states [87, 88, 89, 90, 91, 92, 93, 94, 95], coherent states [96], thermal states [97], and the so-called "calm states" [98]. In this chapter we shall generalize these results to arbitrary initial states that are homogeneous.

Our approach to non Bunch-Davies vacuum initial states starts with a general mixed quantum state describing a statistical distribution of initial inflaton perturbations, constrained only by the assumption of spatial homogeneity. After laying out the foundation and discussing the constraints on these initial states in Sec. 3.4, we will calculate the leading and subleading components of the power spectrum and bispectrum of the comoving curvature perturbation $\mathcal{R}$ in Secs. 3.5 and 3.6. We will then analyse the magnitude of non-Gaussianities generated in this framework in several

momentum configuration limits and discuss their potential detectability in Sec. 3.7.

3.2 Why non Bunch-Davies vacuum initial states?

In Sec. 2.4.I we showed that a natural, symmetry compatible, choice of vacuum state exists for the inflationary perturbation modes in the deep subhorizon limit. The existence of the Bunch-Davies initial states, however, relied on the assumption that the mode dynamics continues to be described by the standard single field prescription at arbitrarily early times. In reality, though, one expects the single field description to remain valid up to a certain cutoff scale, beyond which a more complicated dynamical behaviour, e.g. coupling with other degrees of freedom, becomes relevant.

To demonstrate this phenomenon using a toy model, let us consider the action

$$S[\phi,\chi]=\int\sqrt{-g}\;d^4x\;\bigg[-\frac{1}{2}\nabla_a\phi\nabla^a\phi-\frac{1}{2}m_\phi^2\phi^2-\frac{1}{2}\nabla_a\chi\nabla^a\chi-\frac{1}{2}m_\chi^2\chi^2+g\chi\phi^3\bigg], \tag{3.2.1}$$

for two scalar fields ϕ and χ propagating in a fixed background spacetime. Assuming that $m_\chi\gg m_\phi$ and that we are interested in deriving an effective description for dynamics of ϕ field at energy scales $E\ll m_\chi$, we integrate out the χ field following the standard Wilsonain procedure [1]. Doing this results in the effective ϕ action

$$S_{\text{Eff.}}^{E\ll m_\chi}[\phi]=\int\sqrt{-g}\;d^4x\;\bigg[-\frac{1}{2}\nabla_a\phi\nabla^a\phi-\frac{1}{2}m_\phi^2\phi^2-\frac{g^2}{2}\phi^3\frac{1}{\Box-m_\chi^2}\phi^3\bigg], \tag{3.2.2}$$

where the self-interaction term can be incorporated order by order in powers of $\Box/m_\chi^2$. For instance, it reduces to $g^2\phi^6/(2m_\chi^2)$ at the zeroth order. Now suppose that we are interested in computing the leading piece of the equal time 2-point correlation function of ϕ at some time τ_0 after ϕ decouples from χ. This can be done if the state of ϕ at the time of decoupling is known. In fact, all correlation functions of ϕ at τ_0 can be computed to all orders in powers of $\Box/m_\chi^2$ using the effective action (3.2.2) provided that the state of ϕ at the time of decoupling is known. However, this state is partly determined from the model dependent coupled dynamics of ϕ and χ prior to the time of decoupling. Therefore, the couplings of ϕ to χ not only dictates the initial state to be used in the calculations of the correlation functions of the effective ϕ theory, it also determines the self-interaction Hamiltonian that is needed to fully specify these correlation functions.

[1] See eg. [59].

Given the speculated proximity of the single field inflation energy scale to m_p, the need for considering non Bunch-Davies initial states is particularly important for addressing the inflationary trans-Planckian issue [56, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108]. As we briefly reviewed in Sec. 2.4.II, the modes that we observe today have physical frequencies $\omega_{\rm phys}$ which exceed the Planck scale m_p at sufficiently early times. A complete description of the physics at that stage will only be possible once we have a solid understanding of quantum gravity. Nonetheless, it has been argued that the quantum gravitational corrections to the dynamics at horizon crossing are suppressed by $\sim H^2/m_p^2$ if the evolution is always adiabatic, but could be larger and scale as $\sim H/m_p$ if violations of adiabaticity occur for any given mode. In many toy models of quantum gravity effects, the effects of the Planck scale corrections to the dynamics of the modes can be mimicked by using the standard dynamics but with a modified initial state. Many computations have been performed of modifications to the spectrum and bispectrum of the perturbations due to new physics at high energies [94, 101, 102, 103, 104, 105, 106, 107, 108]. The general results of this chapter could be used as a tool in such computations.

3.3 Action for scalar perturbations

To compute the subleading component of the power spectrum as well as the leading bispectrum piece of $\hat{\mathcal{R}}$ using Eq. (2.5.7), we need the leading order interaction Hamiltonian of $\hat{\mathcal{R}}$. That requires expanding the action (2.1.1) upto third order in the scalar perturbations,

$$S_{\rm scalar}[\mathcal{R}] = \delta_2 S_{\rm scalar}[\mathcal{R}] + \delta_3 S_{\rm scalar}[\mathcal{R}] + \mathcal{O}(\mathcal{R}^4), \tag{3.3.1}$$

where we dropped the zeroth and first order terms as they do not contribute to the dynamics of perturbations. The second order action is just the Mukhanov action (2.3.5) with ν being replaced by $\zeta\mathcal{R}$. The third order action was first computed in [62] and it is given by

$$\delta S_3[\mathcal{R}] = -\int d\tau H_{\rm int} = \int d\tau \int d^3x\, a(\tau)^3 \left(\frac{\dot{\phi}}{H}\right)^4 H m_p^{-2} \mathcal{R}_c'^2 \partial^{-2} \mathcal{R}_c', \tag{3.3.2}$$

where $\mathcal{R}_c$ is defined by

$$\begin{aligned}\mathcal{R} &= \mathcal{R}_c + \left[\frac{1}{2}\frac{\ddot{\phi}}{\dot{\phi}H} + \frac{1}{8m_p^2}\frac{\dot{\phi}^2}{H^2}\right]\mathcal{R}_c^2 + \frac{1}{4m_p^2}\frac{\dot{\phi}^2}{H^2}\partial^{-2}\big(\mathcal{R}_c\partial^2\mathcal{R}_c\big) \\ &= \mathcal{R}_c + \left[\frac{\epsilon}{4} - \frac{\eta}{2}\right]\mathcal{R}_c^2 + \frac{\epsilon}{2}\partial^{-2}\big(\mathcal{R}_c\partial^2\mathcal{R}_c\big),\end{aligned} \tag{3.3.3}$$

where we used $\eta = -\ddot{\bar{\phi}}/(H\dot{\bar{\phi}})$ instead of the definition given in Eq. (2.2.5). Here as before primes denote differentiation with respect to τ. The modes of $\mathcal{R}_c$ are quantized similarly to those of $\mathcal{R}$ by

$$\hat{\mathcal{R}}_c(\tau, \mathbf{k}) = \hat{a}_{\mathbf{k}} \mathcal{R}_c(\tau, k) + \hat{a}^\dagger_{-\mathbf{k}} \mathcal{R}_c^*(\tau, k), \tag{3.3.4}$$

where $\hat{a}$ annihilates the Bunch-Davies vacuum state $|0\rangle$ defined with respect to the modes of $\mathcal{R}$. This is due to the fact that in the interaction picture, which will be used in this chapter, the operators are evolved according to the free theory and $\mathcal{R}$ and $\mathcal{R}_c$ agree at linear order. It follows from Eqs. (2.3.10) and (2.5.1) that to leading order in the slow-roll parameters

$$\mathcal{R}_c = \frac{(i - k\tau)H}{2\epsilon^{1/2} k^{3/2} m_p} e^{-ik\tau}, \qquad \mathcal{R}_c' = \frac{iH\tau k^{1/2}}{2\epsilon^{1/2} m_p} e^{-ik\tau}. \tag{3.3.5}$$

3.4 Parametrization of initial states

We assume initial states that are homogeneous but otherwise arbitrary. We will also discuss the constraints that must be placed on these states in order to avoid large backreaction effects on the background geometry.

I Parametrization of general homogeneous states

For an initial density matrix $\hat{\rho}$, the spectrum and bispectrum of $\hat{\mathcal{R}}$ at late times will be determined by the initial two point, three point, and four point functions at some initial conformal time τ_i, $\langle \hat{a}^\alpha_{\mathbf{k}_1} \hat{a}^\beta_{\mathbf{k}_2} \rangle$, $\langle \hat{a}^\alpha_{\mathbf{k}_1} \hat{a}^\beta_{\mathbf{k}_2} \hat{a}^\gamma_{\mathbf{k}_3} \rangle$, and $\langle \hat{a}^\alpha_{\mathbf{k}_1} \hat{a}^\beta_{\mathbf{k}_2} \hat{a}^\gamma_{\mathbf{k}_3} \hat{a}^\delta_{\mathbf{k}_4} \rangle$. Here $\langle ... \rangle$ means the quantum expectation value $\mathrm{tr}[\hat{\rho}...]$ where $\hat{\rho}$ is an initial density matrix, and $\hat{a}^\alpha_{\mathbf{k}}$ means either the quantum annihilation operator $\hat{a}_{\mathbf{k}}$ (for $\alpha = 0$) or the quantum creation operator $\hat{a}^\dagger_{\mathbf{k}}$ (for $\alpha = 1$) for the quantized modes of $\hat{\mathcal{R}}$ [2].

As mentioned above, we assume that the initial state is homogeneous. It follows that the initial two point function can be parametrized in terms of two functions $\mathcal{F}_1(\mathbf{k})$ and $\mathcal{F}_2(\mathbf{k})$ by

$$\langle \hat{a}_{\mathbf{k}_1} \hat{a}_{\mathbf{k}_2} \rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2) \mathcal{F}_1(\mathbf{k}_1), \tag{3.4.1a}$$

$$\langle \hat{a}^\dagger_{-\mathbf{k}_1} \hat{a}_{\mathbf{k}_2} \rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2) \mathcal{F}_2(\mathbf{k}_1). \tag{3.4.1b}$$

It is clear that $\mathcal{F}_2$ is real and $\mathcal{F}_1$ is even. Under translation $\mathbf{x} \to \mathbf{x} + \mathbf{a}$ we have $\hat{a}_{\mathbf{k}} \to \hat{a}_{\mathbf{k}} e^{-i\mathbf{k}.\mathbf{a}}$,

[2] See Secs. 2.4.I and 2.5.I

which dictates the appearance of the specific delta functions in Eqs. (3.4.1). For an isotropic initial state $\mathcal{F}_1$ and $\mathcal{F}_2$ are just functions of k.

Similarly, the three point function under the assumption of homogeneity can be parametrized in terms of two functions $\mathcal{H}_1(\mathbf{k}_1, \mathbf{k}_2)$ and $\mathcal{H}_2(\mathbf{k}_1, \mathbf{k}_2)$, defined by

$$\left\langle \hat{a}_{\mathbf{k}_1} \hat{a}_{\mathbf{k}_2} \hat{a}_{\mathbf{k}_3} \right\rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \mathcal{H}_1(\mathbf{k}_1, \mathbf{k}_2), \tag{3.4.2a}$$

$$\left\langle \hat{a}^\dagger_{-\mathbf{k}_3} \hat{a}_{\mathbf{k}_1} \hat{a}_{\mathbf{k}_2} \right\rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \mathcal{H}_2(\mathbf{k}_1, \mathbf{k}_2). \tag{3.4.2b}$$

Both $\mathcal{H}_1$ and $\mathcal{H}_2$ are symmetric under interchange of their arguments. Also the other expectation values $\left\langle \hat{a}^\alpha_{\mathbf{k}_1} \hat{a}^\beta_{\mathbf{k}_2} \hat{a}^\gamma_{\mathbf{k}_3} \right\rangle$ can be expressed in terms of $\mathcal{H}_1$ and $\mathcal{H}_2$. In the isotropic case, $\mathcal{H}_1$ and $\mathcal{H}_2$ depend only on k_1, k_2 and on the angle between $\mathbf{k}_1$ and $\mathbf{k}_2$. Equivalently, they are functions of k_1, k_2 and k_3, or of k, k_1/k_3 and k_2/k_3, where $k \equiv (k_1 k_2 k_3)^{1/3}$.

Finally, the initial 4-point function can be parameterized by three functions $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{G}_3$ of 3 vectors:

$$\left\langle \hat{a}_{\mathbf{k}_1} \hat{a}_{\mathbf{k}_2} \hat{a}_{\mathbf{k}_3} \hat{a}_{\mathbf{k}_4} \right\rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) \mathcal{G}_1(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3), \tag{3.4.3a}$$

$$\left\langle \hat{a}^\dagger_{-\mathbf{k}_1} \hat{a}_{\mathbf{k}_2} \hat{a}_{\mathbf{k}_3} \hat{a}_{\mathbf{k}_4} \right\rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) \mathcal{G}_2(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3), \tag{3.4.3b}$$

$$\left\langle \hat{a}^\dagger_{-\mathbf{k}_1} \hat{a}^\dagger_{-\mathbf{k}_2} \hat{a}_{\mathbf{k}_3} \hat{a}_{\mathbf{k}_4} \right\rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) \mathcal{G}_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3). \tag{3.4.3c}$$

These functions inherit some symmetry properties from their definitions. The function $\mathcal{G}_1$ is symmetric in all three of its arguments, and in addition obeys the identity

$$\mathcal{G}_1(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \mathcal{G}_1(\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_1 - \mathbf{k}_2 - \mathbf{k}_3), \tag{3.4.4}$$

which follows from the invariance of the left hand side of Eq.(3.4.3a) under $\mathbf{k}_3 \leftrightarrow \mathbf{k}_4$. Similarly the function $\mathcal{G}_2$ is symmetric in it second and third argument, and obeys the identities

$$\mathcal{G}_2(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \mathcal{G}_2(\mathbf{k}_1, -\mathbf{k}_1 - \mathbf{k}_2 - \mathbf{k}_3, \mathbf{k}_3), \tag{3.4.5a}$$

$$\mathcal{G}_2(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \mathcal{G}_2(\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_1 - \mathbf{k}_2 - \mathbf{k}_3). \tag{3.4.5b}$$

The function $\mathcal{G}_3$ is symmetric in its first two arguments, and obeys the identity

$$\mathcal{G}_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \mathcal{G}_3(\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_1 - \mathbf{k}_2 - \mathbf{k}_3). \tag{3.4.6}$$

In addition taking the complex conjugate of Eq.(3.4.3c) gives the identity

$$\mathcal{G}_3^*(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \mathcal{G}_3(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3, -\mathbf{k}_3, -\mathbf{k}_2). \tag{3.4.7}$$

We also decompose the four point function into its disconnected or Gaussian piece, and its connected piece:

$$\begin{aligned}\mathcal{G}_1(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) &= \delta(\mathbf{k}_1+\mathbf{k}_2)\mathcal{F}_1(\mathbf{k}_1)\mathcal{F}_1(\mathbf{k}_3)+\delta(\mathbf{k}_1+\mathbf{k}_3)\mathcal{F}_1(\mathbf{k}_1)\mathcal{F}_1(\mathbf{k}_2)+\delta(\mathbf{k}_1+\mathbf{k}_4)\mathcal{F}_1(\mathbf{k}_1)\mathcal{F}_1(\mathbf{k}_2)\\ &\quad +\Gamma_1(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3). \qquad (3.4.8a)\end{aligned}$$

$$\begin{aligned}\mathcal{G}_2(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) &= \delta(\mathbf{k}_1+\mathbf{k}_2)\mathcal{F}_2(\mathbf{k}_1)\mathcal{F}_1(\mathbf{k}_3)+\delta(\mathbf{k}_1+\mathbf{k}_3)\mathcal{F}_2(\mathbf{k}_1)\mathcal{F}_1(\mathbf{k}_2)+\delta(\mathbf{k}_1+\mathbf{k}_4)\mathcal{F}_2(\mathbf{k}_1)\mathcal{F}_1(\mathbf{k}_2)\\ &\quad +\Gamma_2(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3). \qquad (3.4.8b)\end{aligned}$$

$$\begin{aligned}\mathcal{G}_3(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) &= \delta(\mathbf{k}_1+\mathbf{k}_2)\mathcal{F}_1^*(\mathbf{k}_1)\mathcal{F}_1(\mathbf{k}_3)+\delta(\mathbf{k}_1+\mathbf{k}_3)\mathcal{F}_2(\mathbf{k}_1)\mathcal{F}_2(\mathbf{k}_2)+\delta(\mathbf{k}_1+\mathbf{k}_4)\mathcal{F}_2(\mathbf{k}_1)\mathcal{F}_2(\mathbf{k}_2)\\ &\quad +\Gamma_3(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3), \qquad (3.4.8c)\end{aligned}$$

Here the functions Γ_1, Γ_2 and Γ_2 parameterize the connected piece.

Note that should we choose an initial density matrix which is diagonal on the eigenbasis of mode occupation number $\hat{N}_{\mathbf{k}} = \hat{a}^\dagger_{\mathbf{k}}\hat{a}_{\mathbf{k}}$ associated with the Bunch-Davies vacuum, the following restrictions on the functions $\mathcal{F}_i$, $\mathcal{H}_i$, and $\mathcal{G}_i$ would result:

$$\mathcal{F}_1 = 0, \qquad (3.4.9a)$$

$$\mathcal{H}_1 = \mathcal{H}_2 = 0, \qquad (3.4.9b)$$

$$\mathcal{G}_1 = \mathcal{G}_2 = 0. \qquad (3.4.9c)$$

This particular choice of initial density matrix was employed in [85]. Our results given in the subsequent sections for the spectrum and bispectrum differ from the ones given in [85] by terms involving the functions above which vanish for their initial state.

Recall from our discussion in Sec. 2.5.I that there are two different versions of the initial vacuum state. There is the Bunch-Davies state $|0\rangle$ which satisfies $\hat{a}_{\mathbf{k}}|0\rangle = 0$, that is the ground state of the free theory. There is also the dressed vacuum $|\Omega\rangle$, defined as the ground state of the Hamiltonian of the full theory including non-linear interaction terms which we provided in Sec. 3.3. The standard result for the bispectrum, first calculated in [62], is the result for the dressed vacuum. The functions $\mathcal{F}_i$, $\mathcal{H}_2$, $\mathcal{G}_i$ and Γ_i all vanish in both vacua, $|0\rangle$ and $|\Omega\rangle$, to leading order in the interactions. [3]. However, the function $\mathcal{H}_1$ vanishes only in the unperturbed vacuum $|0\rangle$, and not in the dressed

[3] See App. C

vacuum $|\Omega\rangle$. We therefore define a modified version of this function by subtracting the dressed vacuum contribution

$$\hat{\mathcal{H}}_1 \equiv \mathcal{H}_1 - \mathcal{H}_1^{0,\mathrm{dr.}}. \tag{3.4.10}$$

Here $\mathcal{H}_1^{0,\mathrm{dr.}}$ means the value of $\mathcal{H}_1$ in the dressed vacuum $|\Omega\rangle$, which we compute in App. C. We will express our result below for the bispectrum in terms of $\mathcal{F}_i$, $\hat{\mathcal{H}}_1$,$\mathcal{H}_2$, and Γ_i, so that when these functions vanish our result reduces to the one given in [62].

II Example: multimode squeezed initial states

An example of a class of initial states to which our analysis applies are multimode squeezed states or generalized vacuum states. These are defined by considering mode operators $\hat{b}_{\mathbf{k}}$ which are related to the operators $\hat{a}_{\mathbf{k}}$ of the $\hat{\mathcal{R}}$ mode expansion given in Eq. (2.5.5) by a linear transformation of the form

$$\begin{pmatrix} \hat{b}_{\mathbf{k}} \\ \hat{b}^\dagger_{-\mathbf{k}} \end{pmatrix} = \begin{pmatrix} \alpha^*(k) & -\beta^*(k) \\ -\beta(k) & \alpha(k) \end{pmatrix} \begin{pmatrix} \hat{a}_{\mathbf{k}} \\ \hat{a}^\dagger_{-\mathbf{k}} \end{pmatrix}, \tag{3.4.11}$$

where $|\alpha(k)|^2 - |\beta(k)|^2 = 1$. The squeezed state at the initial time τ_i is then the state that is annihilated by the operators $\hat{b}_{\mathbf{k}}$. This is the class of states that would be produced by starting at an earlier time with a Bunch-Davies vacuum, and then evolving through a homogeneous isotropic expansion phase with a single scalar field. Non-Gaussianities produced by this class of states have been studied in [87, 88, 89, 90, 91, 92, 93, 94, 95].

For this class of states, we can compute the free functions $\mathcal{F}_i$, $\mathcal{H}_i$ and Γ_i that parametrize the initial state by inserting the definition (3.4.11) into Eqs. (3.4.1) – (3.4.8). This gives

$$\mathcal{H}_i = \Gamma_i = 0, \tag{3.4.12}$$

and

$$\mathcal{F}_2(\mathbf{k}) = |\beta(k)|^2, \quad \mathcal{F}_1(\mathbf{k}) = \alpha(k)\beta^*(k). \tag{3.4.13}$$

III Constraints from backreaction considerations

The energy density and pressure due to the non-vacuum initial state must be small enough that it does not significantly perturb the inflationary background solution. We now review the order of magnitude estimate of this constraint [86, 89, 93]. As long as we are considering modes which are

well inside the horizon, the stress energy tensor associated with the non-vacuum state consists of an energy density ρ_r and a pressure $p_r = \rho_r/3$. These must be small compared to the background energy density $\sim m_p^2 H^2$ during inflation, yielding

$$\rho_r \ll m_p^2 H^2. \tag{3.4.14}$$

A more precise restriction on the initial state can be obtained from the requirement that the Hubble slow-roll parameters of the background are not perturbed to be larger than indicated by observations. From Eqs. (2.2.1) and (2.2.5) we have

$$\dot{H} = -\epsilon H^2 = \frac{-1}{2m_p^2}(p+\rho), \qquad \ddot{H} = 2\epsilon\eta H^3 = \frac{-1}{2m_p^2}[\dot{p} - 3H(p+\rho)].^4 \tag{3.4.15}$$

The pressure p and energy density ρ here consist of the usual contributions from the background inflaton field, together with the contributions ρ_r and p_r from the non-vacuum state of the perturbations. The changes in the slow-roll parameters due to the radiation are then

$$\Delta\epsilon \sim \rho_r/(m_p^2 H^2), \qquad \Delta\eta \sim \rho_r/(\epsilon m_p^2 H^2). \tag{3.4.16}$$

Imposing that $\Delta\epsilon \lesssim \epsilon$ and $\Delta\eta \lesssim \eta$ yields the constraints [89]

$$\rho_r \lesssim \epsilon m_p^2 H^2, \tag{3.4.17a}$$

$$\rho_r \lesssim \epsilon\eta m_p^2 H^2. \tag{3.4.17b}$$

The constraints (3.4.17) are more precise than Eq. (3.4.14), and strongly restrict the number of quanta present in the initial state. Note that the upper bound (3.4.17a) is just the kinetic energy $\dot{\bar{\phi}}^2 \sim \epsilon m_p^2 H^2$ of the background inflaton field.

We impose the constraints (3.4.17) at the initial time τ_i when we start evolving the perturbation modes. If they are satisfied then, they will be satisfied at all subsequent times since the radiation energy density will fall off rapidly while the background inflaton energy density is approximately constant.

We can approximately characterize initial states in terms of the mean mode occupation number $N_{\rm occ}(k)$, related to the functions $\mathcal{F}_1$ and $\mathcal{F}_2$ by [5]

$$N_{\rm occ} \sim \mathcal{F}_1 \sim \mathcal{F}_2. \tag{3.4.18}$$

[4] Here we used $\eta = -\ddot{\bar{\phi}}/(\dot{\bar{\phi}}H)$ instead of the definition given in Eq. (2.2.5).

[5] For the special case of generalized vacuum states, the estimate (3.4.18) is valid for $N_{\rm occ} \gtrsim 1$, but must be replaced by $\mathcal{F}_2 \sim N_{\rm occ}$, $\mathcal{F}_1 \sim \sqrt{N_{\rm occ}}$ when $N_{\rm occ} \ll 1$. See Eq. (3.4.13). Because we are interested in general states we will use the estimate (3.4.18) in the remainder of this chapter.

The energy density per logarithmic wavenumber is $\sim N_{\rm occ}(k)(k/a)^4$. In order to explore the consequences of the constraints (3.4.17), we specialize to a class of states for which the mean occupation is a power law in some interval $k_{\rm min} \leq k \leq k_{\rm max}$ of comoving wavenumber,

$$N_{\rm occ}(k) \sim N_0 \left(\frac{k}{k_{\rm min}}\right)^{n-4} \tag{3.4.19}$$

and vanishes outside that interval. The corresponding energy density is $\rho_r \sim N_0 E_{\rm max}^4 \chi^{-4} f_n(\chi)$, where $E_{\rm max} = k_{\rm max}/a(\tau_i)$ is the maximum physical energy of occupied modes, $\chi = k_{\rm max}/k_{\rm min}$, and

$$f_n(\chi) = \begin{cases} \ln \chi & n = 0, \\ \frac{\chi^n - 1}{n} & n \neq 0. \end{cases} \tag{3.4.20}$$

Combining these estimates with the constraint (3.4.17b) yields an upper bound on the mode occupation number

$$N_0 \lesssim \frac{\epsilon\eta\ m_p^2 H^2 \chi^4}{f_n(\chi) E_{\rm max}^4}. \tag{3.4.21}$$

We now explore the consequences of this bound under two different sets of assumptions, a robust and conservative set of assumptions in the next subsection, and a more speculative set in the following subsection.

III.a Constraints under conservative assumptions

Our conservative assumptions are:

- At the initial time, the standard inflationary action (2.1.1) that we have assumed is valid as an effective field theory up to some cutoff energy scale $\Lambda^{1/4}$.
- At the initial time, the occupied modes are all inside the horizon and below the cutoff, that is,

$$H \leq k_{\rm min}/a \leq k_{\rm max}/a \leq \Lambda^{1/4}. \tag{3.4.22}$$

This assumption is necessary for the validity of our method of computation of the backreaction due to the occupied modes.

It follows from these assumptions that $E_{\rm max} \geq \chi H$, where $\chi = k_{\rm max}/k_{\rm min}$. Combining this with the upper bound (3.4.21) and eliminating ϵ in favour of the measured amplitude of the power spectrum $\Delta_{\mathcal{R}}$ using Eq. (2.5.10) gives

$$N_0 \lesssim \frac{1}{f_n(\chi)} \frac{\eta}{8\pi^2 \Delta_{\mathcal{R}}^2}. \tag{3.4.23}$$

Now the observed value of the scalar spectral index, $n_s = 1 - 2\eta - 4\epsilon \sim 0.96$[6] [2], suggests that $\epsilon \sim \eta \sim 0.01$, absent any fine tuning. Also the observed value of $\Delta_{\mathcal{R}}$ is $\Delta_{\mathcal{R}} \sim 3 \times 10^{-5}$ [11]. Inserting these estimates finally gives the upper bound

$$N_0 \lesssim \frac{10^5}{f_n(\chi)}. \tag{3.4.24}$$

The final upper bound (3.4.24) on the mode occupation number depends strongly on the width $\chi = k_{\rm max}/k_{\rm min}$ of the band of occupied modes, as well as the power law index n. If only a small set of modes is occupied, $\chi \sim 1$, then the mode occupation number can be large compared to unity. However, if the entire range of modes visible in the CMB are occupied, so that $\chi \gtrsim 10^3$, then the constraint depends on the power law index. For example, for $n = 0$, a scale invariant spectrum, N_0 can be large. For $n = 4$ which corresponds to $N_{\rm occ}(k) \sim$ (constant), we obtain the constraint that $N_0 \lesssim 10^{-7}$, which is quite restrictive.

Note that N_0 is the value of $N_{\rm occ}(k)$ at the minimum value of k, $k = k_{\rm min}$, and is also the largest value of $N_{\rm occ}(k)$ as long as $n \leq 4$. Therefore the constraint (3.4.24) is an absolute upper bound for $N_{\rm occ}(k)$ for $n \leq 4$. It will also be useful later to have constraints on the mode occupation number evaluated at the largest value of k,

$$N_1 \equiv N_{\rm occ}(k_{\rm max}) = N_0 \chi^{n-4}, \tag{3.4.25}$$

even though $N_{\rm occ}(k)$ can exceed N_1. From Eq. (3.4.24) we find

$$N_1 \lesssim \frac{10^5}{\chi^{4-n} f_n(\chi)} = \frac{10^5}{\chi^4} \left[\frac{n}{1 - \chi^{-n}} \right]. \tag{3.4.26}$$

The factor in square brackets is either of order unity or small, for $|n| \lesssim$ (a few), so we obtain $N_1 \lesssim 10^5/\chi^4$. In particular $N_1 \ll 1$ if the entire range of modes visible in the CMB are occupied, $\chi \gtrsim 10^3$.

Another interesting quantity to constrain is the maximum bandwidth $\chi_{\rm occ}$ over which the occupation number exceeds unity. In other words, if $N_{\rm occ} \geq 1$ for some range $k_1 \leq k \leq k_2$ of values of k, what is the largest possible value of $\chi_{\rm occ} = k_2/k_1$ that is compatible with the backreaction constraint? We now argue that this maximum bandwidth is ~ 25.

Consider first the case $n < 4$, in which the maximum value of $N_{\rm occ}(k)$ is N_0 and is achieved at $k = k_{\rm min}$. Then if $N_0 < 1$, the occupation number never exceeds unity, and we set $\chi_{\rm occ} = 1$ in this

[6] Recall that we are using a different η than the one given in Eq. (2.2.5).

case (zero bandwidth). If $N_0 > 1$ the bandwidth is given by $1 = N_0 \chi_{\rm occ}^{n-4}$, and combining these gives $\chi_{\rm occ} = \max\left[1, N_0^{1/(4-n)}\right]$. However clearly $\chi_{\rm occ}$ cannot exceed $\chi = k_{\rm max}/k_{\rm min}$, the bandwidth in which we have assumed $N_{\rm occ}$ is nonzero, so we obtain

$$\chi_{\rm occ} = \min\left\{\chi, \max\left[1, N_0^{1/(4-n)}\right]\right\}. \tag{3.4.27}$$

A similar analysis for $n > 4$ yields

$$\chi_{\rm occ} = \min\left\{\chi, \max\left[1, \chi N_0^{1/(n-4)}\right]\right\}. \tag{3.4.28}$$

Combining these results and using the upper bound (3.4.23) for N_0 gives

$$\chi_{\rm occ} \leq \min\left\{\chi, \max\left[1, \chi^{\Theta(n-4)}\left(\frac{Gn}{\chi^n - 1}\right)^{1/|n-4|}\right]\right\}, \tag{3.4.29}$$

where Θ is the step function and $G = \eta/(8\pi^2\Delta_{\mathcal{R}}^2) \sim 10^5$. We now maximize over all values of n and over values of $\chi \geq 1$. The function of n and χ on the right hand side of Eq. (3.4.29) is maximized at $n = 4$ and at $\chi = (1+4G)^{1/4}$, and the maximum value is $(1+4G)^{1/4}$. Thus we obtain

$$\chi_{\rm occ} \lesssim (1+4G)^{1/4} \sim 25. \tag{3.4.30}$$

Finally we note that the upper limit (3.4.30) on the bandwidth of occupied modes is weakened if one uses the less stringent backreaction constraint (3.4.17a) instead of the more stringent constraint (3.4.17b). This yields

$$\chi_{\rm occ} \lesssim (1+4G/\eta)^{1/4} \sim 100, \tag{3.4.31}$$

which was the upper limit obtained in [109].

III.b Constraints under more speculative assumptions

Next, an upper bound on N_0 which is somewhat stronger than the bound (3.4.24) can be obtained if one assumes that the backreaction is small not just at the chosen initial time, but also at all previous times, continuing into the past until the energies of the occupied modes are blueshifted up to the cutoff scale $\Lambda^{1/4}$. This assumption is not particularly well motivated, since it is possible, for example, for a multifield model to behave like a single field model in certain regimes, and going backwards in time a transition from single field behavior to multifield behavior can occur while

$k_{\max}/a(\tau) \ll \Lambda^{1/4}$. Nevertheless, if one makes this assumption, then from Eq. (3.4.21) evaluated at $E_{\max} \sim \Lambda^{1/4}$ we obtain

$$N_0 \lesssim \frac{\epsilon\eta\ m_p^2 H^2 \chi^4}{f_n(\chi)\Lambda}. \tag{3.4.32}$$

Next, we demand that the background inflationary solution be within the domain of the effective field theory. The action (2.1.1) will have correction terms such as $(\nabla\phi)^4/\Lambda$ that are suppressed by powers of the cutoff. This correction term must be small compared to $(\nabla\phi)^2$ when evaluated on the background solution, which yields the condition $\epsilon m_p^2 H^2 \ll \Lambda$. Combining this with Eq. (3.4.32) yields

$$N_0 \leq \eta \frac{\chi^4}{f_n(\chi)} = \begin{cases} \frac{\eta\chi^4}{\ln\chi} & n = 0, \\ \\ \frac{\eta\, n\, \chi^4}{\chi^n - 1} & n \neq 0. \end{cases} \tag{3.4.33}$$

Thus the mode occupation number must be small compared to unity if $\chi \sim 1$, assuming $\eta \sim 0.01$. It can be large however for $\chi \gg 1$ if $n < 4$.

3.5 Scalar power spectrum

As we showed in Sec. 2.5.I.a, the scalar power spectrum $\mathcal{P}_\mathcal{R}$ is given by

$$\left\langle \hat{\mathcal{R}}(\tau_0, \mathbf{k}_1)\hat{\mathcal{R}}(\tau_0, \mathbf{k}_2)\right\rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2)\mathcal{P}_\mathcal{R}(\mathbf{k}_1), \tag{3.5.1}$$

where τ_0 is some conformal time at which both modes are superhorizon. Note that unlike the power spectrum computed in the Bunch-Davies vacuum state given in Eq. (2.5.9), the power spectrum here need not be an isotropic functions of $\mathbf{k}$.

The scalar power spectrum has two contributions,

$$\mathcal{P}_\mathcal{R}(\mathbf{k}) = \mathcal{P}_\mathcal{R}^{(0)}(\mathbf{k}) + \mathcal{P}_\mathcal{R}^{(1)}(\mathbf{k}), \tag{3.5.2}$$

a leading order term $\mathcal{P}_\mathcal{R}^{(0)}(\mathbf{k})$ neglecting the effect of interactions, and a subleading term $\mathcal{P}_\mathcal{R}^{(1)}(\mathbf{k})$ arising from the interactions [7]. The subleading term is smaller by a factor $\sim H\sqrt{\epsilon}/m_p$. We will consider both contributions.

[7] See Sec. 3.3 above.

I Leading order power spectrum

For a general homogeneous initial state, we compute the leading order power spectrum by inserting the $\hat{\mathcal{R}}$ mode expansion given in Eq. (2.5.5) into the power spectrum definition (3.5.1), and simplifying using the definitions (3.4.1) of the two point functions $\mathcal{F}_1$ and $\mathcal{F}_2$. This yields[8] [9]

$$\mathcal{P}^{(0)}_{\mathcal{R}}(\mathbf{k}) = \mathcal{P}^{(0)}_{\mathcal{R},0}(k)\big[1 + \mathcal{F}_2(-\mathbf{k}) + \mathcal{F}_2(\mathbf{k}) - 2\mathrm{Re}\{\mathcal{F}_1(\mathbf{k})\}\big], \tag{3.5.3}$$

where $\mathcal{P}^{(0)}_{\mathcal{R},0}$ is the Bunch-Davies power spectrum given in Eq. (2.5.9) and "Re" means the real part of the given expression. The terms involving the free functions in the square brackets indicate the effect of the presence of quanta in the initial state.

The leading order scalar power spectrum (3.5.3) has been calculated for variety of different choices of initial states [85, 96, 98, 111]. For some specific initial states the deviation from the Bunch-Davies vacuum case vanishes [98, 96]. Current observational constraints on the flatness of the power spectrum constrain the initial state, ruling out models where the scale invariance of the power spectrum is strongly violated, for example in [111].

II Subleading contributions to the power spectrum due to interactions

To calculate the power spectrum (3.5.1) to subleading order we use the time-dependent perturbation theory formalism that we discussed in Sec. 2.5.I. Using Eq. (2.5.7), setting $\tau_0 = 0$ and replacing the Bunch-Davies vacuum state with our choice of general initial state, we find at the leading order in the perturbation theory [10]

$$\big\langle \hat{\mathcal{R}}(0,\mathbf{k}_1)\hat{\mathcal{R}}(0,\mathbf{k}_2)\big\rangle = \big\langle \hat{\mathcal{R}}^I(0,\mathbf{k}_1)\hat{\mathcal{R}}^I(0,\mathbf{k}_2)\big\rangle - i\int_{\tau_i}^{0} d\tau \,\Big\langle \Big[\hat{\mathcal{R}}^I(0,\mathbf{k}_1)\hat{\mathcal{R}}^I(0,\mathbf{k}_2), \hat{H}^I_{\rm int}(\tau)\Big]\Big\rangle\,, \tag{3.5.4}$$

where $\hat{H}^I_{\rm int}$ is given in Eq. (3.3.2). As we mentioned in Sec. 2.5.I, the interaction picture operators $\hat{\mathcal{R}}^I$ and $\hat{H}^I_{\rm int}$ consist of the free mode solutions of the Mukhanov-Sasaki equation (2.3.8). We now rewrite the field operator $\hat{\mathcal{R}}$ on the right hand side in terms of the redefined field operator $\hat{\mathcal{R}}_c$

[8] An additional contribution to the change in the measured spectral index is the backreaction effect (3.4.16). Observations constrain the sum of the backreaction contribution and the direct contribution (3.5.3). One could imagine evading the backreaction constraints (3.4.17) via a cancellation between these two contributions, but this would require fine tuning.

[9] See e.g. [110].

[10] We ignore the trace term that appears in the denominator as we did for the Bunch-Davies vacuum state.

using Eq. (3.3.3). We then insert the mode expansion (3.3.4) for the redefined interaction picture operators together with the explicit expression (3.3.2) for the interaction Hamiltonian, and simplify using the notations introduced in Sec. 3.4.I. We neglect any evolution in the slow roll parameters that occurs between the horizon exits of different modes, that is, we treat these slow roll parameters as constants. The explicit calculation is given in App. C.

The final result for the subleading power spectrum is

$$\mathcal{P}_{\mathcal{R}}^{(1)}(\mathbf{k}) = -\frac{iH^3(\frac{1}{2}\epsilon-\eta)}{16m_p^3k^3\epsilon^{\frac{3}{2}}}\Pi_1(\mathbf{k}) - \frac{iH^3}{16m_p^3k^3\epsilon^{\frac{1}{2}}}\Pi_2(\mathbf{k}) + \frac{H^3}{8\epsilon^{\frac{1}{2}}m_p^3k^3}\Pi_3(\mathbf{k}). \tag{3.5.5}$$

Here the functions Π_1, Π_2 and Π_3 are dimensionless functions of momentum. The function Π_1 is given by

$$\begin{aligned}\Pi_1(\mathbf{k}) &= \tilde{\Phi}_1(\mathbf{k}) - \tilde{\Phi}_{2,1}(\mathbf{k}) - \tilde{\Phi}_{2,2}(\mathbf{k}) - \tilde{\Phi}_{2,3}(\mathbf{k}) + \tilde{\Phi}^*_{2,3}(-\mathbf{k}) + \tilde{\Phi}^*_{2,2}(-\mathbf{k}) \\ &\quad +\tilde{\Phi}^*_{2,1}(-\mathbf{k}) - \tilde{\Phi}^*_1(-\mathbf{k}) + (\mathbf{k}\to-\mathbf{k}),\end{aligned} \tag{3.5.6}$$

where

$$\tilde{\Phi}_1(\mathbf{k}) = \int \frac{d^3p}{(2\pi)^{3/2}}\, p^{-3/2}q^{-3/2}k^{3/2}\mathcal{H}_1(\mathbf{k},\mathbf{p}), \tag{3.5.7}$$

$$\tilde{\Phi}_{2,1}(\mathbf{k}) = \int \frac{d^3p}{(2\pi)^{3/2}}\, p^{-3/2}q^{-3/2}k^{3/2}\mathcal{H}_2(\mathbf{k},\mathbf{p}), \tag{3.5.8}$$

$$\tilde{\Phi}_{2,2}(\mathbf{k}) = \int \frac{d^3p}{(2\pi)^{3/2}}\, p^{-3/2}q^{-3/2}k^{3/2}\mathcal{H}_2(\mathbf{k},\mathbf{q}), \tag{3.5.9}$$

$$\tilde{\Phi}_{2,3}(\mathbf{k}) = \int \frac{d^3p}{(2\pi)^{3/2}}\, p^{-3/2}q^{-3/2}k^{3/2}\mathcal{H}_2(\mathbf{p},\mathbf{q}), \tag{3.5.10}$$

and $\mathbf{q} = -\mathbf{k}-\mathbf{p}$. Similarly the function Π_2 is given by

$$\begin{aligned}\Pi_2(\mathbf{k}) &= \bar{\Phi}_1(\mathbf{k}) - \bar{\Phi}_{2,1}(\mathbf{k}) - \bar{\Phi}_{2,2}(\mathbf{k}) - \bar{\Phi}_{2,3}(\mathbf{k}) + \bar{\Phi}^*_{2,3}(-\mathbf{k}) + \bar{\Phi}^*_{2,2}(-\mathbf{k}) \\ &\quad +\bar{\Phi}^*_{2,1}(-\mathbf{k}_1) - \bar{\Phi}^*_1(-\mathbf{k}_1) + (\mathbf{k}\to-\mathbf{k}),\end{aligned} \tag{3.5.11}$$

where the functions $\bar{\Phi}$ are defined by equations analogous to Eqs. (3.5.7) - (3.5.10) but with the factor $p^{-3/2}q^{-3/2}k^{3/2}$ in the integrand replaced with $p^{-3/2}q^{1/2}k^{-1/2}$. Lastly, the function Π_3 in Eq.

(3.5.5) is given by

$$\begin{aligned}\Pi_3(\mathbf{k}) &= \int \frac{d^3p}{(2\pi)^{3/2}}\sqrt{psk}\Big[\frac{1}{k^2}+\frac{1}{p^2}+\frac{1}{s^2}\Big] \\ &\times\Big\{h(p,s)\Big[\mathcal{H}_1(\mathbf{p},\mathbf{s})-\mathcal{H}_2(\mathbf{p},\mathbf{s})\Big]-h(p,-s)\Big[\mathcal{H}_2(-\mathbf{k},\mathbf{p})-\mathcal{H}_2^*(-\mathbf{s},\mathbf{k})\Big] \\ &-h(-p,s)\Big[\mathcal{H}_2(-\mathbf{k},-\mathbf{s})-\mathcal{H}_2^*(-\mathbf{p},\mathbf{k})\Big]+h(-p,-s)\Big[\mathcal{H}_2^*(-\mathbf{p},-\mathbf{s})-\mathcal{H}_1^*(-\mathbf{p},\mathbf{k})\Big]\Big\} \\ &+(\mathbf{k}\to-\mathbf{k}),\end{aligned} \tag{3.5.12}$$

where $\mathbf{s}=\mathbf{k}-\mathbf{p}$. Here the function $h(a,b)$ is

$$\begin{aligned}h(a,b) &= \int_{\tau_i}^0 d\tau\left[e^{-i\tau(a+b+k)}-e^{-i\tau(a+b-k)}\right] \\ &= -\frac{i}{a+b-k}\left[1-e^{-i\tau_i(a+b-k)}\right]+\frac{i}{a+b+k}\left[1-e^{-i\tau_i(a+b+k)}\right].\end{aligned} \tag{3.5.13}$$

We note that the subleading contribution (3.5.5) to the power spectrum depends only on the three point function of the initial state, parameterized by the functions $\mathcal{H}_i$, and so it vanishes for Gaussian initial states. This includes the Bunch-Davies vacuum state.

3.6 Primordial non-Gaussianities: bispectrum of the scalar perturbations

To predict the non-Gaussianity of the primordial fluctuations one needs to consider higher order correlation functions. Recall from Eq. (2.5.13) that the 3-point function of $\hat{\mathcal{R}}$ modes is parametrized by

$$\left\langle\hat{\mathcal{R}}(\tau_0,\mathbf{k}_1)\hat{\mathcal{R}}(\tau_0,\mathbf{k}_2)\hat{\mathcal{R}}(\tau_0,\mathbf{k}_3)\right\rangle=\delta(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\mathcal{B}_{\mathcal{R}}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3), \tag{3.6.1}$$

where $\mathcal{B}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$ is called the bispectrum. Just as in the case of the power spectrum, the conformal time τ_0 is chosen to be any time at which all three modes are superhorizon. Following Eq. (2.5.17), the 3-point function (3.6.1) including the leading non-linear corrections is given by

$$\begin{aligned}\left\langle\hat{\mathcal{R}}(0,\mathbf{k}_1)\hat{\mathcal{R}}(0,\mathbf{k}_2)\hat{\mathcal{R}}(0,\mathbf{k}_3)\right\rangle &= \left\langle\hat{\mathcal{R}}^I(0,\mathbf{k}_1)\hat{\mathcal{R}}^I(0,\mathbf{k}_2)\hat{\mathcal{R}}^I(0,\mathbf{k}_3)\right\rangle \\ &-i\int_{\tau_i}^0 d\tau\left\langle\left[\hat{\mathcal{R}}^I(0,\mathbf{k}_1)\hat{\mathcal{R}}^I(0,\mathbf{k}_2)\hat{\mathcal{R}}^I(0,\mathbf{k}_3),\hat{H}^I_{\rm int}(\tau)\right]\right\rangle.\end{aligned} \tag{3.6.2}$$

This expression is calculated using the same procedure explained below Eq. (3.5.4). The details of the calculation are provided in App. C.

Note that the interaction Hamiltonian (3.3.2) is cubic in the fields, so the second term in Eq. (3.6.2) contains six factors of field operators. However, these factors occur inside a commutator, and when this commutator is evaluated explicitly the number of factors of field operators is reduced to four. Therefore we only require the two, three and four point functions of the initial state, and not any higher point functions. The final result for the bispectrum can be written as follows:

$$\mathcal{B}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \mathcal{B}_{0,\mathrm{dr}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + \mathcal{B}_{II}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + \mathcal{B}_{III}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + \mathcal{B}_{IV}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3). \quad (3.6.3)$$

Here $\mathcal{B}_{0,\mathrm{dr}}$ is the bispectrum for the initial, dressed, Bunch-Davies vacuum, as calculated in [62]. The remaining terms are corrections due to the non-vacuum initial state. The term $\mathcal{B}_{II}$ is determined by the two point function of the initial state, $\mathcal{B}_{III}$ by the three point function, and $\mathcal{B}_{IV}$ by the connected part of the four point function. We now discuss these various contributions in turn.

I Vacuum contribution

The vacuum contribution is [62]

$$\mathcal{B}_{0,\mathrm{dr}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \frac{H^4}{16 m_p^4 \epsilon^2 k_1^3 k_2^3 k_3^3} \left[\frac{1}{2} \epsilon \sigma_1 \sigma_2 - \eta \sigma_3 + 2\epsilon(\sigma_2^2 - \sigma_4)/\sigma_1 \right], \quad (3.6.4)$$

where we have defined

$$\sigma_p \equiv \frac{1}{(2\pi)^{p/2}} \sum_{i=1}^{3} k_i^p, \quad (3.6.5)$$

for $p = 1, 2, 3, 4$. This is the result that applies when $\mathcal{F}_i = \hat{\mathcal{H}}_1 = \mathcal{H}_2 = \Gamma_i = 0$.

II Contribution from two point function

The piece $\mathcal{B}_{II}$ of the bispectrum depends only on the functions $\mathcal{F}_1(\mathbf{k})$ and $\mathcal{F}_2(\mathbf{k})$ that parameterize the initial two point function, given by Eqs. (3.4.1) and (3.4.10) above. For a Gaussian initial state, as studied previously in Refs. [87, 88, 89, 90, 91, 92, 93, 96, 97], this gives the entire bispectrum. The result is

$$\begin{aligned} \mathcal{B}_{II}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= \frac{H^4(\frac{1}{2}\epsilon - \eta)}{16 m_p^4 \epsilon^2 k_1^2 k_2^2 k_3^2} \mathcal{A}_{II}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + \frac{H^4}{32 m_p^4 \epsilon k_1^2 k_2^2 k_3^2} \tilde{\mathcal{A}}_{II}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \\ &\quad - \frac{iH^4}{8 m_p^4 \epsilon k_1^2 k_2^2 k_3^2} \hat{\mathcal{A}}_{II}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3). \end{aligned} \quad (3.6.6)$$

Here the dimensionless amplitude $\mathcal{A}_{II}$ is given by

$$\mathcal{A}_{II} = \frac{k_3^2}{(2\pi)^{3/2} k_1 k_2} \left[\bar{\mathcal{F}}(\mathbf{k}_1)\bar{\mathcal{F}}(\mathbf{k}_2) + \bar{\mathcal{F}}(\mathbf{k}_1) + \bar{\mathcal{F}}(\mathbf{k}_2)\right] + \text{cyclic perms}, \tag{3.6.7}$$

where "cyclic perms" means two terms obtained from the original term by cyclicly permuting $\mathbf{k}_1, \mathbf{k}_2$ and $\mathbf{k}_3$. Also we have defined the function

$$\bar{\mathcal{F}}(\mathbf{k}) \equiv \mathcal{F}_2(\mathbf{k}) + \mathcal{F}_2(-\mathbf{k}) - 2\mathrm{Re}\,\mathcal{F}_1(\mathbf{k}), \tag{3.6.8}$$

which is same combination of $\mathcal{F}_1$ and $\mathcal{F}_2$ that appears in the power spectrum (3.5.3). The dimensionless amplitude $\tilde{\mathcal{A}}_{II}$ is given by

$$\tilde{\mathcal{A}}_{II} = \frac{k_1^2 + k_2^2}{(2\pi)^{3/2} k_1 k_2} \left[\bar{\mathcal{F}}(\mathbf{k}_1)\bar{\mathcal{F}}(\mathbf{k}_2) + \bar{\mathcal{F}}(\mathbf{k}_1) + \bar{\mathcal{F}}(\mathbf{k}_2)\right] + \text{cyclic perms}. \tag{3.6.9}$$

The dimensionless amplitude $\hat{\mathcal{A}}_{II}$ is given by

$$\begin{aligned}
\hat{\mathcal{A}}_{II} = \quad & k_1 k_2 k_3 \sum_{i=1}^{3} \frac{1}{(2\pi)^{3/2} k_i^2} \Bigg(r(-k_t, -\tilde{k}_3)\Big[\mathcal{F}_2(\mathbf{k}_1)\mathcal{F}_2(\mathbf{k}_2) + \mathcal{F}_1(\mathbf{k}_1)\mathcal{F}_1(\mathbf{k}_2)\Big] \\
& + \Bigg\{ r(\tilde{k}_2, -\tilde{k}_1)\Big[\mathcal{F}_2(-\mathbf{k}_1)\mathcal{F}_1(\mathbf{k}_2) + \mathcal{F}_2(\mathbf{k}_2)\mathcal{F}_1^*(\mathbf{k}_1)\Big] + r(-\tilde{k}_3, -k_t)\mathcal{F}_2(\mathbf{k}_2)\mathcal{F}_1(\mathbf{k}_1) \\
& + r(-\tilde{k}_1, \tilde{k}_2)\Big[\mathcal{F}_1(\mathbf{k}_2)\mathcal{F}_1^*(\mathbf{k}_1) + \mathcal{F}_2(\mathbf{k}_2)\mathcal{F}_2(-\mathbf{k}_1)\Big] + r(k_t, \tilde{k}_3)\mathcal{F}_2(-\mathbf{k}_1)\mathcal{F}_1^*(\mathbf{k}_2) + (\mathbf{k}_1 \leftrightarrow \mathbf{k}_2) \Bigg\} \\
& + r(\tilde{k}_3, k_t)\Big[\mathcal{F}_2(-\mathbf{k}_1)\mathcal{F}_2(-\mathbf{k}_2) + \mathcal{F}_1^*(\mathbf{k}_1)\mathcal{F}_1^*(\mathbf{k}_2)\Big] \\
& + r(-k_t, \tilde{k}_3)\mathcal{F}_2(\mathbf{k}_3) + r(-\tilde{k}_3, k_t)\mathcal{F}_2(-\mathbf{k}_3) + r(\tilde{k}_3, -k_t)\mathcal{F}_1(\mathbf{k}_3) + r(k_t, -\tilde{k}_3)\mathcal{F}_1^*(\mathbf{k}_3) \Bigg) \\
& + \text{cyclic perms},
\end{aligned} \tag{3.6.10}$$

where

$$k_t \equiv k_1 + k_2 + k_3, \qquad \tilde{k_i} \equiv k_t - 2k_i. \tag{3.6.11}$$

Here we have defined the function $r(a, b)$ by

$$\begin{aligned}
r(a,b) &= \int_{\tau_i}^{0} d\tau \left[e^{i\tau a} - e^{i\tau b}\right] \\
&= \frac{1}{ia}\left[1 - e^{i\tau_i a}\right] - \frac{1}{ib}\left[1 - e^{i\tau_i b}\right].
\end{aligned} \tag{3.6.12}$$

III Contribution from three point function

The piece $\mathcal{B}_{III}$ of the bispectrum is

$$\mathcal{B}_{III}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = -\frac{iH^3}{8m_p^3\epsilon^{3/2}k_1^2k_2^2k_3^2}\mathcal{A}_{III}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3), \tag{3.6.13}$$

where the dimensionless amplitude $\mathcal{A}_{III}$ is given by

$$\begin{aligned}\mathcal{A}_{III} &= (k_1k_2k_3)^{1/2}\Big[\hat{\mathcal{H}}_1(\mathbf{k}_1,\mathbf{k}_2) - \hat{\mathcal{H}}_1^*(-\mathbf{k}_1,-\mathbf{k}_2) - \mathcal{H}_2(\mathbf{k}_1,\mathbf{k}_2) + \mathcal{H}_2^*(-\mathbf{k}_1,-\mathbf{k}_2) - \mathcal{H}_2(\mathbf{k}_1,\mathbf{k}_3) \\ &\quad +\mathcal{H}_2^*(-\mathbf{k}_1,-\mathbf{k}_3) - \mathcal{H}_2(\mathbf{k}_2,\mathbf{k}_3) + \mathcal{H}_2^*(-\mathbf{k}_2,-\mathbf{k}_3)\Big],\end{aligned} \tag{3.6.14}$$

where the functions $\hat{\mathcal{H}}_1$ and $\mathcal{H}_2$ are defined in Eqs. (3.4.2) and (3.4.10) above.

IV Contribution from connected piece of four point function

This piece of the bispectrum is determined by the functions $\Gamma_i(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$ for $i = 1,2,3$ that parameterize the connected piece of the four point function of the initial state, defined by Eqs. (3.4.3), (3.4.8) and (3.4.10) above. In order to give the result we first define some notations. We define the functions $\Theta_{i,j}(\mathbf{k}_1,\mathbf{k}_2)$ for $1 \le i \le 3$ and $1 \le j \le 4$ by

$$\Theta_{i,1}(\mathbf{k}_1,\mathbf{k}_2) = \int \frac{d^3p}{(2\pi)^{3/2}}\, p^{-3/2}q^{1/2}\Gamma_i(\mathbf{k}_1,\mathbf{k}_2,\mathbf{p}), \tag{3.6.15a}$$

$$\Theta_{i,2}(\mathbf{k}_1,\mathbf{k}_2) = \int \frac{d^3p}{(2\pi)^{3/2}}\, p^{-3/2}q^{1/2}\Gamma_i(\mathbf{p},\mathbf{k}_1,\mathbf{k}_2), \tag{3.6.15b}$$

$$\Theta_{i,3}(\mathbf{k}_1,\mathbf{k}_2) = \int \frac{d^3p}{(2\pi)^{3/2}}\, p^{-3/2}q^{1/2}\Gamma_i(\mathbf{q},\mathbf{k}_1,\mathbf{k}_2), \tag{3.6.15c}$$

$$\Theta_{i,4}(\mathbf{k}_1,\mathbf{k}_2) = \int \frac{d^3p}{(2\pi)^{3/2}}\, p^{-3/2}q^{1/2}\Gamma_i(\mathbf{p},\mathbf{q},\mathbf{k}_1). \tag{3.6.15d}$$

Here on the right hand sides $\mathbf{k}_3 = -\mathbf{k}_1 - \mathbf{k}_2$ and $\mathbf{q} = \mathbf{k}_3 - \mathbf{p}$. We also define barred versions of some of these functions by modifying the weighting factor in the integrand:

$$\bar{\Theta}_{i,1}(\mathbf{k}_1,\mathbf{k}_2) = \int \frac{d^3p}{(2\pi)^{3/2}}\, p^{-3/2}q^{-3/2}\Gamma_i(\mathbf{k}_1,\mathbf{k}_2,\mathbf{p}), \tag{3.6.16a}$$

$$\bar{\Theta}_{i,2}(\mathbf{k}_1,\mathbf{k}_2) = \int \frac{d^3p}{(2\pi)^{3/2}}\, p^{-3/2}q^{-3/2}\Gamma_i(\mathbf{p},\mathbf{k}_1,\mathbf{k}_2), \tag{3.6.16b}$$

$$\bar{\Theta}_{i,4}(\mathbf{k}_1,\mathbf{k}_2) = \int \frac{d^3p}{(2\pi)^{3/2}}\, p^{-3/2}q^{-3/2}\Gamma_i(\mathbf{p},\mathbf{q},\mathbf{k}_1). \tag{3.6.16c}$$

These functions are not all independent; it follows from the identities (3.4.4) – (3.4.6) that $\bar{\Theta}_{1,1} = \bar{\Theta}_{1,2} = \bar{\Theta}_{1,4}$, $\bar{\Theta}_{2,2} = \bar{\Theta}_{2,4}$ $\Theta_{1,1} = \Theta_{1,2} = \Theta_{1,3} = \Theta_{1,4}$, and $\Theta_{2,2} = \Theta_{2,4}$. Also the functions $\bar{\Theta}_{1,1}$, $\bar{\Theta}_{2,2}$, $\bar{\Theta}_{3,1}$, $\bar{\Theta}_{3,4}$, $\Theta_{1,1}$, $\Theta_{2,2}$, $\Theta_{2,3}$, $\Theta_{3,1}$ and $\Theta_{3,4}$ are all symmetric under the interchange of $\mathbf{k}_1$ and $\mathbf{k}_2$.

The result for the bispectrum is

$$\begin{aligned}\mathcal{B}_{IV}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) &= \frac{H^4(\frac{1}{2}\epsilon-\eta)}{32m_p^4\epsilon^2k_1^2k_2^2k_3^2}\mathcal{A}_{IV}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)+\frac{H^4}{32m_p^4\epsilon k_1^2k_2^2k_3^2}\tilde{\mathcal{A}}_{IV}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\\ &\quad -\frac{iH^4}{16m_p^4\epsilon k_1^2k_2^2k_3^2}\hat{\mathcal{A}}_{IV}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3). \qquad (3.6.17)\end{aligned}$$

Here the dimensionless amplitude $\mathcal{A}_{IV}$ is given by

$$\begin{aligned}\mathcal{A}_{IV} &= k_3^2\sqrt{k_1k_2}\Big[\bar{\Theta}_{1,1}(\mathbf{k}_1,\mathbf{k}_2)+\bar{\Theta}^*_{1,1}(-\mathbf{k}_1,-\mathbf{k}_2)-2\bar{\Theta}_{2,2}(\mathbf{k}_1,\mathbf{k}_2)-\bar{\Theta}_{2,1}(\mathbf{k}_2,\mathbf{k}_1)\\ &\quad -\bar{\Theta}^*_{2,1}(-\mathbf{k}_1,-\mathbf{k}_2)-\bar{\Theta}_{2,1}(\mathbf{k}_1,\mathbf{k}_2)-\bar{\Theta}^*_{2,1}(-\mathbf{k}_2,-\mathbf{k}_1)-2\bar{\Theta}^*_{2,2}(-\mathbf{k}_1,-\mathbf{k}_2)+\bar{\Theta}_{3,4}(\mathbf{k}_1,\mathbf{k}_2)\\ &\quad +2\bar{\Theta}_{3,2}(\mathbf{k}_2,\mathbf{k}_1)+2\bar{\Theta}_{3,2}(\mathbf{k}_1,\mathbf{k}_2)+\bar{\Theta}_{3,1}(\mathbf{k}_1,\mathbf{k}_2)\Big] \quad +\text{cyclic perms.} \qquad (3.6.18)\end{aligned}$$

The dimensionless amplitude $\tilde{\mathcal{A}}_{IV}$ is given by

$$\begin{aligned}\tilde{\mathcal{A}}_{IV} &= \sqrt{k_1k_2}\Big[\Theta_{1,1}(\mathbf{k}_1,\mathbf{k}_2)+\Theta^*_{1,1}(-\mathbf{k}_1,-\mathbf{k}_2)-\Theta_{2,2}(\mathbf{k}_1,\mathbf{k}_2)-\Theta_{2,3}(\mathbf{k}_1,\mathbf{k}_2)-\Theta_{2,1}(\mathbf{k}_2,\mathbf{k}_1)\\ &\quad -\Theta^*_{2,1}(-\mathbf{k}_1,-\mathbf{k}_2)-\Theta_{2,1}(\mathbf{k}_1,\mathbf{k}_2)-\Theta^*_{2,1}(-\mathbf{k}_2,-\mathbf{k}_1)-\Theta^*_{2,2}(-\mathbf{k}_1,-\mathbf{k}_2)\\ &\quad -\Theta^*_{2,3}(-\mathbf{k}_1,-\mathbf{k}_2)+\Theta_{3,4}(\mathbf{k}_1,\mathbf{k}_2)+\Theta_{3,2}(\mathbf{k}_2,\mathbf{k}_1)+\Theta_{3,3}(\mathbf{k}_2,\mathbf{k}_1)+\Theta_{3,2}(\mathbf{k}_1,\mathbf{k}_2)\\ &\quad +\Theta_{3,3}(\mathbf{k}_1,\mathbf{k}_2)+\Theta_{3,1}(\mathbf{k}_1,\mathbf{k}_2)\Big] \quad +\text{cyclic perms.} \qquad (3.6.19)\end{aligned}$$

Finally the dimensionless amplitude $\hat{\mathcal{A}}_{IV}$ is given by

$$\begin{aligned}\hat{\mathcal{A}}_{IV} &= \sqrt{k_1k_2}k_3\int\frac{d^3p}{(2\pi)^{3/2}}\sqrt{pq}\Big[\frac{1}{k_3^2}+\frac{1}{p^2}+\frac{1}{q^2}\Big]\Big\{h(p,q)\Big[\Gamma_1(\mathbf{k}_1,\mathbf{k}_2,\mathbf{p})+\Gamma_3(\mathbf{k}_1,\mathbf{k}_2,\mathbf{p})\Big]\\ &\quad -h(p,-q)\Big[\Gamma_2(\mathbf{q},\mathbf{k}_1,\mathbf{k}_2)+\Gamma_2^*(-\mathbf{p},-\mathbf{k}_1,-\mathbf{k}_2)\Big]-h(-p,q)\Big[\Gamma_2(\mathbf{p},\mathbf{k}_1,\mathbf{k}_2)+\Gamma_2^*(-\mathbf{q},-\mathbf{k}_1,-\mathbf{k}_2)\Big]\\ &\quad +h(-p,-q)\Big[\Gamma_3(\mathbf{q},\mathbf{p},\mathbf{k}_1)+\Gamma_1^*(-\mathbf{p},-\mathbf{k}_1,-\mathbf{k}_2)\Big]-h(p,q)\Big[\Gamma_2(\mathbf{k}_1,\mathbf{k}_2,\mathbf{p})+\Gamma_2(\mathbf{k}_2,\mathbf{k}_1,\mathbf{p})\Big]\\ &\quad +h(p,-q)\Big[\Gamma_3(\mathbf{q},\mathbf{k}_1,\mathbf{k}_2)+\Gamma_3(\mathbf{q},\mathbf{k}_2,\mathbf{k}_1)\Big]+h(-p,q)\Big[\Gamma_3(\mathbf{p},\mathbf{k}_1,\mathbf{k}_2)+\Gamma_3(\mathbf{p},\mathbf{k}_2,\mathbf{k}_1)\Big]\\ &\quad -h(-p,-q)\Big[\Gamma_2^*(-\mathbf{k}_1,-\mathbf{k}_2,-\mathbf{p})+\Gamma_2^*(-\mathbf{k}_2,-\mathbf{k}_1,-\mathbf{p})\Big]\Big\}\\ &\quad +\text{cyclic perms}, \qquad (3.6.20)\end{aligned}$$

where as before $\mathbf{k}_3=-\mathbf{k}_1-\mathbf{k}_2$ and $\mathbf{q}=\mathbf{k}_3-\mathbf{p}$. Also we have defined the function $h(a,b)$ by

$$\begin{aligned}h(a,b) &= \int_0^{\tau_i}d\tau\left[e^{-i\tau(a+b+k_3)}-e^{-i\tau(a+b-k_3)}\right]\\ &= \frac{i}{a+b-k_3}\left[1-e^{-i\tau_i(a+b-k_3)}\right]-\frac{i}{a+b+k_3}\left[1-e^{-i\tau_i(a+b+k_3)}\right]. \qquad (3.6.21)\end{aligned}$$

3.7 Order of magnitude estimates and discussion

We now discuss the implications of our results for the bispectrum by making some order of magnitude estimates. The results depends on a number of parameters and quantities:

- The ratio of the Hubble and Planck scales, H/m_p.
- The Hubble slow-roll parameters ϵ and η. For our order of magnitude estimates we will assume that $\epsilon \sim \eta$.
- In the isotropic case, the dependence on the wave vectors $\mathbf{k}_1$, $\mathbf{k}_2$ and $\mathbf{k}_3$ can be parametrized in terms of a dependence on the overall scale

$$k \equiv (k_1 k_2 k_3)^{1/3}$$

and the dependence on the two parameters $x_2 = k_2/k_1$ and $x_3 = k_3/k_1$ which characterizes the "shape" of the bispectrum [11].

I Generic initial states

Let us start off by neglecting the shape dependence, assuming that $k_1 \sim k_2 \sim k_3 \sim k$. In this case the vacuum bispectrum (3.6.4) scales as $\mathcal{B}_{0,\mathrm{dr}} \sim H^4/(m_p^4 \epsilon k^6)$. Suppose now that the occupation number of the initial state is of order unity. Suppose also that the initial state is "generic" in the sense that all the dimensionless amplitudes $\mathcal{A}$ defined in Sec. 3.6 are of order unity, so that, for example, the three point function is not suppressed compared to the two point function, etc. If we are agnostic about the origin of the non Bunch-Davies initial state, then this genericity assumption is well motivated. However, specific scenarios for generating initial states can give rise to suppression of the three point function with respect to the two and four point functions, and violate our genericity assumption. For example, this can occur in scenarios where the non Bunch-Davies initial states arise as a result of new physics at high energies which has been integrated out [12]. This important class of scenarios yields non generic initial states. Nevertheless, for the remainder of this subsection we will restrict attention to generic initial states.

[11] See Eq. (2.5.14) for definition of the shape function $\mathcal{S}$ and the amplitude of non-Gaussianities f_{NL}.
[12] See e.g. [112].

With the assumption that all the dimensionless amplitudes $\mathcal{A}$ are of order unity, we find that the results (3.6.6) and (3.6.17) for the contributions from the two and four point functions are of the same order as the vacuum contribution:

$$\mathcal{B}_{II} \sim \mathcal{B}_{IV} \sim \frac{H^4}{m_p^4 \epsilon k^6} \sim \mathcal{B}_{0,\mathrm{dr}}. \tag{3.7.1}$$

These contributions are therefore small and hard to detect. By contrast, the contribution (3.6.13) from the three point function scales as

$$\mathcal{B}_{III} \sim \frac{H^3}{m_p^3 \epsilon^{3/2} k^6} \sim \frac{m_p}{H\sqrt{\epsilon}}\, \mathcal{B}_{0,\mathrm{dr}}. \tag{3.7.2}$$

We now eliminate m_p/H in favor of ϵ and the power spectrum $\Delta_{\mathcal{R}}$ using Eq. (2.5.10). This gives

$$\mathcal{B}_{III} \sim \frac{1}{\epsilon\sqrt{8\pi^2 \Delta_{\mathcal{R}}^2}} \mathcal{B}_{0,\mathrm{dr}} \sim (3 \times 10^5)\ \mathcal{B}_{0,\mathrm{dr}}, \tag{3.7.3}$$

where we have used the estimate $\epsilon \sim 0.01$ and the measured value $\Delta_{\mathcal{R}} \sim 3 \times 10^{-5}$.

Thus, the dominant non-Gaussianity for initial states with $N_{\mathrm{occ}} \sim \mathcal{A}_{III} \sim 1$ is due to the three point function of the initial state. This is easy to understand: a nonzero bispectrum is obtained from an initial three point function from just the linear evolution, without requiring any nonlinearities. The contributions to the bispectrum from the initial two and four point functions, on the other hand, are suppressed since they require the nonlinearities in the dynamics. The enhancement factor of $m_p/(H\sqrt{\epsilon})$ in Eq. (3.7.2) was previously obtained in a special case in [86]. Note that this contribution to the bispectrum vanishes identically for Gaussian initial states, as studied in many previous investigations [87, 88, 89, 90, 91, 92, 93, 96, 97].

The result (3.7.3) could yield a large bispectrum, detectable if $\mathcal{A}_{III} \gtrsim 10^{-3}$. Assuming that $\mathcal{A}_{III} \sim N_{\mathrm{occ}}$, the bispectrum could be detectable even for initial occupation numbers small compared to unity, and $N_{\mathrm{occ}} \gtrsim 1$ is compatible with the backreaction constraint (3.4.24) for suitable values of the spectral index n and bandwidth χ [13].

It is not possible to give a generic prediction for the shape dependence of this dominant piece $\mathcal{B}_{III}$ of the bispectrum, since the shape dependence is just inherited from that of the initial state. Different scenarios for the origin of the non-vacuum initial state will yield different shape dependences.

[13] On the other hand $N_{\mathrm{occ}} \gtrsim 1$ would give order-unity corrections to the power spectrum (3.5.3). This would disagree with observations unless the corrected spectrum is nearly scale invariant, which would be a fine tuning.

II Initial states with vanishing three point function

We now specialize to initial states for which the three point function is small or vanishing, for example Gaussian states or statistical mixtures of mode occupation number eigenstates [85]. For such states $\mathcal{B}_{III}$ vanishes, and we argued above that when $k_1 \sim k_2 \sim k_3$ and $N_{\rm occ} \sim 1$ we have

$$\mathcal{B}_{II} \sim \mathcal{B}_{IV} \sim \mathcal{B}_{0,\rm dr}, \tag{3.7.4}$$

so that the non-Gaussianity is small in this regime. Nevertheless it is still possible that $\mathcal{B}_{II}$ and/or $\mathcal{B}_{IV}$ can be large in various limits where $k_1 \sim k_2 \sim k_3$ is violated or $N_{\rm occ} \gg 1$. It is not possible to estimate how $\mathcal{B}_{IV}$ scales in such limits, since it depends on unknown properties of the connected part of the initial four point function. Therefore in the remainder of this section we will restrict attention to states for which the bispectrum is dominated by $\mathcal{B}_{II}$.

II.a Large occupation number regime

Let us first consider the possibilities for an enhanced bispectrum in the regime $N_{\rm occ} \gg 1$, assuming for simplicity that $k_1 \sim k_2 \sim k_3 \sim k$. In this regime the vacuum bispectrum (3.6.4) scales as $\mathcal{B}_{0,\rm dr} \sim H^4/(m_p^4 \epsilon k^6)$, and from Eq. (3.6.6) we obtain

$$\mathcal{B}_{II}/\mathcal{B}_{0,\rm dr} \sim \left[N_{\rm occ}(k) \quad \& \quad N_{\rm occ}^2(k)\right]. \tag{3.7.5}$$

Here the notation inside the square brackets means that there are two types of term that arise, terms proportional to $N_{\rm occ}(k_1) \sim \mathcal{F}_1(k_1) \sim \mathcal{F}_2(k_1) \sim N_{\rm occ}(k_2) \sim N_{\rm occ}(k)$, and terms proportional to the products $N_{\rm occ}(k_1) N_{\rm occ}(k_2) \sim N_{\rm occ}(k)^2$.

We see from Eq. (3.7.5) that a bispectrum much larger than the Bunch-Davies bispectrum requires large occupation numbers, for initial states with vanishing three point function. Occupation numbers that exceed unity are compatible with the backreaction constraint. However, as discussed in Sec. 3.4.III above, large occupation numbers are only possible over very narrow bandwidths, which is a kind of fine tuning. In addition occupation numbers in excess of unity over a range of wavenumbers is only possible if the range of wavenumbers does not exceed about one and half orders of magnitude [14]. Finally, as noted above, occupation numbers $\gtrsim 1$ would give order unity corrections to the power spectrum (3.5.3), which would disagree with observations unless the corrected spectrum

[14] See Eq. (3.4.30) above.

conspires to be nearly scale invariant. Taken together, these constraints imply that it is difficult for the bispectrum to be significantly larger than the vacuum bispectrum in this regime.

II.b Limiting shape regimes

Consider now squeezed triangle configurations, $k_1 \sim k_2 \gg k_3$. The vacuum bispectrum in this regime is, from Eq. (3.6.4),

$$\mathcal{B}_{0,\mathrm{dr}} \sim \frac{H^4}{m_p^4 \epsilon k_1^3 k_3^3}. \tag{3.7.6}$$

The limiting behaviour of $\mathcal{B}_{II}$ in this limit depends on the direction in which the limit is approached, which can be along the edge of the triangle corresponding to elongated configurations, or from the interior of the triangle. If we assume a direction from the interior, the dominant piece of $\mathcal{B}_{II}$ is given by the amplitude $\hat{\mathcal{A}}_{II}$. We find from Eq. (3.6.10) that

$$\mathcal{B}_{II} \sim \frac{H^4}{m_p^4 \epsilon k_1^3 k_3^3} \left\{ \frac{k_1}{k_3} \left[N_{\mathrm{occ}}(k_1) \quad \& \quad N_{\mathrm{occ}}(k_1)^2 \quad \& \quad N_{\mathrm{occ}}(k_1) N_{\mathrm{occ}}(k_3) \right], \quad N_{\mathrm{occ}}(k_3) \right\}. \tag{3.7.7}$$

Here the notation inside the square brackets means that there are three types of term that arise which scale $\propto k_3^{-4}$, terms proportional to $N_{\mathrm{occ}}(k_1) \sim N_{\mathrm{occ}}(k_2) \sim \mathcal{F}_1(k_1) \sim \mathcal{F}_2(k_1)$, terms proportional to the products $N_{\mathrm{occ}}(k_1) N_{\mathrm{occ}}(k_2) \sim N_{\mathrm{occ}}(k_1)^2$, and terms proportional to the product $N_{\mathrm{occ}}(k_1) N_{\mathrm{occ}}(k_3)$. Also there are terms proportional to $\mathcal{F}_1(k_3) \sim \mathcal{F}_2(k_3) \sim N_{\mathrm{occ}}(k_3)$ which scale $\propto k_3^{-3}$.

If we now assume that all the occupation numbers are of order unity, we obtain

$$\mathcal{B}/\mathcal{B}_{0,\mathrm{dr}} \sim k_1/k_3. \tag{3.7.8}$$

Since k_1/k_3 can be as large as ~ 100 for modes probed in the CMB, and even larger for modes probed by the LSS, the estimate (3.7.8) is a significant enhancement of non-Gaussianity over the standard vacuum result, as previously argued in a special case for the class of squeezed vacuum states [85, 92].

However, the assumption that all mode occupation numbers are of order unity is quite restrictive and can conflict with the backreaction constraints discussed in Sec. 3.4.III. We now investigate how large the enhancement factor (3.7.8) can be assuming the power law model (3.4.19) is valid over all scales from k_3 to $k_1 \sim k_2$. We will find that the backreaction constraints imply that the enhancement factor cannot exceed ~ 200.

The backreaction constraints will be weakest when $k_{\rm min} \sim k_3$ and $k_{\rm max} \sim k_1$, so we will assume this in what follows. Consider first the first term in square brackets in Eq. (3.7.7), $N_{\rm occ}(k_1)$, which is constrained by Eq. (3.4.26). Combining Eqs. (3.7.8) and (3.4.26) gives that this contribution to the bispectrum is bounded above by $\mathcal{B}_{II}/\mathcal{B}_{0,\rm dr} \lesssim 10^5 n/(\chi^3(1-\chi^{-n}))$, which is of order unity or smaller for $|n| \leq 10$ and $\chi = k_{\rm max}/k_{\rm min} \sim k_1/k_3 \gtrsim 10^2$. A similar argument applies to the second term in the square brackets in Eq. (3.7.8). The largest term is the third term, which from Eqs. (3.4.24), (3.4.26) and (3.7.8) is bounded above by

$$\frac{\mathcal{B}_{II}}{\mathcal{B}_{0,\rm dr}} \sim \chi N_0 N_1 \lesssim \frac{10^{10}\chi^{n-3}}{f_n(\chi)^2} \sim \frac{10^{10}\chi^{n-3}n^2}{(\chi^n-1)^2} \lesssim \frac{10^{10}}{\chi^3(\ln\chi)^2}, \tag{3.7.9}$$

where we have used the fact that the maximum of the function of n is achieved at $n = 0$. The upper bound (3.7.9) can be quite large for $\chi \sim 1$, but that regime corresponds to measuring the bispectrum over a narrow range of scales, and choosing the initial state to populate only those scales, which would be a considerable fine tuning. The upper bound (3.7.9) is smaller than χ, invalidating the estimate (3.7.8), when $\chi \gtrsim 200$.

We now turn to a discussion of the the edge of the triangle corresponding to elongated configurations, i.e. the line $k_1 = k_2 + k_3$. There is an enhancement of the bispectrum for these configurations. For the special case of squeezed states [15], this enhancement was previously pointed out in [89], and studied in more detail in [93]. The enhancement arises from the four terms in Eq. (3.6.10) in which one of the arguments of the function $r(a,b)$ is $\pm\tilde{k}_1$. Normally the phases $k\tau_i$ in the formula (3.6.12) for $r(a,b)$ have absolute values which are large compared to unity, since we have assumed that all the modes are inside the horizon at the initial time, $k_j|\tau_i| \gg 1$. Therefore the function r scales as $r \sim 1/k$. However, for elongated configurations, $\tilde{k}_1 \to 0$ from Eq. (3.6.11), and so for example the function $r(\tilde{k}_1, -\tilde{k}_2)$ scales as $\sim |\tau_i|$, enhanced by a factor $k|\tau_i| \gg 1$. We find from Eqs. (3.6.10) and (3.6.12) that in the elongated limit[16]

$$\frac{\mathcal{B}_{II}}{\mathcal{B}_{0,\rm dr}} \sim |\tau_i| k_1 W[k_1|\tau_i|(x_2+x_3-1)]\big[N_{\rm occ} \;\;\&\;\; N_{\rm occ}^2\big], \tag{3.7.10}$$

where $W(x) = (1-e^{-ix})/(ix)$. The factor $W[k_1|\tau_i|(x_2+x_3-1)]$ goes to unity along the edge $k_2 + k_3 = k_1$ of the triangle.

[15] See Sec. 3.4.II.

[16] The same estimate also applies to the squeezed limit approached along the edge of the triangle, i.e. via elongated configurations. See Fig. 7.

How large can the enhancement factor (3.7.10) be? The factors of occupation number are limited by the backreaction constraint, just as discussed above for squeezed configurations. The function W is of order unity or smaller. The factor $k_1|\tau_i|$ can apparently be arbitrarily large, since the initial conformal time τ_i can be chosen to be as early as desired. However, this large τ_i divergence is only apparent, it is an artifact of how we have chosen to parametrize the initial state, since it diverges as $\tau_i \to -\infty$ [17]. If we specify the initial state at some time τ_i, then changing the initial time to some earlier time τ_j will have no affect on the bispectrum if we choose the initial state at τ_j to be that obtained from the initial state at τ_i by evolving backwards in time from τ_i to τ_j. Our simple power law model (3.4.18) and (3.4.19) for the initial state is not invariant under time evolution when interactions are included, which is the reason for the τ_i dependence of the bispectrum (3.6.3). To avoid this spurious time dependence, we will choose the value of τ_i to be the latest possible time for which all the modes under consideration are inside the horizon. This choice gives $|\tau_i| \sim 1/k_{\rm min}$, so $|\tau_i|k_1 \sim |\tau_i|k_{\rm max} \sim k_{\rm max}/k_{\rm min} \sim \chi$. With this choice the enhancement factor (3.7.10) for elongated configurations is of the same order as the enhancement factor (3.7.9) for the squeezed configurations approached from the interior of the triangle [18].

II.c Observational constraints

We now discuss observational constraints on the class of states under discussion in this section, where the bispectrum is dominated by the initial two point function. As discussed above, it is not possible to make detailed predictions for the cases where the bispectrum is dominated by the initial three or four point functions.

For simplicity, we make the following assumptions and specializations. We assume that the function $\mathcal{F}_1$ vanishes and $\mathcal{F}_2 = N_{\rm occ}$. For the occupation number $N_{\rm occ}$ we use the model (3.4.19) specialized to $n = 4$. Then the last piece [19] of the result (3.6.6) for the bispectrum simplifies to [20]

$$\mathcal{B}_{II} \approx \frac{18}{5}\, \mathcal{P}^2(k)\big(10N_0[N_0+1]\epsilon\big)\frac{1}{9k^3}\left[(k_1^2k_2^2 + k_3^2k_1^2 + k_3^2k_2^2)\sum_{i=1}^{3}\frac{1-\cos\tilde{k}_i\tau_i}{\tilde{k}_i}\right], \qquad (3.7.11)$$

where we have omitted terms proportional to $(1-\cos k_t\tau_i)/k_t$ by virtue of being subdominant to

[17] This aspect was missed in the analysis of [89].
[18] See Fig. 7.
[19] Given by $\hat{\mathcal{A}}_{II}$.
[20] We dropped factors of 2π as they emerge due to our choice of conventions.

the other terms in the triangular regimes of interest. Also given our specializations $\mathcal{F}_1 = 0$ and $\mathcal{F}_2 = N_{\rm occ}$, the power spectrum $\mathcal{P}(k)$ in Eq. (3.7.11) above differs from the Bunch-Davies power spectrum (2.5.9). From Eq. (3.7.11) and the definition of $f_{\rm NL}$ provided in Eq. (2.5.14) we have

$$f_{\rm NL} = 10N_0(N_0+1)\epsilon. \tag{3.7.12}$$

The square bracket on the right of Eq. (3.7.11) includes the shape functions

$$\mathcal{S}^1_{\rm NBD} \equiv \frac{1}{2}k_1^2(k_2^2+k_3^2)\ \frac{1-\cos\tilde{k_1}\tau_i}{\tilde{k_1}} + (\text{cyclic perm.}) \tag{3.7.13}$$

and

$$\mathcal{S}^2_{\rm NBD} \equiv (k_2^2k_3^2)\ \frac{1-\cos\tilde{k_1}\tau_i}{\tilde{k_1}} + (\text{cyclic perm.}) \tag{3.7.14}$$

which were first introduced as shape function ansatzs in [85]. Here "NBD" stands for non Bunch-Davies. They have been analysed by the Planck collaboration in [61]. They obtained upper bounds on the corresponding $f_{\rm NL}$ parameters which they define as

$$\mathcal{B}_{\rm NBD} = \frac{2}{k^3}\ \mathcal{P}^2(k) f^{\rm NBDi}_{\rm NL}\ \mathcal{S}^i_{\rm NBD}. \tag{3.7.15}$$

The parameter $f^{\rm NBD1}_{\rm NL}$ has a shape that peaks in folded configurations, while the parameter $f^{\rm NBD2}_{\rm NL}$ has a shape that peaks in squeezed configurations. Comparing our bispectrum in Eq. (3.7.11) to the Bunch-Davies bispectrum ansatz in Eqs. (3.7.13) and (3.7.14) we find

$$f^{\rm NBD1}_{\rm NL} = \frac{2}{5}f_{\rm NL}, \qquad f^{\rm NBD2}_{\rm NL} = \frac{1}{5}f_{\rm NL}. \tag{3.7.16}$$

The Planck collaboration's combined temperature and E-mode polarization anisotropy bounds are

$$f^{\rm NBD1}_{\rm NL} = -10 \pm 19, \qquad f^{\rm NBD2}_{\rm NL} = -0.8 \pm 5.5. \tag{3.7.17}$$

The corresponding constraint on the initial occupation number N_0 obtained from $f^{\rm NBD2}_{\rm NL}$ is more stringent than the one obtained from $f^{\rm NBD1}_{\rm NL}$. Specifically, for $\epsilon \sim 0.01$, the bounds (3.7.17) give the upper bound $N_0 \lesssim 12$ on the initial occupation number. This bound is only more stringent than the one given in Eq. (3.4.24) if $\chi \lesssim 10$. When both N_0 and χ are ~ 10, the power spectrum (3.5.3) can have a large scale dependence unless a cancellation between various scale dependent terms occurs.

Chapter 4

Robustness Of The Standard Inflationary Predictions

4.1 Overview

As we mentioned at the beginning of chapter 3, the robustness of the standard predictions of inflation depends in part on the assumption that single field slow-roll inflation started considerably before the horizon exit of the largest modes we can observe today. This occurs in any model where the total number N of e-folds of inflation is significantly larger than the minimum number $N_{\rm min} \approx 60$ [1]. In many models $N \gg N_{\rm min}$ is very natural, while $N \approx N_{\rm min} +$ (a few) is somewhat unnatural or fine tuned. Nevertheless it is still interesting to consider the possibility that $N \approx N_{\rm min} +$ (a few), since it opens up opportunities for testing the predictions of a variety of pre-inflationary dynamics. The well-known discrepancy between the predicted and observed spectrum of the CMB temperature anisotropies for multipoles $l \lesssim 40$ serves as a motivation in this direction [2].

In chapter 3, we used the non Bunch-Davies vacuum initial states formalism to model the influence of complicated pre-inflationary phases on the inflationary correlation functions. In this chapter, we will approach this problem by explicitly incorporating an extra degree of freedom in a model where a 60 e-fold single field inflation is preceded by an era of radiation fluid domination [3]. Here we will compute the 2-point correlation functions of the primordial perturbations this time generated by the radiation-inflaton-gravity system and discuss the observability of their distinct

[1] Assuming GUT-scale inflation.

[2] Several different scenarios aiming to explain this anomaly have been explored in the literature, including preceding multifield phases [113], initial non slow-roll phases [114], and quantum mechanical tunneling from other vacua in a string theory landscape [115, 116].

[3] This cosmological model has a long history. See for instance [117, 118, 119] for the study of the effects of this cosmological phase transition on the stability of quantum fields.

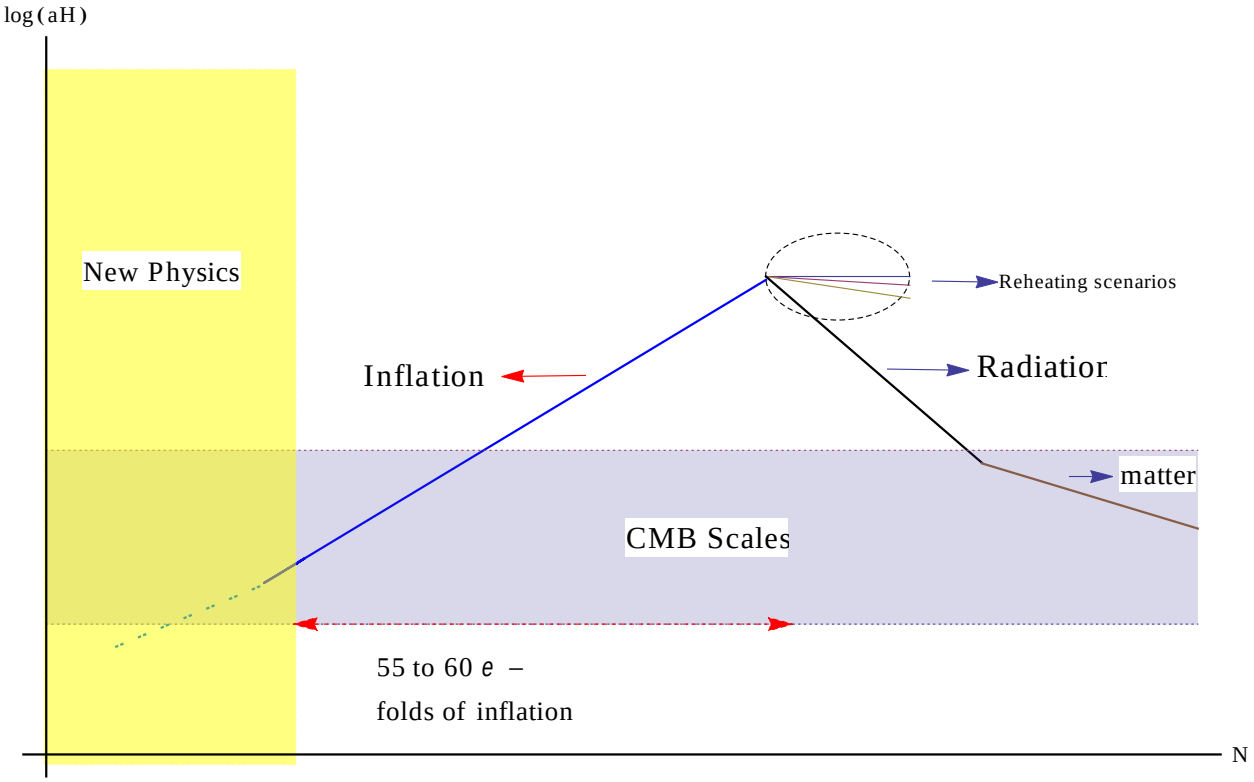


Figure 14: The is a qualitative plot of the comoving Hubble scale aH during inflation and afterwards. The currently observable long-wavelength modes could have exited the Hubble horizon about 60 e-folds prior to the end of inflation. For a slow-roll inflation that lasts for about 60 e-folds, these long wavelength modes could carry valuable information regarding pre-inflationary dynamics.

signatures.

This chapter is organized as follows: In Sec. 4.2 we will discuss the properties of the background dynamics of this model, highlighting the existence of two distinguished transitions in the background evolution. Next, in Sec. 4.3 we will derive the equations of motion for the scalar and tensor perturbations. In Sec. 4.4 we will discuss the adiabaticity of the perturbations in the distant past. We will then argue for the suppression of the entropic perturbation modes in Sec. 4.5. We will conclude in Secs. 4.6 and 4.7 with a few numerical case studies of the spectrum of tensor and scalar perturbations in this model.

4.2 The fluid-inflaton model

We consider a cosmological model for before and during inflation consisting of a perfect isentropic fluid and a scalar inflaton field ϕ that are coupled only gravitationally. The action for the model is

$$S[g_{ab},\phi,J^a]=S_{\rm EH}[g_{ab}]+S_{\rm inf}[g_{ab},\phi]+S_{\rm fluid}[g_{ab},J^a]=\int d^4x\,\sqrt{-g}\,\Big[\frac{m_p^2}{2}R-\frac{1}{2}(\nabla\phi)^2-V(\phi)-\rho_{\rm f}(|\vec{J}|)\Big], \tag{4.2.1}$$

where R is the Ricci scalar, $\rho_{\rm f}$ is the fluid energy density, and $\vec{J}$ is the number density current of the fluid [4]. The corresponding equations of motion are the Klein-Gordon equation and the Einstein equations given in Eq. (2.1.2) with the stress-energy tensor being

$$T_{ab}^{\rm total}=T_{ab}^{\rm inf}+T_{ab}^{\rm fluid}, \tag{4.2.2}$$

where $T_{ab}^{\rm inf}$ is given in Eq. (2.1.3) and

$$T_{ab}^{\rm fluid}=(\rho_{\rm f}+p_{\rm f})u_a u_b+p_{\rm f}g_{ab}, \tag{4.2.3}$$

and the fluid equations

$$\nabla_a T^{{\rm fluid}\,ab}=0. \tag{4.2.4}$$

Here $p_{\rm f}$ and u^a are the fluid pressure and 4-velocity. We will later specialize the equation of state $p_{\rm f}=p_{\rm f}(\rho_{\rm f})$ of the fluid to be that of radiation, $p_{\rm f}=\rho_{\rm f}/3$.

In this section we will analyse the background cosmological solutions of this model, and in Sec. 4.3 below we will analyse the dynamics of the linear scalar and tensor perturbations.

I Background cosmology

For a homogeneous, isotropic and spatially flat background spacetime we use the FRW metric (1.1.1) with $k=0$. The background equations are the Friedmann equations

$$3m_p^2\frac{\mathcal{H}^2}{a^2}=\bar{\rho}_{\rm f}+\frac{1}{2a^2}\bar{\phi}'^2+V(\bar{\phi}) \tag{4.2.5}$$

$$\frac{m_p^2}{a^2}\big(2\mathcal{H}'+\mathcal{H}^2\big)=-\bar{p}_{\rm f}-\frac{1}{2a^2}\bar{\phi}'^2+V(\bar{\phi}), \tag{4.2.6}$$

[4] See Sec. 4.3.I below.

where primes denotes differentiation with respect to the conformal time τ and $\mathcal{H} \equiv a'/a$ is the conformal Hubble parameter, and the background Klein-Gordon and fluid continuity equations

$$\bar{\phi}'' + 2\mathcal{H}\bar{\phi}' + a^2 V_{,\phi}(\bar{\phi}) = 0, \tag{4.2.7}$$

$$\bar{\rho}_{\rm f}' + 3\mathcal{H}(\bar{\rho}_{\rm f} + \bar{p}_{\rm f}) = 0, \tag{4.2.8}$$

where $\bar{\rho}_{\rm f}$ and $\bar{p}_{\rm f}$ are the background energy density and pressure of the fluid. The continuity equation (4.2.8) yields

$$\bar{\rho}_{\rm f} = \frac{\mathcal{E}_0}{a^{3(1+w)}}, \tag{4.2.9}$$

for an equation of state $\bar{p}_{\rm f} = w\bar{\rho}_{\rm f}$, and a constant of integration $\mathcal{E}_0$. In our analysis of the background dynamics we will mainly use the first Friedmann equation (4.2.5) and the two equations (4.2.7) and (4.2.8).

II The background dynamics and parameters

The background fluid-inflaton Lagrangian depends on a number of parameters, namely the equation of state parameter w of the fluid as well as all of the coupling parameters in the inflaton potential. The space of solutions of the differential equations (4.2.5), (4.2.7) and (4.2.8) is four dimensional, requiring four initial conditions to determine a solution. One of these parameters is gauge, since we can resale the coordinates to set the value of a at the initial time of integration to unity. The remaining three parameters can be taken to be the initial value and the initial time derivative of the inflaton, and the ratio $\bar{\rho}_{\rm inf}/\bar{\rho}_{\rm f}$ of the initilal energy densities of the inflaton and fluid.

In exploring this space of model parameters, our strategy will be to choose models that agree with observations in the absence of a fluid, and then gradually modify the model by dialing the parameters related to the inflaton at some initial time when the energy density of the fluid is Planckian. In particular we will restrict attention to models in which, in the absence of fluid, inflation occurs at the GUT scale of $\sim 10^{16}\,$GeV. We will use as an example the quadratic potential $V = m^2\phi^2/2$ to construct numerical examples of the power spectra of the scalar and tensor perturbations in sections 4.6.I and 4.7.I.

To illustrate the qualitative features of the the background dynamics, Fig. 15 shows the evolution of the fluid and inflaton energy densities, the Hubble parameter $H = \mathcal{H}/a$, and the inflaton field $\bar{\phi}$.

For this example we have taken $V(\phi) = m^2\phi^2/2$ with $m = 10^{-5}m_p$, and for the fluid we have taken $w = 1/3$ (radiation) and used a Planckian initial energy density. The initial inflaton kinetic and potential energies are

$$\frac{\bar{\phi}'^2}{2a^2} \approx V(\bar{\phi}) \approx 10^{-8}m_p^4, \tag{4.2.10}$$

which is about the GUT energy density. As seen in Fig. 15, the background dynamics starts in a radiation dominated era. The energy densities become equal at $N \approx 5$, after which the energy density of the radiation keeps decreasing rapidly while the background dynamics becomes similar to that of single field slow-roll inflation. The ensuing near constancy of the Hubble parameter bears evidence to the Universe expanding quasi-exponentially, as expected in the standard inflationary paradigm.

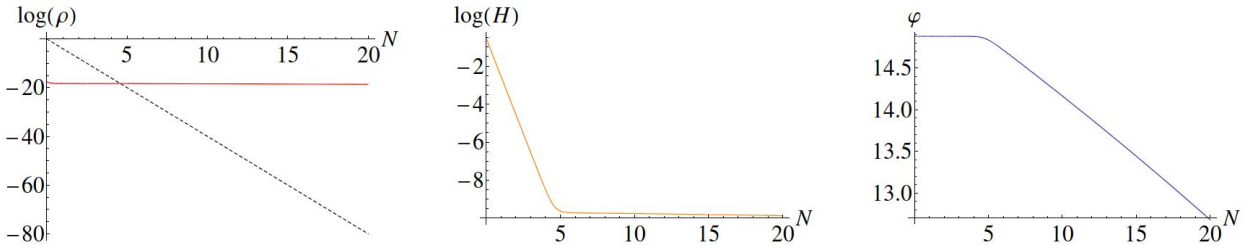


Figure 15: *Left:* Logarithm of energy density ρ as a function of number N of e-folds of inflation. The radiation fluid is shown in black and the inflaton in red. Note that the radiation fluid is initially dominant but rapidly becomes subdominant. *Middle:* Log of the Hubble parameter H as a function of N. Once the inflaton starts to dominate at $N \sim 5$, the Hubble parameter becomes approximately constant and the expansion becomes quasi-exponential. *Right:* The background inflaton field as a function of N. Note that the evolution of the inflaton is nearly frozen during the fluid dominated era. Here we have set $m_p = 1$.

During the radiation dominated era, the inflaton field evolves very little and is effectively frozen. This can be seen in the third panel of Fig. 15. The freezing of the inflaton is easy to understand: defining $N = \ln a$, we have $d\ln\bar{\phi}/dN = \dot{\bar{\phi}}/(\bar{\phi}H)$. When $\bar{\rho}_{\rm f} \gg \bar{\rho}_{\rm inf}$, the fluid dominates the right hand side of the Friedman equation (4.2.5), giving

$$\frac{d\ln\bar{\phi}}{dN} \sim \frac{m_p\dot{\bar{\phi}}}{\bar{\phi}\sqrt{\bar{\rho}_{\rm f}}} \lesssim \frac{m_p}{\bar{\phi}}\sqrt{\frac{\bar{\rho}_{\rm inf}}{\bar{\rho}_{\rm f}}}. \tag{4.2.11}$$

Thus we have $d\ln\bar{\phi}/dN \ll 1$ whenever $\bar{\rho}_{\rm inf} \ll \bar{\rho}_{\rm f}$, as long as the inflaton field is larger than the Planck scale. One can also think of this effect in terms of an enhanced Hubble damping: the Hubble

damping term in the Klein Gordon equation (4.2.7) is much larger than it would be without the fluid, since H^2 is much larger than $\rho_{\rm inf}$. Because of this enhanced damping, the inflaton kinetic energy rapidly becomes much smaller than its potential energy. For example, in the model shown in Fig. 15, the kinetic and potential energies are initially equal, but by the time of fluid-inflaton equality the kinetic energy has decreased by more than three orders of magnitude. This then implies that the inflaton dominated phase starts out in the slow-roll regime: the value of the slow-roll parameter $\varepsilon_{\rm inf} \equiv \bar{\phi}'^2/(2m_p^2\mathcal{H}^2)$ when the inflaton begins to dominate is $\sim 10^{-3}$.

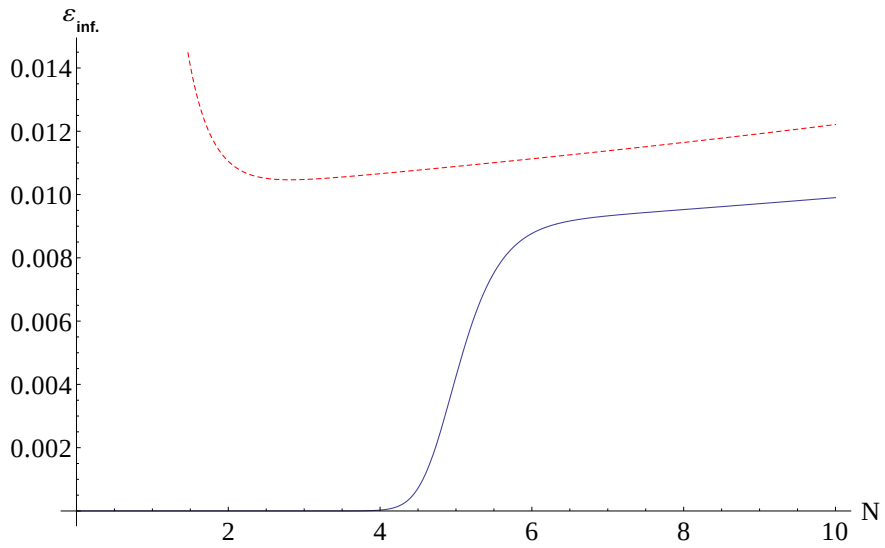
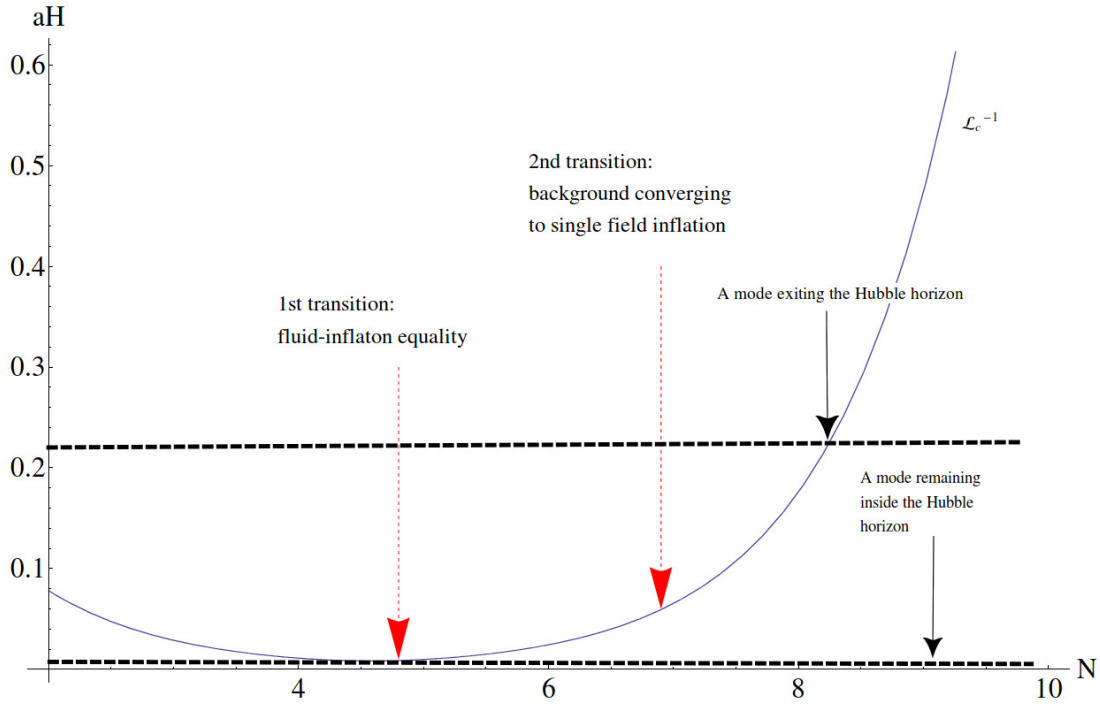


Figure 16: The inflaton slow roll parameter $\varepsilon_{\rm inf}$ as a function of N, for the fluid-inflaton model with potential $V = m^2\phi^2/2$ (blue), and for the corresponding inflaton-only mode (dashed red). In this example, the initial fluid energy density is the Planck scale, while that of the inflaton is about the GUT scale $10^{-2}m_p$. The slow-roll parameter in the inflation-only model starts at some value set by the initial conditions, and rapidly decays to the asymptotic track determined by the potential. By contrast, in the fluid-inflation model, the slow roll parameter is initially driven to a very small value during the fluid domination regime, and then relaxes upwards to the asymptotic track determined by the potential in the inflaton dominated regime.

As discussed in the overview, there are two important transitions that occur in the fluid-inflaton model. The first is the transition when $\bar{\rho}_{\rm inf} \sim \bar{\rho}_{\rm f}$, where the background expansion changes from the fluid dominated power law expansion to the inflaton dominated quasi-exponential expansion.

At the first transition, the comoving horizon scale $\mathcal{L}_c \equiv \mathcal{H}^{-1}$ has a local maximum. It increases

Figure 17: The comoving scale aH versus N for a fluid with an equation of state parameter $w = 1/3$ and an inflaton with a ϕ^2 potential. The two important transitions are shown.



during the fluid dominated era, since

$$\frac{d}{da}\mathcal{L}_c \propto \bar{\rho} + 3\bar{p}, \tag{4.2.12}$$

and during this era $\bar{\rho}$ and $\bar{p}$ are dominated by the fluid which we assume obeys the strong energy condition $w > -1/3$. It decreases during the inflaton dominated era as usual. Because of the existence of a local maximum, there exists a minimum comoving wavenumber $k_{\rm min}$ such that modes with $k \geq k_{\rm min}$ leave the Hubble horizon during the inflaton dominated phase, but modes with $k \leq k_{\rm min}$ are outside of the Hubble horizon throughout the fluid and inflaton dominated phases. Using [5]

$$3m_p^2\mathcal{H}^2 = \frac{\mathcal{E}_0}{a^{1+3w}} + \Lambda_0 a^2, \tag{4.2.13}$$

minimizing with respect to a and noting that we choose $\mathcal{E}_0 = m_p^4$ at $a = 1$ gives

$$k_{\rm min} \sim m_p^{\frac{1-3w}{3+3w}} \Lambda_0^{\frac{1+3w}{6(1+w)}}. \tag{4.2.14}$$

For example, for radiation with $w = 1/3$ we have $k_{\rm min} \sim \Lambda_0^{1/4}$.

[5] We have approximated $\bar{\rho}_{\rm inf} \approx \Lambda_0$ at early times, where Λ_0 is some constant.

The comoving Hubble horizon defined above is completely determined by the background dynamics. However, in this model there are different scales analogous to the Hubble horizon that are more relevant in discussing the evolution of the perturbations. To see this, we use some of the results derived in Sec. 4.3 below. Let us for now neglect the interaction terms between the fluid and the inflaton perturbation modes in Eq. (4.3.24). We see that each mode obeys a harmonic oscillator equation with the effective masses $m^2_{\rm inf}(k,\tau)$ and $m^2_{\rm f}(k,\tau)$ given by the diagonal elements of the matrix $\mathbf{\Omega}$ defined in Eq. (4.4.2). One can see that the perturbations evolve differently in two distinct regimes. The first regime is when $m^2_{\rm inf} \sim m^2_{\rm f} \sim k^2$, and the second regime is when k^2 is much smaller than $|m^2_{\rm inf} - k^2|$ and $|m^2_{\rm f} - c_s^2 k^2|$. The transition occurs roughly when

$$k^2 \sim |m^2_{\rm inf} - k^2| \text{ or } |m^2_{\rm f}/c_s^2 - k^2|. \tag{4.2.15}$$

Based on this transition property, we define

$$\mathcal{L}_{c-\rm inf} \equiv \frac{1}{\sqrt{|k^2 - \Omega_{11}|}}, \qquad \text{and} \qquad \mathcal{L}_{c-\rm f} \equiv \frac{1}{\sqrt{|k^2 - \Omega_{22}/c_s^2|}} \tag{4.2.16}$$

in analogy with the comoving Hubble horizon.

Nevertheless, we will indicate in Sec. 4.4.I that the scalar modes with wavenumbers larger than $k_{\rm min}$ are the ones that are of primary interest to us. These scalar perturbation modes are initially in an adiabatic regime. Furthermore, it can be shown that after the first transition, while the background geometry is nearly de Sitter, the above defined adiabatic horizon scales are within a factor unity of the comoving Hubble horizon. Thus any time we refer to a mode exiting the horizon, we shall simply mean its wavenumber k becoming equal to $\mathcal{H}$. See Fig. 18 for a plot of all three horizon scales.

We now turn to a discussion of the second important transition in the model. This transition occurs when the breaking of the time translation symmetry (characterizing the deviation of the spacetime from the de Sitter spacetime) switches from being predominantly due to the fluid to being predominantly due to the inflaton. Recall that

$$m_p^2 \dot{H} = \frac{m_p^2}{a}\Big(\frac{\mathcal{H}}{a}\Big)' = -\frac{1}{2}\Big[(1+w)\frac{\mathcal{E}_0}{a^{3(1+w)}} + \frac{\bar{\phi}'^2}{a^2}\Big], \tag{4.2.17}$$

where H is the Hubble parameter and dot denotes differentiation with respect to the coordinate time variable t. Before the transition, the first term in Eq. (4.2.17) dominates, and afterwards the

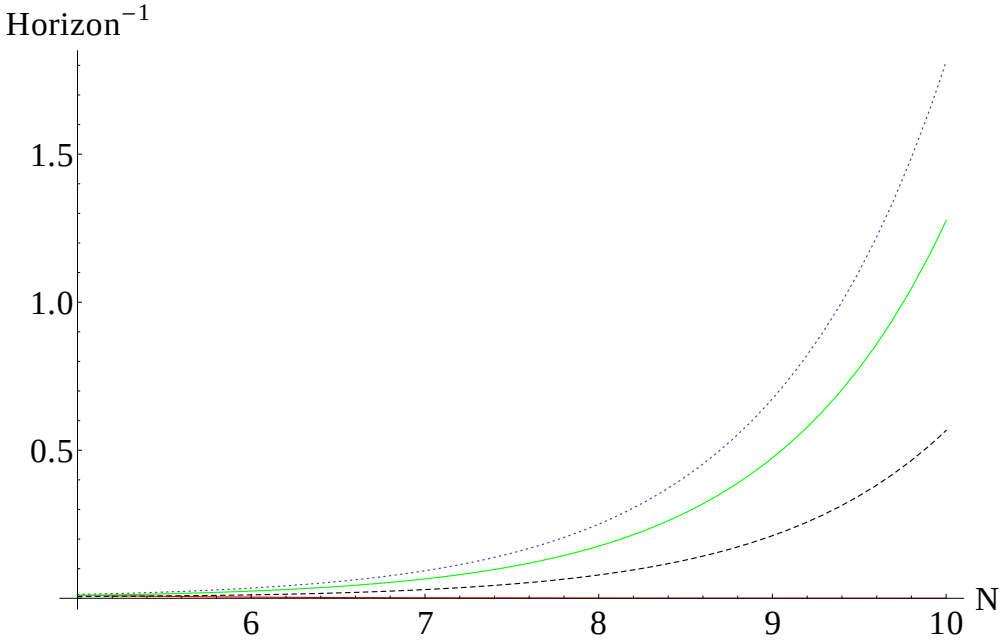


Figure 18: The inverses of the three horizon scales $\mathcal{L}_c$, $\mathcal{L}_{c-\rm inf}$ and $\mathcal{L}_{c-\rm f}$ versus N. The green plot is the plot of $\mathcal{L}_c^{-1} = \mathcal{H}$, which is a background function used to define a local-in-time horizon. The dotted blue plot is the plot of the inverse of the inflaton adiabatic horizon scale $\mathcal{L}_{c-\rm inf}^{-1}$, and the dashed black plot is the plot of the inverse of the fluid adiabatic horizon scale $\mathcal{L}_{c-\rm f}^{-1}$ associated with a fluid with an equation of state parameter $w = 0$. The adiabatic horizon scales roughly signify the boundary between the two distinct regimes for the perturbations; namely the highly oscillatory regime and usually the non-oscillatory regime. We have also included a plot of the inverse of the fluid adiabatic horizon scale for the case of radiation $w = 1/3$ (the red plot vaguely seen at the bottom). Note that the radiation case is an exception.

second term dominates . One expects the background dynamics to begin to converge to that of single field inflation once the time translation symmetry is broken by the inflaton kinetic energy. We can also quantify this transition in terms of the fluid and the inflaton contributions $\varepsilon_{\rm f}$ and $\varepsilon_{\rm inf}$ to the Hubble slow roll parameter $\epsilon \equiv -\dot{H}/H^2$, given by

$$\varepsilon_{\rm f} \sim \frac{\bar{\rho}_{\rm f}}{m_p^2 H^2}, \qquad \varepsilon_{\rm inf} \sim \frac{\dot{\bar{\phi}}^2}{m_p^2 H^2}. \tag{4.2.18}$$

The second transition occurs when $\varepsilon_{\rm inf} \gg \varepsilon_{\rm f}$. More precisely, we note that slow-roll inflation requires the smallness of the other Hubble slow-roll parameter $\eta \equiv -\ddot{H}/(2\dot{H}H)$ [6]. This parameter ensures the smallness of the inflaton field acceleration. Like ϵ, this parameter also has contributions due to

[6]Here we use this definition in lieu of the one given in Eq. (2.2.5). This definition coincides with $-\ddot{\bar{\phi}}/(\dot{\bar{\phi}}H)$ in absence of fluid.

the fluid, $\eta_{\rm f}$, as well as the inflaton, $\eta_{\rm inf}$, that are given by

$$\eta_{\rm f} \sim \frac{\bar{\rho}_{\rm f}}{\epsilon m_p^2 H^2}, \qquad \eta_{\rm inf} \sim \left| \frac{\ddot{\bar{\phi}}\dot{\bar{\phi}}}{\epsilon m_p^2 H^3} \right|. \tag{4.2.19}$$

We expect to recover a single field slow-roll background dynamics by the time $\epsilon \approx \varepsilon_{\rm inf}$ and $\eta \approx \eta_{\rm inf}$. Thus, we take these two conditions to be the defining conditions for the second transition. These conditions together place an upper bound on the energy density of the fluid at the moment of the second transition, that is

$$\bar{\rho}_{\rm f} \lesssim \varepsilon_{\rm inf}\eta_{\rm inf} m_p^2 H^2. \tag{4.2.20}$$

In what follows we shall take $\bar{\rho}_{\rm f} = \varepsilon_{\rm inf}\eta_{\rm inf} m_p^2 H^2$ as the defining moment for the second transition. The fluid energy density falls below the upper bound of Eq. (4.2.20) at an e-fold around [7]

$$\log\left[\left(\frac{m_p^4}{\varepsilon_{\rm inf}\eta_{\rm inf}\Lambda_1}\right)^{\frac{1}{3(1+w)}}\right], \tag{4.2.21}$$

where Λ_1 is the approximate value of the inflaton energy density during inflation. For a fluid dominated phase with $w = 1/3$ and an $\varepsilon_{\rm inf} = |\eta_{\rm inf}| = 0.01$, we find that after $N \approx 6.9$ e-folds the background dynamics is practically identical to that of a single field slow-roll inflation [8].

III Analytic solutions to the background equations in asymptotic limits

The background Eqs. (4.2.5) to (4.2.7) can be solved analytically in the limits when either one of the energy densities is sufficiently subdominant to the other one. The strong Hubble damping imposed on the motion of the inflaton field up until the second transition allows one to approximate the energy density of the inflaton as a cosmological constant to leading order. This means that in either regime, we expect to approximate the inflaton potential using the first few terms in its Taylor series expansion

$$V(\phi) = \Lambda_* + V'(\phi_*)(\phi - \phi_*) + \mathcal{O}\big[(\phi - \phi_*)^2\big], \tag{4.2.22}$$

where ϕ_* is the value of the inflaton field either at the beginning of the fluid dominated era, or at the onset of the second transition, when the inflaton energy density becomes dominant. This property will be particularly useful when we discuss the decoupling and adiabaticity of the perturbation modes at early times in Sec. 4.4.

[7] Assuming $\mathcal{E}_0 = m_p^4$.

[8] We are assuming GUT-scale inflation.

Based on this insight, we regard the background cosmology at early times to be the fluid plus a cosmological constant and we study the perturbations around this model induced by the variations in the inflaton field. Examining Eqs. (4.2.5) and (4.2.7) in this limit reveals that we can solve the background equations perturbatively using the following expansions for the scale factor and the inflaton field

$$a = \hat{a}\big[1 + \sigma\delta a^{(1)} + \sigma^2\delta a^{(2)} + \mathcal{O}(\sigma^3)\big], \qquad \bar{\phi} = \bar{\phi}_0\big[1 + \sigma\delta\bar{\phi}^{(1)} + \sigma^2\delta\bar{\phi}^{(2)} + \mathcal{O}(\sigma^3)\big], \tag{4.2.23}$$

where $\bar{\phi}_0$ is the initial value of the inflaton field, $\hat{a}$ is the scale factor due to the dominant fluid, and σ is the perturbation variable taken to be the dimensionless ratio of the energy density of the inflaton over the fluid when the fluid is Planckian. It turns out that the background equations are easily solved using this perturbation scheme. What makes these equations particularly simple to solve is the fact that to each order in σ, the Eqs. (4.2.5) and (4.2.7) decouple. In other words, the σ^{n-1} order term in a is used to solve for the σ^n order term in $\bar{\phi}$. The reason is simple to see; as we commented earlier, this perturbation method is really a perturbation around the fluid plus cosmological constant model generated by the variations in the background inflaton field. Indeed the first order correction to the scale factor, $\delta a^{(1)}$, is due to the zeroth order inflaton field acting as a cosmological constant with $\bar{\rho}_{\rm inf} = V(\bar{\phi}_0) = \Lambda_0$. For fluids with $w > -1/3$, one finds that $\delta a^{(1)}$ decays like $\tau^{\frac{8+6w}{1+3w}}$ as τ (or $\hat{a}$) approaches zero.

Using the zeroth order scale factor $\hat{a}$, one then finds that there exists a solution for which $\delta\bar{\phi}^{(1)}$ decays like $\tau^{\frac{6+6w}{1+3w}}$. This confirms that the expansion (4.2.22) is valid at early times. Other solutions for $\delta\bar{\phi}^{(1)}$ diverge in the limit of τ going to zero, spoiling the fluid domination at early times. We therefore specialize to this class of solutions for the inflaton field.

To provide a concrete example, we apply the above formalism to the case of a radiation fluid ($w = 1/3$) coupled to an inflaton with $V = m^2\phi^2/2$. In this case, we find

$$\begin{aligned} a(\tau) &= \sqrt{\frac{\mathcal{E}_0}{3}}\frac{\tau}{m_p}\left[1 + \frac{\bar{\phi}_0^2 m^2\mathcal{E}_0}{180 m_p^4}\tau^4 + \frac{\bar{\phi}_0^2(5\bar{\phi}_0^2/m_p^2 - 12)m^4\mathcal{E}_0^2}{194400 m_p^6}\tau^8 + ...\right], \\ \bar{\phi}(\tau) &= \bar{\phi}_0\left[1 - \frac{1}{60}\frac{m^2}{m_p^2}\mathcal{E}_0\tau^4 - \frac{(2\bar{\phi}_0^2/m_p^2 - 15)m^4\mathcal{E}_0^2}{194400 m_p^4}\tau^8 + ...\right]. \end{aligned} \tag{4.2.24}$$

Next, we analyze the background cosmology after the second transition, when the background dynamics begins to converge to single field slow-roll inflation. As a result of the initially fluid-induced Hubble damping, the inflaton field behaves as a cosmological constant to a very good approximation

for a while even after the second transition. Thus, we can employ a similar perturbation scheme to study the perturbations to this cosmological constant model induced by both the inflaton field variations and the remaining fluid energy density. To do this, consider the following expansions for the scale factor and the inflaton field,

$$a = \bar{a}\big[1+\bar{\varepsilon}_{\rm inf}\Delta a^{(1)}+\bar{\varepsilon}^2_{\rm inf}\Delta a^{(2)}+\mathcal{O}(\bar{\varepsilon}^3_{\rm inf})\big], \qquad \bar{\phi} = \bar{\phi}_1\big[1+\bar{\varepsilon}_{\rm inf}\Delta\bar{\phi}^{(1)}+\bar{\varepsilon}^2_{\rm inf}\Delta\bar{\phi}^{(2)}+\mathcal{O}(\bar{\varepsilon}^3_{\rm inf})\big], \quad (4.2.25)$$

where $\bar{\phi}_1$ is the value of the inflaton field at the second transition, $\bar{a}$ is the scale factor due to the dominant cosmological constant, and the perturbation parameter $\bar{\varepsilon}_{\rm inf}$ is the inflaton slow-roll parameter introduced in Eq. (4.2.18) that is evaluated at the second transition. As mentioned in the last subsection, this parameter breaks the time-translation symmetry during the ensuing slow-roll inflation.

To proceed, we have to determine at what orders in $\bar{\varepsilon}_{\rm inf}$ the added fluid will contribute to the perturbative expansions in Eq. (4.2.25). This can be understood using the upper bound derived for $\bar{\rho}_{\rm f}$ in Eq. (4.2.20). There we found that once $\bar{\rho}_{\rm f} \sim \bar{\varepsilon}^2_{\rm inf}\Lambda_1$, the background dynamics begins to converge to that of single field inflation [9]. This regime is precisely where our perturbation scheme is valid. It can then be seen from Eqs. (4.2.5) and (4.2.7) that the leading fluid corrections to the scale factor and the scalar field are at most of order $\bar{\varepsilon}^2_{\rm inf}$.

It actually turns out that the leading fluid correction to the scale factor is of order $\bar{\varepsilon}^2_{\rm inf}$, while its leading correction to the scalar field is of order $\bar{\varepsilon}^3_{\rm inf}$. This is a result of the same decoupling of Eqs. (4.2.5) and (4.2.7) to each order in perturbations that also occurred in the distant past regime. In this case, the decoupling occurs because the zeroth order inflaton field induces variations in itself via the Klein-Gordon equation (4.2.7). Therefore, it is easy to see that the $\bar{\varepsilon}^{n-1}_{\rm inf}$ order term in a can be used to solve for the $\bar{\varepsilon}^n_{\rm inf}$ order term in $\bar{\phi}$.

For the case of a radiation fluid ($w = 1/3$) coupled to an inflaton with $V = m^2\phi^2/2$ a straightforward calculation gives

$$\begin{aligned} a(\tau) &= m_p\sqrt{\frac{3}{\Lambda_1}}\Delta\tau^{-1}\left[1-\frac{27-21\hat{H}_{\rm inf}\Delta\tau-7\hat{H}^4_{\rm inf}\Delta\tau^4+\hat{H}^7_{\rm inf}\Delta\tau^7+42\hat{H}_{\rm inf}\Delta\tau\log|\hat{H}_{\rm inf}\Delta\tau|}{42\hat{H}_{\rm inf}\Delta\tau}\bar{\varepsilon}_{\rm inf}+...\right], \\ \bar{\phi}(\tau) &= \bar{\phi}_1\left[1+\frac{2-2\hat{H}^3_{\rm inf}\Delta\tau^3+6\log|\hat{H}_{\rm inf}\Delta\tau|}{6}\bar{\varepsilon}_{\rm inf}+...\right]. \end{aligned} \quad (4.2.26)$$

Here $\Lambda_1 = m^2\bar{\phi}^2_1/2$, $\hat{H}_{\rm inf} \equiv \sqrt{\Lambda_1/(3m_p^2)}$ is the Hubble scale at the onset of the second transition

[9] We have taken $\eta_{\rm inf} = \varepsilon_{\rm inf}$ for simplicity.

phase, and $\Delta\tau \equiv \tau - \tau_*$ with τ_* being a conformal time prior to which inflation ends. Also, in the above calculation we have set $\Delta\tau$ equal to $\hat{H}_{\rm inf}^{-1}$ at the onset of second transition.

We will use the equations derived here in Sec. 4.4 to show that the radiation and inflaton scalar perturbations will dynamically decouple in the distant past (radiation dominated) and the future (inflaton dominated) regimes.

4.3 Dynamics of the perturbations

I Fluid action

In this section we derive an action principle for the coupled fluid-inflaton-gravity perturbations. We start by describing in more detail our action principle for the fluid.

We describe the fluid flow by a set of comoving coordinates $x^a(\alpha^i, \lambda)$. Here α^i are a set of Lagrange coordinates chosen on an arbitrary spacelike hypersurface and λ is an affine parameter, which together provide sufficient data to label a fluid flow line passing through a given point x^a in spacetime. The flow of a perfect fluid is constrained by the conservation of the number density current along all flow lines,

$$\nabla_a J^a = 0. \tag{4.3.1}$$

The above J^a is the number density current of the fluid defined as nu^a, with u^a being the fluid four-velocity and n the fluid number density. One then defines the fluid four-velocity by $dx^a/d\lambda\ (-dx^a/d\lambda\ dx_a/d\lambda)^{-1/2}$, thereby arriving at [40]

$$J^a = \frac{\mathcal{F}(\alpha^i)}{\sqrt{-g}\mathcal{J}} \frac{dx^a}{d\lambda}, \tag{4.3.2}$$

where $\mathcal{F}(\alpha^i)$ is some function of the Lagrange coordinates and $\mathcal{J} \equiv \mathcal{D}(x^a)/\mathcal{D}(\alpha^i, \lambda)$ is the Jacobian associated with the coordinate transformation from the Lagrange coordinates and λ to the comoving coordinates. It can be easily seen that $\vec{J}$ defined in Eq. (4.3.2) solves the continuity equation (4.3.1).

With this description of the fluid flow, fluid Lagrangian density is given by [10]

$$\mathcal{L}_{\rm fluid}(|\vec{J}|) = -\sqrt{-g}\ \rho_{\rm f}\big(|\vec{J}|\big), \tag{4.3.3}$$

where $\rho_{\rm f}$ is the energy density of the fluid. Note that the fluid pressure $p_{\rm f}$ can be defined in terms of the fluid energy density $\rho_{\rm f}$ via [120]

$$p_{\rm f} \equiv n\frac{\partial \rho_{\rm f}}{\partial n} - \rho_{\rm f}. \tag{4.3.4}$$

Alternatively, a fluid with an equation of state parameter w has its energy density and pressure related by $p_{\rm f} = w\rho_{\rm f}$.

II Scalar perturbations

In this section we derive the linear equations of motion for the gauge invariant variables associated with the scalar perturbations to the background fluid-inflaton model introduced in Sec. 4.2. It is convenient to derive these equations by expanding the action (4.2.1) to second order in the scalar perturbations. Working with the action has privileges over the perturbed linearized Einstein equations. The clarity of the process of integrating out the non-dynamical components of the metric as well as the quantization of the perturbations are the main advantages of working with the action principle.

We insert the linearly perturbed scalar field and the flat FRW metric given in Eqs. (2.3.1) and (2.3.2) into the action (4.2.1) and expand to second order in the scalar perturbations. The first order terms must vanish identically by virtue of the background equations of motion. The second

[10] One formulation of the Lagrangian density of a perfect isentropic fluid is given by

$$-\sqrt{-g}\rho_{\rm f}\big(|\vec{J}|\big) - \delta\nabla_a J^a,$$

where δ is a Lagrange multiplier. If we use the Lagrangian fluid variables explained above, then $\nabla_a J^a = 0$ by virtue of Eq. (4.3.2). In this case, the fluid Lagrangian density reduces to $-\sqrt{-g}\rho_{\rm f}\big(|\vec{J}|\big)$. We refer the reader to Sec. 5 of [120] for an extensive discussion of the Lagrangian fluid variables. The reader might also find various common formulations of the fluid action principle discussed on pages 31-36 of [120] insightful.

order gravitational and inflaton terms are

$$\begin{aligned}
&\delta_2 S_{\rm EH}\\
&= \int d\tau \int d^3x\, \frac{a^2 m_p^2}{2}\Big\{ -6\Psi'^2 - 12\mathcal{H}(\Phi+\Psi)\Psi' - 9\mathcal{H}^2(\Phi+\Psi)^2 - 2\partial\Psi.(2\partial\Phi - \partial\Psi)\\
&-4\mathcal{H}(\Phi+\Psi)\partial^2(B-E') + 4\mathcal{H}\Psi'\partial^2 E - 4\Psi'\partial^2(B-E') - 4\mathcal{H}\partial\Psi.\partial B + 6\mathcal{H}^2(\Phi+\Psi)\partial^2 E\\
&-4\mathcal{H}\partial^2 E\partial^2(B-E') + 4\mathcal{H}\partial^2 E\partial^2 B + 3\mathcal{H}^2(\partial^2 E)^2 + 3\mathcal{H}^2(\partial B)^2\Big\},
\end{aligned} \tag{4.3.5}$$

and

$$\begin{aligned}
&\delta_2 S_{\rm inf}\\
&= \int d\tau \int d^3x\, \frac{a^2}{2}\Big\{\big(a^2 V - \frac{\bar\phi'^2}{2}\big)\big(\Phi^2 - 3\Psi^2 - (\partial B)^2 + (\partial^2 E)^2 + 2\Psi\partial^2 E + 6\Phi\Psi\\
&-2\Phi\partial^2 E\big) - 2\big(\Phi - 3\Psi + \partial^2 E\big)\big(a^2 V_{,\phi}\delta\phi + \bar\phi'[\Phi\bar\phi' - \delta\phi']\big) - \big[(\partial\delta\phi)^2 + 2\bar\phi'\partial B.\partial\delta\phi\\
&+a^2 V_{,\phi\phi}\delta\phi^2 - 4\Phi^2\bar\phi'^2 + \bar\phi'^2(\partial B)^2 + 4\Phi\bar\phi'\delta\phi' - (\delta\phi')^2\big]\Big\},
\end{aligned} \tag{4.3.6}$$

where $\partial^2 \equiv \delta^{ij}\partial_i\partial_j$ and $\partial A\ .\partial B \equiv \delta^{ij}A_{,i}B_{,j}$, and we used the conformal time τ instead of the coordinate time t.

Perturbing the fluid Lagrangian is a more delicate task. Note that the energy density of the fluid $\rho_{\rm f}$ itself is not a dynamical variable. For a perfect isentropic fluid, the energy density is a function of the fluid number density n. One needs to perturb the fluid Lagrangian by perturbing n, and use (4.3.1) to relate the perturbations in n to the metric perturbations and the perturbations in the fluid four-velocity, the latter being related to the perturbations of the comoving frame of the fluid. Following [40], we denote the perturbations to the comoving frame of the fluid by a shift vector $\xi^i(\tau, x^j)$. Since we are interested in scalar perturbations here, we define $\xi_i \equiv \partial_i\xi$ for some function ξ. In short, we are writing the number density perturbations, $\delta n(\eta, x^i)$, in terms of $\xi(\eta, x^i)$ and other metric perturbation functions. The details of this calculation can be found in [40]. The final result is

$$\begin{aligned}
&\delta_2 S_{\rm fluid}\\
&= \int d\tau \int d^3x\ a^4\Big\{\frac{1}{2}\bar\rho_{\rm f}\Phi^2 + \bar p_{\rm f}\Big(\frac{3}{2}\Psi^2 - 3\Phi\Psi + \Phi\partial^2 E - \Psi\partial^2 E + \frac{1}{2}(\partial^2 E)^2 - E_{,ij}E_{,ij}\\
&+\frac{1}{2}(\partial B)^2\Big) + (\bar\rho_{\rm f} + \bar p_{\rm f})\Big(\frac{1}{2}\xi'_{,i}\xi'_{,i} + B_{,i}\xi'_{,i} + \Phi\partial^2\xi\Big) - \frac{1}{2}c_s^2(\bar\rho_{\rm f} + \bar p_{\rm f})\big(3\Psi - \partial^2 E - \partial^2\xi\big)^2\Big\},
\end{aligned} \tag{4.3.7}$$

where c_s is the sound speed of the fluid defined by $c_s^2 = \bar{p}_{\rm f}'/\bar{\rho}_{\rm f}'$. Adding the contributions (4.3.5) to (4.3.7) and using the background equations of motion (4.2.5) and (4.2.6) together with integrations by parts in time and space, the total variation in the action simplifies to

$$\delta_2 S = \int d\tau \int d^3x \, \frac{a^2}{2}\Big\{ m_p^2\Big[-6\big(\Psi'^2 + 2\mathcal{H}\Phi\Psi' + \mathcal{H}^2\Phi^2\big) - 4\big(\Psi' + \mathcal{H}\Phi\big)\partial^2(B - E') \\ -4\partial\Psi.\partial\Phi + 2(\partial\Psi)^2\Big] + a^2(\bar{\rho}_{\rm f} + \bar{p}_{\rm f})(\xi_{,i}' + B_{,i})^2 + a^2(\bar{\rho}_{\rm f} + \bar{p}_{\rm f})(2\Phi\partial^2 E - 6\Phi\Psi + 2\Phi\partial^2\xi) \\ -c_s^2 a^2(\bar{\rho}_{\rm f} + \bar{p}_{\rm f})(3\Psi - \partial^2 E - \partial^2\xi)^2 + \bar{\phi}'^2\Phi^2 + \frac{2}{a^2}(\Phi - 3\Psi + \partial^2 E)(\bar{\phi}'\delta\phi a^2)' - (\partial\delta\phi)^2 \\ -2\bar{\phi}'\partial B.\partial\delta\phi - a^2 V_{,\phi\phi}\delta\phi^2 - 4\Phi\bar{\phi}'\delta\phi' + (\delta\phi')^2\Big\}. \tag{4.3.8}$$

In Eq. (4.3.8) six perturbation functions appear; four associated with the metric, one associated with the scalar field, and one associated with the fluid. It is convenient to replace the fluid perturbation function ξ with a three-velocity potential defined as [40]

$$\theta \equiv \frac{2\sqrt{\beta}a}{c_s}(\xi' + B) \tag{4.3.9}$$

which is related to the perturbation of the four-velocity δu^i of the fluid by

$$\theta_{,i} = -\frac{2\sqrt{\beta}a^2}{c_s}\delta u_i. \tag{4.3.10}$$

In Eq. (4.3.9) β is defined as $(a^2/2)(\bar{\rho}_{\rm f} + \bar{p}_{\rm f})$. Using the definition (4.3.9) and the constraint

$$\big(\frac{ac_s}{2}\sqrt{\beta}\theta\big)' = -a^2 c_s^2\beta\Big(3\Psi - \partial^2 E - \partial^2\xi + \frac{\Phi}{c_s^2}\Big) \tag{4.3.11}$$

obtained from varying the action (4.3.8) with respect to ξ, we eliminate all ξ' and $\partial^2\xi$ appearing in Eq. (4.3.8) in favor of θ and the other metric perturbation functions. The final result is

$$\delta_2 S = \int d\tau \int d^3x \, \frac{1}{2}\Big\{ -6m_p^2 a^2(\Psi' + \mathcal{H}\Phi)^2 + 2\frac{\beta a^2}{c_s^2}\Phi^2 - 2a^2 m_p^2\partial\Psi.(2\partial\Phi - \partial\Psi) \\ +\frac{1}{2}\frac{(c_s a\sqrt{\beta}\theta)'^2}{a^2\beta c_s^2} - \frac{1}{2}c_s^2(\partial\theta)^2 - 6c_s a\sqrt{\beta}\theta\Psi' - 2a\sqrt{\beta}c_s\theta\Big(\frac{\Phi}{c_s^2}\Big)' - 4a^2 m_p^2\partial^2(B - E') \\ \Big(\Psi' + \mathcal{H}\Phi - \frac{1}{2m_p^2}\bar{\phi}'\delta\phi + \frac{c_s}{2am_p^2}\sqrt{\beta}\theta\Big) + a^2\bar{\phi}'^2\Phi^2 + 2(\Phi - 3\Psi)(\bar{\phi}'\delta\phi a^2)' - a^2(\partial\delta\phi)^2 \\ -a^4 V_{,\phi\phi}\delta\phi^2 - 4a^2\Phi\bar{\phi}'\delta\phi' + a^2(\delta\phi')^2\Big\}. \tag{4.3.12}$$

Note that the action (4.3.12) depends on the perturbation functions B and E only through the combination $B - E'$. Varying the action (4.3.12) with respect to $B - E'$ gives

$$m_p^2(\Psi' + \mathcal{H}\Phi) = \frac{1}{2}\bar{\phi}'\delta\phi - \frac{c_s}{2a}\sqrt{\beta}\theta. \tag{4.3.13}$$

As can be seen from Eq. (4.3.12), Φ is non-dynamical and it can be integrated out using Eq. (4.3.13), i.e. Φ can be eliminated in favor of Ψ, θ, and $\delta\phi$. Observe that by doing this we eliminate E, B, and Φ. This way we are left with three dynamical fields; Ψ, θ, and $\delta\phi$. The action (4.3.12) after integrating out B, E, and Φ becomes

$$\begin{aligned}
\delta_2 S = \int d\tau \int d^3x\, \frac{1}{2}\Bigg\{ & \frac{\theta^2}{m_p^2}\Big[-\frac{3}{2}c_s^2\beta + \frac{\beta^2}{2m_p^2\mathcal{H}^2} - \frac{\beta'\mathcal{H}}{2\mathcal{H}^2} + \frac{\beta\mathcal{H}'}{2\mathcal{H}^2} + \frac{m_p^2(c_s\sqrt{\beta}a)'^2}{c_s^2\beta a^2} - \frac{m_p^2(c_s\sqrt{\beta}a)''}{2c_s\sqrt{\beta}a} \\
& + ac_s\sqrt{\beta}\Big(\frac{\sqrt{\beta}}{a\mathcal{H}c_s}\Big)' + \frac{\bar{\phi}'^2}{4m_p^2\mathcal{H}^2}c_s^2\beta\Big] + \theta\Psi'\Big(\frac{2a\beta^{3/2}}{c_s m_p^2\mathcal{H}^2} - 6c_s a\sqrt{\beta} - \frac{2(c_s\sqrt{\beta}a)'}{\mathcal{H}c_s^2} + \frac{a\bar{\phi}'^2}{m_p^2\mathcal{H}^2}c_s\sqrt{\beta}\Big) \\
& + \theta\delta\phi\Big(\frac{a^3}{m_p^2\mathcal{H}}c_s\sqrt{\beta}V_{,\phi}\Big) + \theta\delta\phi'\Big(\frac{a}{m_p^2\mathcal{H}}c_s\sqrt{\beta}\bar{\phi}'\Big) + \Psi'\delta\phi'\Big(\frac{2a^2\bar{\phi}'}{\mathcal{H}}\Big) + \Psi'\delta\phi\Big(\frac{2a^4V_{,\phi}}{\mathcal{H}} + 6a^2\bar{\phi}'\Big) \\
& - \frac{2\sqrt{\beta}a}{\mathcal{H}c_s}\theta'\Psi' + \Psi'^2\Big(\frac{2a^2\beta}{c_s^2\mathcal{H}^2} + \frac{a^2\bar{\phi}'^2}{\mathcal{H}^2}\Big) + \frac{1}{2}\theta'^2 + a^2(\delta\phi')^2 + \frac{4a^2m_p^2}{\mathcal{H}}\partial\Psi.\partial\Psi' + 2m_p^2a^2(\partial\Psi)^2 \\
& + 2\frac{a}{\mathcal{H}}c_s\sqrt{\beta}\partial\Psi.\partial\theta - \frac{1}{2}c_s^2(\partial\theta)^2 - a^2(\partial\delta\phi)^2 + \delta\phi^2\Big[-a^4V_{,\phi\phi} - \frac{a^4}{m_p^2\mathcal{H}}\bar{\phi}'V_{,\phi} + \Big(\frac{a^2\bar{\phi}'^2}{2m_p^2\mathcal{H}}\Big)'\Big]\Bigg\}.
\end{aligned} \tag{4.3.14}$$

The functions $\delta\phi$, Ψ, and θ form the following two gauge-invariant variables

$$\nu = \frac{1}{\sqrt{2}}(\theta - 2z\Psi), \qquad \pi = a\delta\phi + \zeta\Psi, \tag{4.3.15}$$

where $z = a\sqrt{\beta}/(\mathcal{H}c_s)$ and $\zeta = a\bar{\phi}'/\mathcal{H}$. The function ν is the gauge invariant fluid velocity potential, and the function π is the gauge invariant inflaton perturbation.

So far in this computation we have not made any gauge specialization but kept the gauge arbitrary. One might imagine that this computation would be easier if one imposes some gauge conditions at the start, for example the Newtonian gauge conditions $B = E = 0$. However, this gauge specialization is too restrictive and does not allow us to properly integrate out the non-dynamical variables. With that gauge choice, we would not have been able to derive the constraint equation (4.3.13) needed to integrate out Φ. Similar difficulties arise in the synchronous gauge where $\Phi = B = 0$ (but not in the inflaton comoving gauge where $\delta\phi = E = 0$).

Rewriting the action (4.3.14) in terms of ν and π defined in Eq. (4.3.15) gives [11] [12]

$$\delta_2 S = \int d\tau \int d^3x \left\{ \mathcal{L}_{\rm fluid} + \mathcal{L}_{\rm inf} + \mathcal{L}_{\rm int} \right\}, \tag{4.3.16}$$

where

$$\mathcal{L}_{\rm fluid} = \frac{1}{2}\left(\nu'^2 - c_s^2 (\partial\nu)^2 + \left[\frac{z''}{z} + 4A \right] \nu^2 \right) \tag{4.3.17}$$

is the fluid Lagrangian density,

$$\mathcal{L}_{\rm inf} = \frac{1}{2}\left(\pi'^2 - (\partial\pi)^2 + \left[\frac{\zeta''}{\zeta} + \frac{2}{\zeta^2}(4z^2 A - (zB)') \right] \pi^2 \right) \tag{4.3.18}$$

is the inflaton Lagrangian density, and

$$\mathcal{L}_{\rm int} = \left(4\sqrt{2} A \frac{z}{\zeta} - \sqrt{2} B \frac{\zeta'}{\zeta^2} \right) \nu\pi + \sqrt{2} \frac{B}{\zeta} \nu\pi' \tag{4.3.19}$$

is the interaction Lagrangian density. Here

$$A \equiv \frac{\mathcal{H}'}{4m_p^2 \mathcal{H}^2} \bar{\phi}'^2 - \frac{(\bar{\phi}'^2)'}{8m_p^2 \mathcal{H}} - \frac{\bar{\phi}'^2 \beta}{8m_p^4 \mathcal{H}^2}(1 - c_s^2) + \frac{c_s'}{2c_s} \frac{\bar{\phi}'^2}{2m_p^2 \mathcal{H}} \tag{4.3.20}$$

and

$$B \equiv \left(c_s - \frac{1}{c_s} \right) \frac{a \bar{\phi}'^2 \sqrt{\beta}}{2m_p^2 \mathcal{H}^2} \tag{4.3.21}$$

are functions of the background variables. We remark here that the fluid influences the inflaton perturbations in three steps: first through the explicit interaction terms given in Eq. (4.3.19), second by modifying the expressions for the background functions A and B that involve the fluid variables, and third by modifying the background solutions which also influence the inflaton perturbations. Finally, observe that turning off the inflaton by setting $\pi = 0$ gives the standard gauge invariant Lagrangian density for the isentropic fluid [40], while turning off the fluid by setting $\nu = 0$ gives the standard gauge invariant action for the inflaton perturbations [48, 49, 50].

We can write the action (4.3.16) more compactly as

$$\delta_2 S = \int d\tau \int d^3x \, \frac{1}{2}\left[\mathbf{\Pi}'^{\rm T} \mathbf{\Pi}' - \partial_i \mathbf{\Pi}^{\rm T} \boldsymbol{c}_s^2 \partial^i \mathbf{\Pi} + \mathbf{\Pi}^{\rm T} \mathbf{\Gamma} \mathbf{\Pi} + \mathbf{\Pi}'^{\rm T} \boldsymbol{\lambda} \mathbf{\Pi} \right], \tag{4.3.22}$$

[11] The action presented in this form is not applicable for the case of pressureless dust fluid. A redefinition of variables is necessary for this special case.

[12] For scalar cosmological perturbations, isentropic fluids admit a simple scalar field description. For the case of radiation fluid, one can replace the fluid Lagrangian density $-\rho_f$ with $(\nabla\psi)^4/4$. Cosmologies of scalar fields with non-canonical kinetic terms of this form had been previously studied (see e.g. [27]).

where the matrices are defined as

$$\boldsymbol{\Pi} \equiv \begin{pmatrix} \pi \\ \nu \end{pmatrix}, \qquad \boldsymbol{c}_s^2 \equiv \begin{pmatrix} 1 & 0 \\ 0 & c_s^2 \end{pmatrix}, \qquad \boldsymbol{\lambda} \equiv \begin{pmatrix} 0 & \sqrt{2}\frac{B}{\zeta} \\ -\sqrt{2}\frac{B}{\zeta} & 0 \end{pmatrix},$$

$$\boldsymbol{\Gamma} \equiv \begin{pmatrix} \frac{\zeta''}{\zeta} + \frac{2}{\zeta^2}\left[4z^2A - (zB)'\right] & 4\sqrt{2}A\frac{z}{\zeta} - \frac{\sqrt{2}}{2}B\frac{\zeta'}{\zeta^2} - \frac{\sqrt{2}}{2}\frac{B'}{\zeta} \\ 4\sqrt{2}A\frac{z}{\zeta} - \frac{\sqrt{2}}{2}B\frac{\zeta'}{\zeta^2} - \frac{\sqrt{2}}{2}\frac{B'}{\zeta} & \frac{z''}{z} + 4A \end{pmatrix}. \tag{4.3.23}$$

We have suppressed the dependences on τ for all the background functions above. The equation of motion for this action is

$$\boldsymbol{\Pi}'' - \left[\boldsymbol{c}_s^2\nabla^2 + \boldsymbol{\Gamma} - \frac{1}{2}\boldsymbol{\lambda}'\right]\boldsymbol{\Pi} + \boldsymbol{\lambda}\boldsymbol{\Pi}' = 0, \tag{4.3.24}$$

with the corresponding equations of motion for each mode being

$$\nu'' + \left(c_s^2k^2 - \frac{z''}{z} - 4A\right)\nu - \left(4\sqrt{2}A\frac{z}{\zeta} - \sqrt{2}B\frac{\zeta'}{\zeta^2}\right)\pi - \sqrt{2}\frac{B}{\zeta}\pi' = 0, \tag{4.3.25a}$$

$$\pi'' + \left[k^2 - \frac{\zeta''}{\zeta} - \frac{2}{\zeta^2}(4z^2A - (zB)')\right]\pi + \left(\sqrt{2}\frac{B'}{\zeta} - 4\sqrt{2}A\frac{z}{\zeta}\right)\nu + \sqrt{2}\frac{B}{\zeta}\nu' = 0. \tag{4.3.25b}$$

We have omitted ν and π dependences on τ and k. This is a system of coupled oscillators with time dependent coefficients.

III Tensor perturbations

To derive the linear equations of motion for the gravitational waves produced in this model, we expand the first term of the action (4.2.1) to second order in the tensorial metric perturbations. Unlike the scalar perturbations, the metric tensorial perturbations do not couple to the matter perturbations. The final result, therefore, is identical to the one derived in Sec. 2.3.III. The equation of motion for the gravitational waves in the Fourier momentum space is then given by

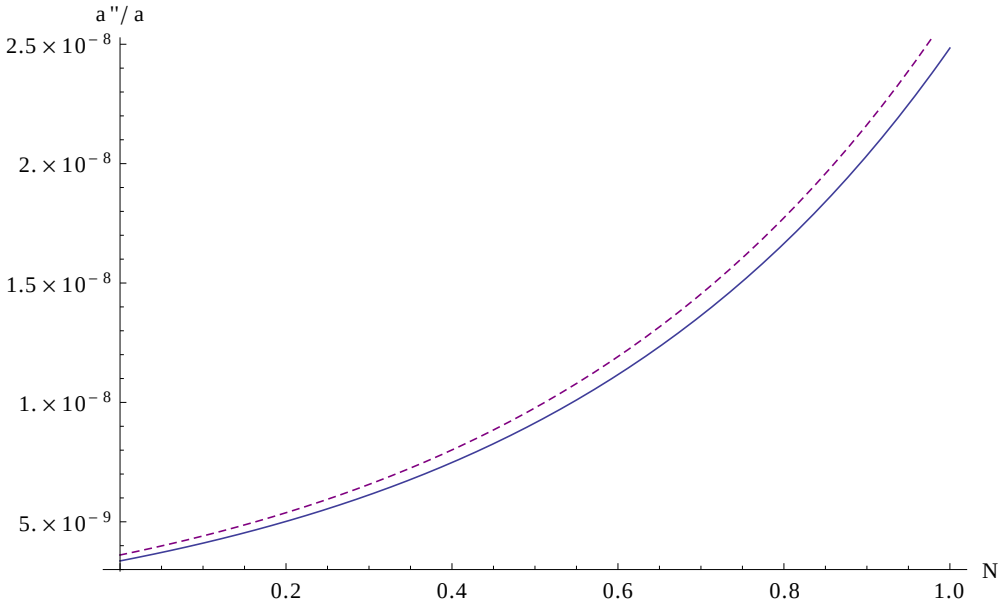
$$h^{''A} + \left[k^2 - \frac{a''}{a}\right]h^A = 0, \tag{4.3.26}$$

where h^A is the properly normalized tensor mode function for the polarization state A. This equation is shared in all models where the matter content is minimally coupled to gravity and the gravitational field action is the Einstein-Hilbert term. What renders the evolution of the gravitational waves different in various models is the difference in the background dynamics, i.e. the difference in a''/a ratio. From the two Friedman equations (4.2.5) and (4.2.6) we have

$$m_p^2\frac{a''}{a} = \frac{a^2}{6}\left(\bar{\rho} - 3\bar{p}\right) = \frac{a^2}{6}\Big[\left(\bar{\rho}_{\rm inf} - 3\bar{p}_{\rm inf}\right) + \bar{\rho}_{\rm f}\left(1 - 3w\right)\Big]. \tag{4.3.27}$$

For the case of radiation ($w = 1/3$) coupled to inflaton, the second term in the bracket vanishes and a''/a is given by the same expression as in the case of the inflaton-only model. However, the background geometry and the dynamics of the inflaton field ϕ is still influenced by the presence of the radiation fluid. We will explore these differences in Sec. 4.7.I where we numerically compute the spectrum of the primordial gravitational waves in the fluid-inflaton model.

Figure 19: Below is the plot of a''/a ($m_p = 1$) for the fluid-inflaton model (blue) and the inflaton-only model (dashed purple) with the ϕ^2 potential. The initial values in the inflaton-only model are chosen such that the background dynamics becomes identical to the fluid-inflaton model after the second transition.



4.4 Past and future decoupling of the scalar perturbations

An interesting property of the fluid and inflaton scalar perturbations in this model is that they are only effectively coupled while the background energy densities are within a few orders of magnitude of one another. To establish this, we must examine the equations of motion (4.3.24) both at early times and at later times when the background dynamics converges to that of single field inflation. The equations of motion (4.3.24) are of the form (after Fourier transforming to the momentum space)

$$\Pi_i'' + \left[\Omega_{ij} + \frac{1}{2}\lambda_{ij}'\right]\Pi_j + \lambda_{ij}\Pi_j' = 0, \tag{4.4.1}$$

where $\mathbf{\Pi}$ and $\boldsymbol{\lambda}$ are defined in Eq. (4.3.23), and

$$\mathbf{\Omega} \equiv -\boldsymbol{c}_s^2 k^2 - \mathbf{\Gamma} = \begin{pmatrix} k^2 - \frac{\zeta''}{\zeta} - \frac{2}{\zeta^2}\left[4z^2 A - (zB)'\right] & -4\sqrt{2}A\frac{z}{\zeta} + \frac{\sqrt{2}}{2}B\frac{\zeta'}{\zeta^2} + \frac{\sqrt{2}}{2}\frac{B'}{\zeta} \\ -4\sqrt{2}A\frac{z}{\zeta} + \frac{\sqrt{2}}{2}B\frac{\zeta'}{\zeta^2} + \frac{\sqrt{2}}{2}\frac{B'}{\zeta} & c_s^2 k^2 - \frac{z''}{z} - 4A \end{pmatrix}. \tag{4.4.2}$$

As can be seen from Eqs. (4.4.1) and (4.4.2), Ω_{11} and Ω_{22} can be interpreted as the effective masses of the inflaton and fluid in this model. Turning off the interactions, they become

$$\Omega_{11} \to k^2 - \zeta''/\zeta \qquad \text{and} \quad \Omega_{22} \to c_s^2 k^2 - z''/z \tag{4.4.3}$$

as they would be in inflaton-only and fluid-only models. From Eq. (4.4.1), one expects that the perturbations decouple once

$$\lambda'_{12},\ \Omega_{12} + \frac{1}{2}\lambda'_{12} \ll\ \Omega_{11}, \Omega_{22}. \tag{4.4.4}$$

We can check this condition using the asymptotic background solutions obtained in Sec. 4.2.III. The result for the distant past limit is given in Table. 4.

	$\frac{-1}{3} < w < \frac{1}{3}$	$w = \frac{1}{3}$
Ω_{11}, Ω_{22}	$c_s^2 k^2 + (\text{constant})\ \tau^{-2} + ...$	$c_s^2 k^2 + (\text{constant})\ \tau^{2} + ...$
$\Omega_{12} + \frac{1}{2}\lambda'_{12}$	$\mathcal{O}\left(\tau^{\frac{8+6w}{1+3w}}\right)$	$\mathcal{O}\left(\tau^{\frac{8+6w}{1+3w}}\right)$
λ'_{12}	$\mathcal{O}\left(\tau^{\frac{4}{1+3w}}\right)$	$\mathcal{O}\left(\tau^{\frac{4}{1+3w}}\right)$

Table 4: Distant past forms of the background functions appearing in Eq. (4.4.1).

This calculation shows that the perturbation modes decouple in the limit $\tau \to 0$. Another question that could be asked is: does a mode of a given wavenumber k decouple before the Planck scale is reached? To address this question, one could numerically compute the dimensionless ratio

$$\frac{\Omega_{12} + \frac{1}{2}\lambda'_{12}}{\sqrt{\Omega_{11}\Omega_{22}}} \tag{4.4.5}$$

just as $\bar{\rho}_{\rm f}$ becomes Planckian, and see whether it is small or large compared to unity.

Next we would like to show that this dynamical decoupling of the scalar perturbations also occurs when the background dynamics begins to look like single field slow-roll inflation, i.e. when $\bar{\rho}_{\rm f} \lesssim \varepsilon_{\rm inf}\eta_{\rm inf} m_p^2 H^2$. This can be done using the late time solutions of the background equations (4.2.26). The result is provided in Tab. 5 for the case of radiation fluid. From this we see that

	$w = \frac{1}{3}$
$\Omega_{11} - k^2$	$-\frac{2}{\Delta\tau^2} + \mathcal{O}(\bar{\varepsilon}_{\rm inf})$
$\Omega_{22} - c_s^2 k^2$	$\big(\frac{2}{\Delta\tau^2} - 4\hat{H}_{\rm inf}^3\Delta\tau + 2\hat{H}_{\rm inf}^6\Delta\tau^4\big)\bar{\varepsilon}_{\rm inf} + \mathcal{O}(\bar{\varepsilon}_{\rm inf}^2)$
$\Omega_{12} + \frac{1}{2}\lambda'_{12}$	$\hat{H}_{\rm inf}^2\big(-\frac{1}{\sqrt{6}} + \frac{28}{\sqrt{6}}\hat{H}_{\rm inf}^3\Delta\tau^3\big)\bar{\varepsilon}_{\rm inf}^{3/2} + \mathcal{O}(\bar{\varepsilon}_{\rm inf}^2)$
λ'_{12}	$\hat{H}_{\rm inf}^2\big(\frac{2}{\sqrt{6}} - \frac{8}{\sqrt{6}}\hat{H}_{\rm inf}^3\Delta\tau^3\big)\bar{\varepsilon}_{\rm inf}^{3/2} + \mathcal{O}(\bar{\varepsilon}_{\rm inf}^2)$

Table 5: Future forms of the background functions appearing in Eq. (4.4.1).

early on during the second transition we have

$$\begin{aligned}
&\Omega_{11} - k^2 \sim \Delta\tau^{-2} \sim \hat{H}_{\rm inf}^2 \gg \Omega_{12} + \frac{1}{2}\lambda'_{12} \sim \lambda'_{12} \sim \hat{H}_{\rm inf}^2\bar{\varepsilon}_{\rm inf}^{3/2},\\
&\Omega_{22} - c_s^2 k^2 \sim \hat{H}_{\rm inf}^2\mathcal{O}(\bar{\varepsilon}_{\rm inf}^2) \lesssim \Omega_{12} + \frac{1}{2}\lambda'_{12} \sim \lambda'_{12} \sim \hat{H}_{\rm inf}^2\bar{\varepsilon}_{\rm inf}^{3/2},
\end{aligned} \tag{4.4.6}$$

where we have used $\Delta\tau = \hat{H}_{\rm inf}^{-1}$ at the second transition consistent with our choice of scale factor normalization in Sec. 4.2.III. From Eq. (4.4.6) we see that the inflaton mode has effectively decoupled from the fluid mode by the second transition. However, this does not appear to be the case for the radiation fluid mode. It is in fact simple to show analytically that

$$\Omega_{22} - c_s^2 k^2 = \frac{z''}{z} + 4A \to 0, \qquad \text{in the limit of } a \to \frac{1}{\hat{H}_{\rm inf}\Delta\tau}, \tag{4.4.7}$$

i.e., the effective mass of the radiation fluid becomes equal to its wavenumber squared in the limit of the background geometry becoming de Sitter. Thus, when considering the issue of the decoupling of the radiation fluid modes at the second transition, one must consider the full radiation fluid effective mass Ω_{22} instead. As we will see in the next subsection, we shall be primarily interested in the scalar modes with wavenumbers $k \gtrsim k_{\rm min}$. Using the value of $k_{\rm min}$ obtained in Sec. 4.2.II, we have

$$\Omega_{22} \gtrsim k_{\rm min}^2 \approx m_p^{-2}(\mathcal{E}_0\Lambda_0)^{1/2} \sim \hat{H}_{\rm inf}^2\bar{\varepsilon}_{\rm inf} \gg \Omega_{12} + \frac{1}{2}\lambda'_{12} \sim \lambda'_{12} \sim \hat{H}_{\rm inf}^2\bar{\varepsilon}_{\rm inf}^{3/2}, \tag{4.4.8}$$

where we used $k_{\rm min} \sim m_p^{-1}(\mathcal{E}_0\Lambda_0)^{1/4}$ [13] and $\mathcal{E}_0 \sim \bar{\varepsilon}_{\rm inf}^2\Lambda_0$ consistent with our choice of $\Delta\tau = \hat{H}_{\rm inf}^{-1}$ at the second transition (also note that we are assuming $\Lambda_0 \approx \Lambda_1$ for simplicity). Therefore, it appears that the radiation and inflaton perturbation modes of primary interest will eventually decouple again early on during the second transition.

[13] Here we waived the assumption that $\mathcal{E}_0 = m_p^4$ because we are taking $a = 1$ at the onset of the second transition.

I Adiabaticity of the scalar perturbations ; $w = \frac{1}{3}$

The analysis done in Sec. 4.4 indicates that the fluid and inflaton modes dynamically decouple at early times, reducing to two uncoupled harmonic oscillators with time dependent frequencies. We now show that if the fluid is radiation ($w = 1/3$), both the fluid and inflaton modes are in an adiabatic regime at early times, that is the effective frequencies $\sqrt{\Omega_{ii}}$ of those modes evolve in timescales much longer than the frequencies themselves. This concept of adiabaticity replaces and generalizes the usual concept of modes being "inside the horizon". Indeed for both the fluid and inflaton modes, the effect of the background geometry on subhorizon perturbations diminishes as τ^2 with $\tau \to 0$. This can be seen in Table. 4, where we showed

$$\Omega_{22} \equiv c_s^2 k^2 - \frac{z''}{z} - 4A = c_s^2 k^2 + \mathcal{O}(\tau^2), \qquad \Omega_{11} \equiv k^2 - \frac{\zeta''}{\zeta} - \frac{2}{\zeta^2}\left[4z^2 A - (zB)'\right] = k^2 + \mathcal{O}(\tau^2). \quad (4.4.9)$$

This indicates

$$\left|\frac{\Omega_{22}'}{\Omega_{22}^{3/2}}\right|, \left|\frac{\Omega_{11}'}{\Omega_{11}^{3/2}}\right| \ll 1 \quad (4.4.10)$$

as $\tau \to 0$. Therefore in this limit the scalar modes behave like simple, time-independent harmonic oscillator for which the general state furnishes a natural choice of initial state, a generalization of the usual Bunch-Davies vacuum.

Although all modes become adiabatic sufficiently early in time, not all modes will have become adiabatic by the time the Planck scale is reached, which is the earliest time our model can be trusted. We look into this matter by computing the adiabaticity ratio (4.4.10) evaluated for $k_{\rm min}$ when the fluid becomes Planckian. Using the background perturbative solutions found in Sec. 4.2.III, we find (for $w = 1/3$)

$$\left|\frac{\Omega_{11}'}{\Omega_{11}^{3/2}}\right|_{k_{\rm min}} \approx m_p^{-3}\Lambda_0^{3/4}, \qquad \left|\frac{\Omega_{22}'}{\Omega_{22}^{3/2}}\right|_{k_{\rm min}} \approx 10\ m_p^{-1}\Lambda_0^{1/4}, \quad (4.4.11)$$

where $\Lambda_0 = V(\bar{\phi}_0)$. Note that here we have specialized to the class of solutions for the inflaton field for which the field freezes at early times. Since the inflaton field is expected to be at the GUT scale during inflation, we find

$$\left|\frac{\Omega_{11}'}{\Omega_{11}^{3/2}}\right|_{k} \gtrsim 10^{-6}, \qquad \left|\frac{\Omega_{22}'}{\Omega_{22}^{3/2}}\right|_{k} \gtrsim 0.1, \quad \text{for all } \ k \lesssim k_{\rm min}. \quad (4.4.12)$$

Evidently, the fluid modes with wavenumbers smaller than $k_{\rm min}$ will not enjoy a strong adiabaticity at the Planck scale. We will therefore be interested in having the observable scales correspond to wavenumbers larger than $k_{\rm min}$.

II Adiabaticity of the tensor perturbations ; $w = \frac{1}{3}$

Just as in the scalar perturbation case, the tensor modes are in an adiabatic regime at early times. For tensor perturbations to be in an adiabatic regime at early times, the frequency of the gravitational waves ω must satisfy

$$\left|\frac{\omega'}{\omega^2}\right| \ll 1. \tag{4.4.13}$$

Using the background solution for the scale factor found in Sec. 4.2.III one finds

$$\omega^2(k) = k^2 - \frac{a''}{a} = k^2 + \mathcal{O}(\tau^2) \tag{4.4.14}$$

indicating that $\omega^2(k) \to k^2$ for all tensor modes as $\tau \to 0$. However, we must see for what modes the conditions given in Eq. (4.4.13) are met by the time the energy density of the fluid becomes Planckian. A short calculation similar to the one done for the scalar modes shows

$$\frac{\omega'(k_{\rm min})}{\omega^2(k_{\rm min})} \approx m_p^{-1}\Lambda_0^{1/4}, \tag{4.4.15}$$

where the expression is calculated when the fluid is at the Planck scale. Therefore for all gravitational waves with wavenumbers $k \gtrsim k_{\rm min}$, there exists a natural choice of ground state at the initial time when the fluid energy density is Planckian.

III Non-adiabaticity of perturbations ; $w \neq \frac{1}{3}$

In the case of a fluid with an equation of state parameter $w \neq 1/3$, a short calculation using the results of Sec. 4.2.III gives

$$\Omega_{22} = c_s^2 k^2 - \frac{2-6w}{(1+3w)^2}\tau^{-2} + \mathcal{O}(\tau^2), \qquad \Omega_{11} = k^2 - \frac{2-6w}{(1+3w)^2}\tau^{-2} + \mathcal{O}(\tau^2),$$
$$\omega^2 = k^2 - \frac{2-6w}{(1+3w)^2}\tau^{-2} + \mathcal{O}(\tau^2). \tag{4.4.16}$$

Here one cannot achieve initial adiabaticity in the sense of (4.4.10) for all modes of interest, particularly for modes near $k_{\rm min}$ which are of primary interest in this paper.

The non-adiabaticity for a mode is an indication of its interaction with the background geometry. In such cases one no longer has a natural choice of ground state. One could take different approaches to parameterize such initial states as excited states, the core idea being to constrain such states using backreaction considerations of the energy density stored in these initial quanta. Computations

using excited initial states are more speculative due to our ignorance of the correct initial occupation numbers for the perturbation quanta. Therefore we refrain from constructing numerical examples for such scenarios in sections 4.6.I and 4.7.I.

4.5 Curvature and entropy perturbations in this model

In the cosmological theory of small fluctuations, conservation laws can exist for certain gauge invariant variables in some circumstances. The existence of conserved quantities is the primary reason why predictions of some inflationary models can be tested despite our ignorance of what went on immediately after inflation. As mentioned in Sec. 2.5.I, the comoving curvature perturbation $\mathcal{R}$ is of primary interest in this regard. This gauge invariant quantity is generally defined by

$$\mathcal{R} \equiv \Psi - \mathcal{H}\frac{\delta q}{a(\bar{\rho}+\bar{p})} = \Psi - \mathcal{H}\frac{\delta u}{a}, \tag{4.5.1}$$

where δq and δu are the three-momentum and the three-velocity potentials associated with a fluid [14]. Taking $\mathcal{R}$ to the Fourier space using $\mathcal{R}(\tau,\boldsymbol{x}) = \int\ d^3k/(2\pi)^{3/2}\ \mathcal{R}(\tau,\boldsymbol{k})e^{i\boldsymbol{k}.\boldsymbol{x}}$, one can show [11]

$$\mathcal{R}'(\tau,k) = -3\frac{\mathcal{H}\bar{p}'}{\bar{\rho}'}\mathcal{S}(\tau,k) + \mathcal{O}\left(\frac{k^2}{\mathcal{H}^2}\right), \tag{4.5.2}$$

where $\mathcal{S}(\tau,k)$ is the Fourier transform of a gauge invariant function called the entropy perturbation,

$$\mathcal{S} \equiv \mathcal{H}\Big(\frac{\delta\rho}{\bar{\rho}'} - \frac{\delta p}{\bar{p}'}\Big), \tag{4.5.3}$$

where δp and $\delta\rho$ are the pressure and energy density perturbations. This function is the gauge invariant measure of the non-adiabaticity of scalar perturbations. What we are primarily interested in is whether the comoving curvature perturbation for the fluid-inflaton model is conserved on superhorizon scales.

It is a remarkable fact that for any background cosmology, one can find solutions of the perturbation equations (4.3.24) for which $\mathcal{R}(k)$ is conserved as long as it remains in the superhorizon regime, i.e. as long as $k \ll \mathcal{H}$ [51, 52]. This fact remains valid regardless of the specifics of all subsequent cosmological eras. These perturbations, which are known as the adiabatic perturbations, can be intuitively realized as local spatial curvature fluctuations inducing a time-delay effect in the

[14] See App. B.

background geometry. All linear fluctuations in single component cosmological models are known to be of this kind. On the other hand, for more complicated cosmological models, such as multifield inflationary models, non-adiabatic solutions can exist. For these solutions, $\mathcal{S}(k)$ will generally not vanish in the superhorizon regime. Non-vanishing of $\mathcal{S}(k)$ causes the comoving curvature perturbation $\mathcal{R}(k)$ to change even in the superhorizon regime. Without a quantity that remains conserved in the superhorizon regime, predictions of an inflationary model are only testable if one chooses a specific theory of reheating.

Luckily, the entropy perturbations are unimportant in our model. We establish this fact by showing that the entropy perturbation function $\mathcal{S}$ reduces to its single field counterpart after the second transition. Using the definition (4.5.3) we have

$$\mathcal{S} = \mathcal{H}\Big[\frac{\delta\rho_{\rm f} + \delta\rho_{\rm inf}}{\bar{\rho}'} - \frac{\delta p_{\rm f} + \delta p_{\rm inf}}{\bar{p}'}\Big], \tag{4.5.4}$$

where $\bar{\rho}$ and $\bar{p}$ are the overall background energy density and pressure. A short computation using the Friedmann equations (4.2.5) and (4.2.6) gives

$$\bar{\rho}' = -\frac{6\mathcal{H}^3 m_p^2}{a^2}\epsilon, \qquad\qquad \bar{p}' = \frac{2\mathcal{H}^3 m_p^2}{a^2}\epsilon(3-2\eta), \tag{4.5.5}$$

where $\epsilon \equiv -\dot{H}/H^2$ and $\eta \equiv -\ddot{H}/(2\dot{H}H)$ are the two Hubble slow-roll functions parameterizing the degree to which the background de Sitter symmetry is broken. It is also convenient to define

$$\begin{aligned} \eta^0_{\rm inf} &\equiv \frac{-1}{2H^3}\frac{\ddot{H}_\phi}{\varepsilon_{\rm inf}}, \qquad & \eta^0_{\rm f} &\equiv \frac{-1}{2H^3}\frac{\ddot{H}_{\rm f}}{\varepsilon_{\rm inf}} \\ \varepsilon_{\rm inf} &\equiv -\frac{\dot{H}_\phi}{H^2}, \qquad & \varepsilon_{\rm f} &\equiv -\frac{\dot{H}_f}{H^2}, \end{aligned} \tag{4.5.6}$$

where the subscript ϕ(f) on derivatives of the Hubble parameter denotes the contribution due to the inflaton (fluid).

Next we examine the fluid and inflaton perturbation functions. The inflaton perturbation functions are [11]

$$\delta\rho_{\rm inf} = \delta X + V_{,\phi}\delta\phi, \qquad \delta p_{\rm inf} = \delta X - V_{,\phi}\delta\phi, \tag{4.5.7}$$

where we defined $\delta X \equiv \bar{\phi}'\delta\phi'/a^2 - \bar{\phi}'^2\Phi/a^2$. The fluid perturbation functions can be derived from the linearized Einstein equations [40]

$$\delta\rho_{\rm f} = \frac{2\beta}{a^2}[3\Psi - \partial^2 E - \partial^2\xi], \qquad \delta p_{\rm f} = c_s^2\delta\rho_{\rm f}, \tag{4.5.8}$$

which using the constraint equation (4.3.11) reduces to

$$\delta\rho_{\rm f} = -\frac{2}{a^4 c_s^2}\left[\left(\frac{ac_s}{2}\sqrt{\beta}\theta\right)' + a^2\beta\Phi\right]. \tag{4.5.9}$$

Using Eqs. (4.5.6) to (4.5.9), we write $\mathcal{S}$ as a power series in the fluid slow-roll parameters $\eta_{\rm f}^0$ and $\varepsilon_{\rm f}$,

$$\mathcal{S} = \mathcal{S}^{(0)} + \mathcal{S}^{(1)} + ... \tag{4.5.10}$$

with

$$\mathcal{S}^{(0)} = \frac{1}{\mathcal{H}}\frac{2(\eta_{\rm inf}^0 - 3)}{3(3-2\eta_{\rm inf}^0)}\mathcal{R}'_{\rm inf} \tag{4.5.11}$$

and

$$\begin{aligned}\mathcal{S}^{(1)} = {}& \frac{1}{\mathcal{H}}\left[\frac{2(\eta_{\rm inf}^0 - 3)}{3(3-2\eta_{\rm inf}^0)\varepsilon_{\rm inf}}\varepsilon_{\rm f} - \frac{2\eta_{\rm inf}^0}{(3-2\eta_{\rm inf}^0)^2}\left(\frac{\eta_{\rm f}^0}{\eta_{\rm inf}^0} - \frac{\varepsilon_{\rm f}}{\varepsilon_{\rm inf}}\right)\right]\mathcal{R}'_{\rm inf} \\ & -\frac{1}{\mathcal{H}}\left[\frac{\varepsilon_{\rm f}}{3\varepsilon_{\rm inf}c_s^2}\left(1 + \frac{3c_s^2}{3-2\eta_{\rm inf}^0}\right)\right]\mathcal{R}'_{\rm f} + (\mathcal{R}_{\rm inf} - \mathcal{R}_{\rm f})\left[\left(\frac{1}{3c_s^2} - \frac{2}{3}\right)\varepsilon_{\rm f} - \frac{2\eta_{\rm inf}^0}{3-2\eta_{\rm inf}^0}\left(\frac{\eta_{\rm f}^0}{\eta_{\rm inf}^0} - \frac{\varepsilon_{\rm f}}{\varepsilon_{\rm inf}}\right)\right].\end{aligned} \tag{4.5.12}$$

Here we have used the definitions

$$\mathcal{R}_{\rm inf} \equiv \frac{\pi}{\zeta} \qquad \text{and} \qquad \mathcal{R}_{\rm f} \equiv -\frac{\nu}{\sqrt{2}z} \tag{4.5.13}$$

which are the comoving curvature perturbations in absence of the other species. Judging from the slow-roll parameters appearing in Eq. (4.5.12), one can see that after the second transition when $\eta_{\rm f}^0 \ll \eta_{\rm inf}^0$ and $\varepsilon_{\rm f} \ll \varepsilon_{\rm inf}$, $\mathcal{S}$ converges to $\mathcal{S}^{(0)}$, which is the entropy perturbation function of a single inflaton field. This suggests that as the fluid energy density becomes subdominant to that of inflaton, the entropy between the two species diminishes and the overall entropy perturbation function reduces to its value in the absence of fluid. This indicates that our model does not produce non-negligible entropy perturbations, just as in the case of single field inflation.

4.6 Power spectrum of the scalar perturbations

The power spectrum of the comoving curvature perturbation is the primary tool by which predictions of inflationary models are experimentally tested. Using the definition (4.5.1) we can write

$$\mathcal{R} \equiv \Psi - a\mathcal{H}\frac{\delta q_{\rm f} + \delta q_{\rm inf}}{\bar{\phi}'^2 + 2\beta} = \Psi - \frac{\mathcal{H}}{\bar{\phi}'^2 + 2\beta}\left[-\bar{\phi}'\delta\phi + c_s\frac{\sqrt{\beta}}{a}\theta\right] = \frac{\bar{\phi}'^2\frac{\pi}{\zeta}}{\bar{\phi}'^2 + 2\beta} - \frac{2\beta\frac{\nu}{\sqrt{2}z}}{\bar{\phi}'^2 + 2\beta}, \tag{4.6.1}$$

where the sum of momentum perturbations $\delta q_{\rm f} + \delta q_{\rm inf}$ was obtained from the right hand side of Eq. (4.3.13), being $-T^0{}_i/2$. Also the gauge invariant variables ν and π were introduced in Eq. (4.3.15). Note that π/ζ and $\nu/\sqrt{2}z$ would be the comoving curvature perturbations in inflaton or fluid-only models.

Since we are interested in understanding how primordial quantum fluctuations seed the formation of the CMB anisotropies, we need to compute the quantum correlation functions of the comoving curvature perturbation $\mathcal{R}$. The presence of derivative interactions between ν and π in action (4.3.16) renders the quantization procedure non-trivial. The details of this procedure is provided in App. D. It turns out that one can write the quantized modes of the comoving curvature perturbation $\mathcal{R}$ as

$$\hat{\mathcal{R}}(\tau, \mathbf{k}) = \mathbf{W}^T(\tau)\hat{\mathbf{\Pi}}(\tau, \mathbf{k}). \tag{4.6.2}$$

Here

$$\mathbf{W} \equiv \begin{pmatrix} \frac{\bar{\phi}'^2}{\bar{\phi}'^2+2\beta}\zeta^{-1} \\ \frac{-2\beta}{\bar{\phi}'^2+2\beta}(\sqrt{2}z)^{-1} \end{pmatrix}, \tag{4.6.3}$$

and the quantized modes for the generalized solutions $\hat{\mathbf{\Pi}}(\tau, \mathbf{k})$ are defined as

$$\hat{\mathbf{\Pi}}(\tau, \mathbf{k}) \equiv \sum_{A=1,2} \mathbf{\Pi}_A(\tau, k)\hat{a}_{A,\mathbf{k}} + \mathbf{\Pi}^*_A(\tau, k)\hat{a}^\dagger_{A,-\mathbf{k}}, \tag{4.6.4}$$

where similar to the notation used in Sec. 4.4

$$\mathbf{\Pi}_A \equiv \begin{pmatrix} \pi_A \\ \nu_A \end{pmatrix}, \qquad A = 1, 2 \tag{4.6.5}$$

with the set $\{\mathbf{\Pi}_1, \mathbf{\Pi}_2, \mathbf{\Pi}^*_1, \mathbf{\Pi}^*_2\}$ of four linearly independent solutions spanning the four dimensional space of solutions for the coupled system of differential equations (4.3.24). Also, the annihilation operators $\hat{a}_{A\mathbf{k}}$ eliminate the quantum mechanical ground states $|\Omega\rangle_A$ for the inflaton and the fluid perturbations. The existence of natural ground states is guaranteed by virtue of the adiabaticity of the scalar modes at early times discussed in Sec. 4.4 for the case of inflaton coupled to radiation, which is our primary interest.

Given the fact that there is a natural choice of ground state for the inflaton and fluid perturbations, we compute the power spectrum of the comoving curvature perturbation in the initial state

$$|\mathcal{I}\rangle \equiv |\Omega_1\rangle \otimes |\Omega_2\rangle, \tag{4.6.6}$$

where $|\Omega_{1(2)}\rangle$ is the initial ground state for inflaton (fluid) perturbations. The power spectrum of the comoving curvature perturbation then becomes

$$\langle\mathcal{I}|\hat{\mathcal{R}}(\tau_{\rm out},\mathbf{k})\hat{\mathcal{R}}(\tau_{\rm out},\mathbf{k}')|\mathcal{I}\rangle \equiv \delta(\mathbf{k}+\mathbf{k}')\mathcal{P}_{\mathcal{R}}(k) \tag{4.6.7}$$

where $\tau_{\rm out}$ is conventionally taken to be the conformal time for which $k=\mathcal{H}$, and

$$\begin{aligned}\mathcal{P}_{\mathcal{R}}(k) &= \sum_A \Big(\mathbf{W}^T\mathbf{\Pi}_A(k)\Big)\Big(\mathbf{W}^T\mathbf{\Pi}_A^*(k)\Big)\\ &= \sum_A\left[\left(\frac{\bar{\phi}'^2}{\bar{\phi}'^2+2\beta}\right)^2\frac{|\pi_A(k)|^2}{\zeta^2}+\left(\frac{2\beta}{\bar{\phi}'^2+2\beta}\right)^2\frac{|\nu_A(k)|^2}{2z^2}-2\,\mathrm{Re}\left\{\frac{2\beta\bar{\phi}'^2}{(\bar{\phi}'^2+2\beta)^2}\frac{\pi_A(k)\nu_A^*(k)}{\sqrt{2}z\zeta}\right\}\right],\end{aligned} \tag{4.6.8}$$

where all the background and perturbation functions are evaluated at $\tau_{\rm out}$. Note that the scalar power spectrum receives most of its contribution from the first term in the square bracket, since $\sqrt{\beta}/\bar{\phi}'\to 0$ with the fluid energy density becoming subdominant to that of inflaton.

I Case study

We now numerically evaluate the power spectrum (4.6.8) and then the dimensionless power spectrum $\Delta^2_{\mathcal{R}}\equiv k^3/(2\pi^2)\mathcal{P}_{\mathcal{R}}$ for the case of radiation coupled to an inflaton with a ϕ^2 potential.

As already discussed in Sec. 4.2, we shall fix the initial energy scale of the radiation at the Planck scale and specialize to the class of background cosmology for which $\bar{\phi}$ approaches a constant value in the $\tau\to 0$ limit. In this case, the remaining free parameters are the initial value of the inflaton field and the coupling parameter of the inflaton potential $m^2/2$. A super-Planckian field variation of order $\Delta\bar{\phi}\sim 15$ during inflation together with the measured values of $\Delta^2_{\mathcal{R}}$ and its scale dependence parameter n_s require $10^{-6}m_p\lesssim m\lesssim 10^{-5}m_p$ [11]. Knowing the range of these two parameters, we look for values that result in a suitable fit to the reconstructed values of the amplitude A_s and the scale dependence parameter n_s for the ϕ^2 potential model at the pivot scale $k_*=0.05$ Mpc^{-1}, as reported by the Planck collaboration [2].

In Fig. 20 we produce a plot of $\Delta^2_{\mathcal{R}}$ for a range of low multipoles $2\lesssim l\lesssim 28$, assuming that inflation is followed by instantaneous reheating. Different plots of the scalar power spectrum correspond to different values of an observationally relevant parameter $\mathfrak{R}$, which is the dimensionless

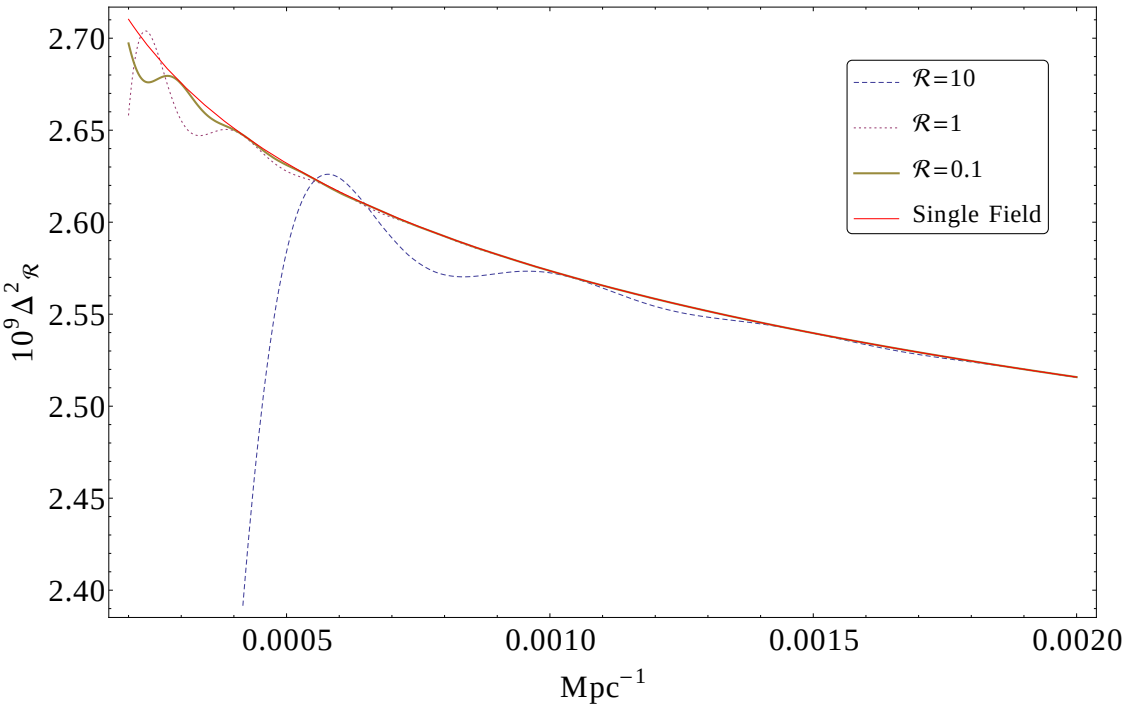


Figure 20: The power spectrum $\Delta^2_{\mathcal{R}}$ versus k for radiation coupled to an inflaton with the potential $V = m^2\phi^2/2$ for $2 \lesssim l \lesssim 28$ assuming instantaneous reheating. The parameter $\mathfrak{R}$ is the ratio of the fluid energy density to the inflaton energy density when the $l = 2$ mode left the horizon.

ratio of the fluid to inflaton energy densities when the quadrupole scale $l = 2$ exists the horizon. Fixing the value of m, we change $\mathfrak{R}$ by changing the initial values of the inflaton field.

A number remarks are in order here:

- In Sec. 4.4.I we argued that modes with wavenumbers $k \lesssim k_{\rm min}$ have never been in an adiabatic regime since the Planck era. This means that the quadrupole scale is not initially in an adiabatic regime for $\mathfrak{R} > 1$. For the case of $\mathfrak{R} = 10$, one can show that multipoles $l \gtrsim 5$ are initially in an adiabatic regime.

- The coupling of inflaton to radiation produces two main features in the scalar power spectrum: large amplitude suppression followed by small oscillations for low multipoles. The oscillations die off rapidly as k increases and the scalar power spectrum quickly converges to the standard single field result. The period of oscillations, which is larger for larger values of $\mathfrak{R}$, can extend over a few multipoles. Nonetheless, the amplitude of oscillations does not exceed a few percent.

- Observational data for $l \lesssim 5$ are not reliable due to a high value of cosmic variance. For values

of $\mathfrak{R} \gtrsim 1$, one can expect deviations from the standard single field result for $l \gtrsim 5$, as indicated in Fig. 20 for $\mathfrak{R} = 10$. Even though the small oscillations produced in this model do not appear to be detectable, the sharp suppression at the end tail of the power spectrum could very well be. Nonetheless, this suppression is only slightly outside of the range of multipoles with large cosmic variance. Also one should bear in mind that these modes are barely adiabatic at the initial time.

- Oscillatory features for low multipoles have previously been discovered in a variety of models, including trans-Planckian models [121, 122, 123, 124], axion monodromy models with additional instanton induced oscillatory features in the potential [125], models with brief non slow-roll phases [114, 126, 127, 128, 129], models of cascade inflation [130, 131], and models with a sudden change in the speed of sound [132, 133, 134]. In our case, the oscillations are produced while the adiabaticity for all modes within the observable range as well as the slow-roll nature of the background dynamics are preserved.

- Sharp suppressions of the scalar power spectrum for small multipoles have been previously suggested in cutoff models [135], models of cascade inflation [130, 131], as well as models with an initial non-slow-roll dynamics [114]. Suppressions in the former class of models extend over a larger range of multipoles than in our case. However, the ones observed in the latter class of models are quite similar to ours.

- The features observed in this particular case study are in a good agreement with the ones seen in the previous analyses of this model. In particular, both the oscillatory features as well as the sharp suppressions pertaining to the low multipole scalar perturbation modes were noticed in [136] and [137]. The large suppressions were also noticed in [138].

- Finally we comment on the effects of a non-trivial reheating scenario on the aforementioned results. A non-trivial reheating scenario in which inflation is not immediately succeeded by an era of radiation domination results in a smaller ratio a_i/a_0 of the initial scale factor and the scale factor today. Therefore, the quadrupole scale today would correspond to a larger momentum scale when the fluid energy density is Planckian. Larger values of k exit the horizon at later times, when the ratio of the fluid energy density over the inflaton energy density is smaller. Hence lowering the temperature of reheating in non-trivial reheating models

correspond to having smaller deviations from the standard power spectrum at low multipoles.

4.7 Power spectrum of gravitational waves

In this section we compute the power spectrum of the stochastic background of gravitational waves produced in this model. The quantization procedure for the gravitational waves in this model is quite standard and can be found in chapter 2.

The power spectrum of the gravitational waves is defined as

$$\langle \mathcal{J}|\hat{h}_{ij}(\boldsymbol{k},\tau_{\rm out})\hat{h}^{ij}(\boldsymbol{k'},\tau_{\rm out})|\mathcal{J}\rangle \equiv \delta(\boldsymbol{k}+\boldsymbol{k'})\mathcal{P}_h(k), \tag{4.7.1}$$

where $\tau_{\rm out}$ is the conformal time when $k = \mathcal{H}$, and $|\mathcal{J}\rangle$ is the Bunch Davies initial state for the tensor modes. Here we have

$$\mathcal{P}_h(k) = 4\frac{|h(k,\tau_{\rm out})|^2}{a^2(\tau_{\rm out})}. \tag{4.7.2}$$

Just as in the case of scalar power spectrum, we are interested in computing the dimensionless tensor power spectrum $\Delta_h^2 \equiv k^3/(2\pi^2)\mathcal{P}_h$.

I Case study

We numerically compute the tensor power spectrum from Eq. (4.7.1) for the same model studied in Sec. 4.6.I. We choose

$$H_{\rm inf} = 3 \times \sqrt{\frac{r}{0.1}} \times 10^{-5} m_p \approx 4.24 \times 10^{-5} m_p \tag{4.7.3}$$

for the Hubble parameter during inflation, which corresponds to a tensor-to-scalar ratio of $r = 0.2$. Using this value of the Hubble parameter, we calculate the amplitude A_t of the primordial gravitational waves during the standard single field slow-roll inflation to be

$$A_t = \frac{2}{\pi^2}\frac{H_{\rm inf}^2}{m_p^2} \approx 3.64 \times 10^{-10}. \tag{4.7.4}$$

As in Sec. 4.6.I, we select values of the inflaton mass m and the initial value of the inflaton field that give both an A_t close to the value above and $n_t \approx -2\epsilon$, as required from theoretical considerations. Fixing the value of m, we vary the initial value of the inflaton field to produce various values of $\mathfrak{R}$. Fig. 21 presents our result for different values of $\mathfrak{R}$.

We now make a number of remarks with regard to this result:

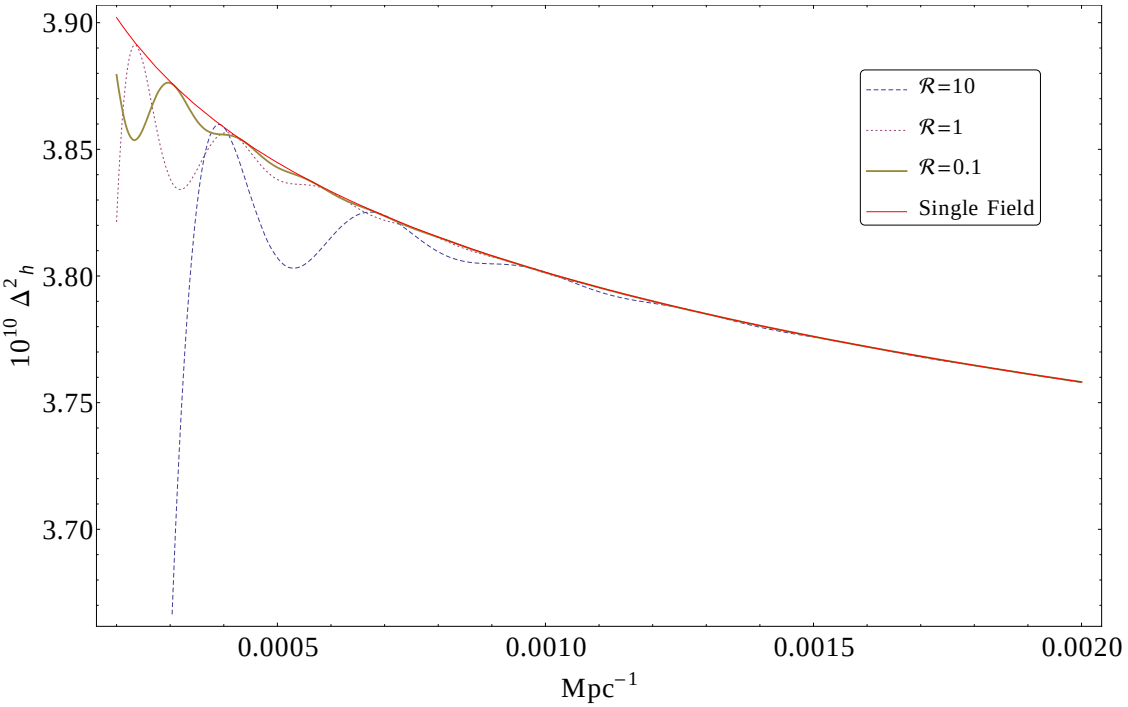


Figure 21: The tensor power spectrum Δ^2_h versus k for a model with radiation coupled to an inflaton with the potential $V = m^2\phi^2/2$ for $2 \lesssim l \lesssim 28$ assuming instantaneous reheating. The initial value of the inflaton field is slightly changed to produce different values of $\mathfrak{R}$.

- For tensor perturbations, one can show by explicit calculation that the quadrupole scale is initially in an adiabatic regime even for $\mathfrak{R} = 10$. Recall that for $\mathfrak{R} > 1$, the quadrupole scale is smaller than $k_{\rm min}$.
- The tensor power spectrum exhibits the same qualitative features as the scalar power spectrum; namely small oscillations and sharp suppression for order unity multipoles. The amplitude of oscillations however seem somewhat larger in this case, though not large enough to merit a phenomenological significance. Both of these features become insignificant for $l \gtrsim 5$.
- Similar oscillatory features were noticed in a study of trans-Planckian models in [139].
- The observed features are in a good qualitative agreement with the ones found in an earlier study of this model in [136].
- As in the case of scalar spectrum, having a non-trivial model of reheating corresponds to less deviations from the standard result in the observable window of scales.

Summary Of Results And Future Outlook

In part I we reviewed the fundamentals of single field inflation as well as the main predictions of inflation and their consistency with current observations. We also mentioned some of the forecasts on the main inflationary parameters [15] that are expected from the currently running and upcoming CMB and LSS observations. We then moved on to exploring two separate approaches to parametrizing the influence of pre-inflationary physics on single field slow-roll models of inflation, namely by using non trivial initial states for inflationary perturbations and by explicitly incorporating an extra degree of freedom that dominates the dynamics at energy scales above the energy scale of inflation. We found in particular that non trivial initial states are capable of producing large corrections to inflationary scalar correlation functions, though some fine tuning is required to evade the current observational constraints on the inflationary power spectrum.

Overall, the predictions of single field slow-roll inflation are in excellent agreement with all cosmological observations. The most prominent area of tension between the standard inflationary predictions and observations is perhaps the low multipole end [16] of the CMB temperature anisotropy power spectrum. As lower multipoles correspond to the modes that exited the Hubble horizon earlier during inflation, such a tension could provide some hints about the dynamics of pre-inflationary phases. As CMB measurements improve in the next few years [17] we expect to learn more about this particular anomaly and the status of various proposals that attempt to address it. Additionally, a potential detection of primordial gravitational waves using the B-mode polarization pattern on the CMB can not only determine the energy scale of inflation, but also it can provide detailed information regarding the inflationary Lagrangian by ruling out a large number of single field models while favouring a few.

[15] namely the spectral index, the running of the spectral index, and the amplitude of non-Gaussianities for scalar perturbations as well as the tensor-to-scalar ratio.

[16] corresponding to $l \lesssim 40$.

[17] Particularly through the CMB-S4 experiment, if implemented.

Appendix A

Horizons and Distances In Cosmology

Particle and event horizons

Null trajectories define boundaries of communicating regions in spacetimes. For a homogeneous and isotropic spacetime in four dimensions, a radial null trajectory $t(r)$ originating from $r = 0$ at t_1 is given by

$$ds^2 = 0 \Rightarrow \int_{t_1}^{t} \frac{dt'}{a(t')} = \int_0^r \frac{dr'}{\sqrt{1 - kr'^2}}, \tag{A.1}$$

where we used Eq. (1.1.1). Let us for simplicity only consider $k = 0$. The physical distance travelled by the null ray at the time t is then given by

$$d_p(t) = a(t)r = a(t) \int_{t_1}^{t} \frac{dt'}{a(t')}. \tag{A.2}$$

The above equation defines the particle horizon $d_p(t)$ at time t. This notion of horizon sets the boundary of the physical region outside of which the observer at the origin at time t could not have received any information that had originated after time t_1. As an example, consider the de Sitter metric where $a(t) = e^{Ht}$. The particle horizon at time t is given by

$$d_p(t) = e^{Ht} \int_{t_1}^{t} e^{-Ht'} dt' = \frac{1}{H}\Big[e^{H(t-t_1)} - 1\Big], \tag{A.3}$$

which diverges as $t_1 \to -\infty$, or as the de Sitter geometry extends to the Big Bang singularity. This is the statement that a sufficient period of inflation puts all regions of the spacetime in causal contact with one another.

If we now flip the limits of the integral given in Eq. (A.2) and take t_1 to correspond to the final coordinate time of the spacetime, we arrive at the event horizon $d_e(t)$ for an observer at the origin

at time t. If a spacetime has a finite timelike existence, then the time t_1 is usually taken to be the time of the Big Crunch, otherwise one takes $t_1 \to +\infty$. Doing the latter for the de Sitter geometry, we find

$$d_e(t) = e^{Ht} \int_t^{+\infty} e^{-Ht'} dt' = H^{-1}. \tag{A.4}$$

The above implies that all events outside the physical radius H^{-1} of an observer at any given time always remain unbeknownst to her.

In addition to the above notions of horizon in cosmology, it is also particularly common in the inflationary cosmology to use the notion of the Hubble horizon H^{-1}. The Hubble horizon is not a horizon in the kinematic sense. Its significance for the study of the inflationary perturbations is that it roughly defines a transition boundary for a perturbation mode as it changes its character from oscillatory to decaying with time [1].

Angular diameter distance

There are various notions of distance in cosmology that one can define. One of those, the angular diameter distance d_A, was used in chapter 1 in discussing the horizon problem. To see how this type of distance is defined, note that the angular interval $\Delta\theta$ subtended by a luminous object on the sky can be written as [41]

$$\Delta\theta = \frac{L}{d_A}, \tag{A.5}$$

where L is the transverse size of the object and d_A is the physical distance of the object to the observer at the time of light emission. The latter, d_A, is the angular diameter distance and is defined by

$$d_A(t) = a(t_1) \int_{t_1}^{t} \frac{dt'}{a(t')} = a(t_1) \int_{a(t_1)}^{a_0} \frac{da}{a^2 H(a)}, \tag{A.6}$$

where t_1 is the time of light emission, $H(a)$ is the Hubble parameter as a function of the scale factor, and we assumed spatial flatness for the FRW metric.

[1] See Eq. (2.3.8). As long as $k^2 \gg \zeta''/\zeta$, the perturbation mode oscillates with time. Once k^2 falls below $\zeta''/\zeta \sim H^2$, it no longer oscillates, rather it decays with time.

Appendix B

Metrics, Gauges, And Tensors

Einstein equations for the background FRW metric

Let us review how the Friedmann equations (1.1.2) and (1.1.3) are derived from the background Einstein equations

$$\bar{G}_{ab} = m_p^{-2}\bar{T}_{ab}. \tag{B.1}$$

For a homogeneous and isotropic Universe, the background stress energy tensor is given by

$$\bar{T}_{ab} = (\bar{\rho}_{\text{total}} + \bar{p}_{\text{total}})\bar{u}_a\bar{u}_b + \bar{p}_{\text{total}}\bar{g}_{ab}, \tag{B.2}$$

where $\bar{\rho}_{\text{total}}$ and $\bar{p}_{\text{total}}$ are the background energy density and pressure of all gravitational sources, $\bar{g}_{ab}$ is the FRW metric given in Eq. (1.1.1), and $\vec{u} = \vec{\partial}_t$ defines the background comoving frame that is given by the symmetries of the FRW metric. The non-vanishing components of the background stress energy tensor are

$$\bar{T}_{00} = \bar{\rho}_{\text{total}}, \qquad \bar{T}_{ii} = \bar{p}_{\text{total}}\bar{g}_{ii}, \tag{B.3}$$

where subscript 0 and i are time and space indices respectively.

The non vanishing components of the background Einstein tensor are given by [14]

$$\bar{G}_{00} = 3\left[H^2 + \frac{k}{a^2}\right], \qquad \bar{G}_{ii} = -\bar{g}_{ii}\left[3H^2 + 2\dot{H} + \frac{k}{a^2}\right] = -\bar{g}_{ii}\left[H^2 + 2\frac{\ddot{a}}{a} + \frac{k}{a^2}\right]. \tag{B.4}$$

The two Friedmann equations are then

$$\begin{aligned} &m_p^2\bar{G}_{00} = \bar{T}_{00} \Rightarrow 3m_p^2\left[H^2 + \frac{k}{a^2}\right] = \bar{\rho}_{\text{total}}, \\ &\frac{m_p^2}{6}\left[\bar{G}^0{}_0 - \bar{G}^i{}_i\right] = \frac{1}{6}\left[\bar{T}^0{}_0 - \bar{T}^i{}_i\right] \Rightarrow m_p^2\frac{\ddot{a}}{a} = -\frac{1}{6}(\bar{\rho}_{\text{total}} + 3\bar{p}_{\text{total}}), \end{aligned} \tag{B.5}$$

where summation over upper and lower spatial indices is assumed.

Review of stress energy tensor perturbations

Here we review how the stress energy tensor for a perfect fluid can be linearly perturbed. Recall that the stress energy tensor for a perfect fluid can be written as

$$T_{ab} = (\rho + p)u_a u_b + p g_{ab}. \tag{B.6}$$

where p, ρ, and u^a are the pressure, energy density, and the four-velocity of the fluid. An everywhere isotropic Universe requires the components of this tensor to only depend on time. However, allowing small spatial dependences for the fluid functions [1],

$$\begin{aligned} p(\tau, \boldsymbol{x}) \equiv \bar{p}(\tau) + \delta p(\tau, \boldsymbol{x}), \qquad \rho(\tau, \boldsymbol{x}) \equiv \bar{\rho}(\tau) + \delta\rho(\tau, \boldsymbol{x}), \\ u^a(\tau, \boldsymbol{x}) \equiv \bar{u}^a(\tau) + \delta u^a(\tau, \boldsymbol{x}), \end{aligned} \tag{B.7}$$

one is left with a slightly inhomogeneous and anisotropic Universe described by the linearly perturbed FRW metric

$$ds^2 = a(\tau)^2\Big[-(1+2\Phi)d\tau^2 + 2B_i d\tau dx^i + \big\{(1-2\Psi)\delta_{ij} + 2E_{ij}\big\}dx^i dx^j\Big]. \tag{B.8}$$

We then define the background fluid four-velocity vector $\bar{u}^a$ to be that of the frame in which the three-momentum density vanishes. Recalling that $g_{ab}u^a u^b = -1$ to all orders in perturbation, we find

$$\vec{\bar{u}} = \frac{1}{a}\vec{\partial}_\tau, \qquad \vec{\delta u} = -\frac{\Phi}{a}\vec{\partial}_\tau + \frac{v^i - B^i}{a}\vec{\partial}_i, \tag{B.9}$$

where $(v^i - B^i)/a$ is the fluid three-velocity vector to leading non-vanishing order for some three-vector $\boldsymbol{v}$. With these definitions, the linearly-perturbed stress energy tensor becomes

$$\begin{aligned} T_{\tau\tau} &= a^2[\bar{\rho} + \delta\rho + 2\bar{\rho}\Phi], \\ T_{i\tau} &= a^2\big[\bar{p}(B_i - v_i) - \bar{\rho}v_i\big], \\ T_{ij} &= a^2\big[\bar{p}\delta_{ij} + \delta p + 2\bar{p}(E_{ij} - \Psi)\big]. \end{aligned} \tag{B.10}$$

Finally we define a useful additive function known as the three-momentum density perturbations δq^i using [11]

$$\delta q^i \equiv (\bar{p} + \bar{\rho})a v^i. \tag{B.11}$$

[1] We are using the conformal time τ instead of the coordinate time t.

This quantity is proportional to the time-space component of the stress energy tensor,

$$T^{\tau}{}_{i} = \frac{1}{a}\delta q_i. \tag{B.12}$$

For scalar perturbations, one can define a three-velocity potential as $\delta u_{,i} \equiv a v_i$. This in turn gives rise to the three-momentum density potential δq defined by

$$\delta q_{,i} \equiv (\bar{p} + \bar{\rho})\delta u_{,i}. \tag{B.13}$$

Review of gauge transformations and gauge invariant functions

In this section we review the basics of gauge transformations in cosmological perturbation theory. We will then provide the gauge transformation relations for the perturbation functions used in this thesis. For an excellent discussion of this subject see [40] or a more recent review article [11].

Gauge transformations in cosmological perturbation theory are concerned with how perturbation functions change under coordinate transformations on some spacetime manifold. Given a set of coordinate system (τ, x^i), we define an infinitesimally shifted coordinate system using

$$\hat{\tau} \equiv \tau + \xi^0(\tau, x^i) \qquad \hat{x}^i \equiv x^i + \xi^i(\tau, x^i), \tag{B.14}$$

for some smooth functions ξ^0 and ξ^i. For scalar perturbations $\xi^i \equiv \partial^i \xi$, for some smooth function ξ. The line element defined in Eq. (B.8) is a geometric invariant, independent of choices of coordinate system. We should therefore have

$$ds^2 = g_{ab}(\tau, x^i)dx^a dx^b = g_{ab}(\hat{\tau}, \hat{x}^i)d\hat{x}^a d\hat{x}^b, \tag{B.15}$$

where $g_{ab}(\hat{\tau}, \hat{x}^i)$ and $d\hat{x}^a$ are the metric and differential 1-form defined using the new coordinate system $(\hat{\tau}, \hat{x}^i)$. We use Eq. (B.15) to derive the changes in the scalar metric perturbation functions,

$$\begin{aligned}
\hat{\Phi} &= \Phi - \mathcal{H}\xi^0 - \xi'^0, \\
\hat{\Psi} &= \Psi + \mathcal{H}\xi^0, \\
\hat{B} &= B + \xi^0 - \xi', \\
\hat{E} &= E - \xi.
\end{aligned} \tag{B.16}$$

Tensorial metric perturbations h_{ij} are defined to be traceless and transverse, rendering them gauge invariant. Imposing the traceless-transverse condition ensures that no piece of h_{ij} transforms as a scalar or a vector under the coordinate transformation (B.14).

In addition to the metric perturbation functions, scalar perturbation functions also undergo transformations upon changes in coordinate systems. For a scalar function $\alpha(\tau, x^i)$, we can define its perturbation as

$$\delta\alpha(\tau, x^i) \equiv \alpha(\tau, x^i) - \bar{\alpha}(\tau), \tag{B.17}$$

where $\bar{\alpha}(\tau)$ is the background value of $\alpha(\tau, x^i)$. Keeping $\alpha(\tau, x^i)$ invariant under the coordinate transformations, we get

$$\hat{\delta\alpha} = \delta\alpha - \bar{\alpha}'\xi^0. \tag{B.18}$$

The scalar perturbation functions in this thesis are the fluid three-velocity potential θ, the inflaton perturbation $\delta\phi$, the three-momentum density potential δq for the fluid and inflaton, and the energy density and pressure of both species. These scalar perturbation functions transform under an arbitrary coordinate transformation as

$$\begin{aligned}
\hat{\theta} &= \theta - \sqrt{2}z\mathcal{H}\xi^0, \\
\hat{\delta\phi} &= \delta\phi - \bar{\phi}'\xi^0, \\
\hat{\delta q} &= \delta q - a(\bar{\rho} + \bar{p})\xi^0, \\
\hat{\delta\rho} &= \delta\rho - \bar{\rho}'\xi^0, \\
\hat{\delta p} &= \delta p - \bar{p}'\xi^0,
\end{aligned} \tag{B.19}$$

where $z = a\sqrt{\beta}/(\mathcal{H}c_s)$ with $\beta = (a^2/2)(\bar{\rho}_{\rm f} + \bar{p}_{\rm f})$.

Using the gauge transformation relations derived above for the metric and scalar perturbation functions, one can form linear gauge-invariant perturbation functions. The ones used in this thesis are the Mukhanov variables ν and π for the fluid and inflaton,

$$\nu \equiv \frac{1}{\sqrt{2}}(\theta - 2z\Psi), \qquad \pi \equiv a\delta\phi + \zeta\Psi, \tag{B.20}$$

the comoving curvature perturbation $\mathcal{R}$,

$$\mathcal{R} \equiv \Psi - \mathcal{H}\frac{\delta q}{a(\bar{\rho} + \bar{p})}, \tag{B.21}$$

and the entropy perturbation function $\mathcal{S}$

$$\mathcal{S} \equiv \mathcal{H}\left(\frac{\delta\rho}{\bar{\rho}'} - \frac{\delta p}{\bar{p}'}\right). \tag{B.22}$$

Appendix C

Appendices to chapter 3

Free functions for the dressed vacuum state

In this appendix we compute the free functions $\mathcal{F}_i$, $\mathcal{H}_i$ and Γ_i that characterize the initial state, for the ground state of the interacting theory, the dressed vacuum.

We are interested in evaluating the initial two point, three point and four point functions. The general n-point function is

$$\langle\Omega|\hat{a}^{\alpha}_{\mathbf{k}_1}...\hat{a}^{\delta}_{\mathbf{k}_n}|\Omega\rangle, \tag{C.1}$$

where $|\Omega\rangle$ is the dressed vacuum state. Here the indices α, β etc. can be 0 or 1; the notation is that $\hat{a}^{\alpha}$ means $\hat{a}$ for $\alpha = 0$ and $\hat{a}^{\dagger}$ for $\alpha = 1$. Note that the n-point function (C.1) is time dependent. We will evaluate this quantity at some initial conformal time τ_i.

As usual, we can evolve expectation values in the dressed vacuum by starting with the free vacuum at early times, with the early time slightly complexified [59]:

$$\langle\Omega|\hat{a}^{\alpha}_{\mathbf{k}_1}...\hat{a}^{\delta}_{\mathbf{k}_n}|\Omega\rangle = \lim_{\tau\to-\infty(1-i\epsilon)}\langle 0|\hat{U}^{\dagger}(\tau_i,\tau)\hat{a}^{\alpha}_{\mathbf{k}_1}...\hat{a}^{\delta}_{\mathbf{k}_n}\hat{U}(\tau_i,\tau)|0\rangle. \tag{C.2}$$

Here $\hat{U}(\tau_i,\tau)$ is the unitary evolution operator that maps interaction picture states at conformal time τ to conformal time τ_i, given by

$$\hat{U}(\tau_i,\tau) = T\exp\left\{-i\int_{\tau}^{\tau_i}d\tau'\hat{H}^{I}_{\rm int}(\tau')\right\}. \tag{C.3}$$

Expanding to the leading order in interactions gives

$$\langle\Omega|\hat{a}^{\alpha}_{\mathbf{k}_1}...\hat{a}^{\delta}_{\mathbf{k}_n}|\Omega\rangle = \langle 0|\hat{a}^{\alpha}_{\mathbf{k}_1}...\hat{a}^{\delta}_{\mathbf{k}_n}|0\rangle + i\int_{-\infty(1\pm i\epsilon)}^{\tau_i}d\tau\left\langle 0\left|\left[\hat{H}_{\rm int}(\tau),\hat{a}^{\alpha}_{\mathbf{k}_1}...\hat{a}^{\delta}_{\mathbf{k}_n}\right]\right|0\right\rangle, \tag{C.4}$$

where we dropped the superscript I.

Next, we use the explicit formula (3.3.2) for the interaction Hamiltonian. Since this interaction Hamiltonian is cubic in the fields, and since the vacuum expectation value of the product of an odd number of creation or annihilation operators vanishes, it follows from Eq. (C.4) that all of the free functions (3.4.1) and (3.4.3) vanish. The only possible non-vanishing free functions are $\mathcal{H}_1$ and $\mathcal{H}_2$, defined by Eq. (3.4.2). We now evaluate $\mathcal{H}_1^{0,\mathrm{dr}}$ to leading order:

$$\begin{aligned}
\langle\Omega|\hat{a}_{\mathbf{k}_1}\hat{a}_{\mathbf{k}_2}\hat{a}_{\mathbf{k}_3}|\Omega\rangle &= \delta(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\mathcal{H}_1^{0,\mathrm{dr}}(\mathbf{k}_1,\mathbf{k}_2) \\
&= i\int_{-\infty(1\pm i\epsilon)}^{\tau_i} d\tau \left\langle 0\right| \left[\hat{H}_{\mathrm{int}}, \hat{a}_{\mathbf{k}_1}\hat{a}_{\mathbf{k}_2}\hat{a}_{\mathbf{k}_3}\right] \left|0\right\rangle \\
&= -\frac{\sqrt{\epsilon}H}{2m_p}\int d\tau \int \frac{d^3k'}{(2\pi)^{3/2}} \int d^3p \frac{\sqrt{k'|\mathbf{p}-\mathbf{k}'|p}}{p^2} \left\langle 0\right| \Big[(\hat{a}_{\mathbf{k}'}e^{-i\tau k'} - \hat{a}^\dagger_{-\mathbf{k}'}e^{i\tau k'}) \\
&\quad (\hat{a}_{\mathbf{p}-\mathbf{k}'}e^{-i\tau|\mathbf{p}-\mathbf{k}'|} - \hat{a}^\dagger_{\mathbf{p}-\mathbf{k}'}e^{i\tau|\mathbf{p}-\mathbf{k}'|})(\hat{a}_{-\mathbf{p}}e^{-i\tau p} - \hat{a}^\dagger_{\mathbf{p}}e^{i\tau p}), \hat{a}_{\mathbf{k}_1}\hat{a}_{\mathbf{k}_2}\hat{a}_{\mathbf{k}_3}\Big] \left|0\right\rangle \\
&= -\frac{\sqrt{\epsilon}H}{2m_p}\int_{-\infty(1-i\epsilon)}^{\tau_i} d\tau \int \frac{d^3k'}{(2\pi)^{3/2}} \int d^3p \frac{\sqrt{k'|\mathbf{p}-\mathbf{k}'|p}}{p^2} \\
&\quad \times \left\langle 0\right| \hat{a}_{\mathbf{k}_1}\hat{a}_{\mathbf{k}_2}\hat{a}_{\mathbf{k}_3}\hat{a}^\dagger_{-\mathbf{k}'}\hat{a}^\dagger_{\mathbf{p}-\mathbf{k}'}\hat{a}^\dagger_{\mathbf{p}} \left|0\right\rangle e^{i\tau(k'+|\mathbf{p}-\mathbf{k}'|+p)}. \qquad \text{(C.5)}
\end{aligned}$$

Here we have used the interaction Hamiltonian (3.3.2), the mode expansion (3.3.4) and Eq. (3.3.5) for the mode functions. Simplifying Eq. (C.5) gives

$$\mathcal{H}_1^{0,\mathrm{dr}}(\mathbf{k}_1,\mathbf{k}_2) = i\frac{H}{(2\pi)^{3/2}m_p}\sqrt{\epsilon}\frac{(k_1k_2k_3)^{1/2}}{k_t}\sum_{i=1}^{3}\frac{1}{k_i^2}e^{i\tau_i k_t}, \qquad \text{(C.6)}$$

where $\mathbf{k}_3 = -\mathbf{k}_1 - \mathbf{k}_2$ and $k_t = k_1 + k_2 + k_3$. A similar calculation shows that $\mathcal{H}_2^{0,\mathrm{dr}}$ vanishes.

We note that there is no contribution to these computations from the field redefinition (3.3.3), since the operators $\hat{a}_{\mathbf{k}}$ are defined in terms of the mode expansion (3.3.4) of the redefined field operator $\hat{\mathcal{R}}_c$.

Subleading corrections to the power spectrum

Here we compute the subleading corrections to the power spectrum given in Eq. (3.5.4). Note that there are two separate contributions; first due to the field redefinitions, and then due to Hamiltonian evolution.

Field redefinition contribution

To leading order in the slow-roll parameters, this piece of the subleading power spectrum is given by

$$\begin{aligned}
&\langle\hat{\mathcal{R}}(0,\mathbf{k}_1)\hat{\mathcal{R}}(0,\mathbf{k}_2)\rangle - \langle\hat{\mathcal{R}}_c(0,\mathbf{k}_1)\hat{\mathcal{R}}_c(0,\mathbf{k}_2)\rangle \\
&= \left[\frac{\epsilon}{4}-\frac{\eta}{2}\right]\langle\hat{\mathcal{R}}_c(0,\mathbf{k}_1)\hat{\mathcal{R}}_c^2(0,\mathbf{k}_2)\rangle + \frac{\epsilon}{2}\langle\hat{\mathcal{R}}_c(0,\mathbf{k}_1)\partial^{-2}[\hat{\mathcal{R}}_c\partial^2\hat{\mathcal{R}}_c](0,\mathbf{k}_2)\rangle + (1\leftrightarrow 2). \qquad \text{(C.7)}
\end{aligned}$$

The first piece is

$$\begin{aligned}
&\left[\frac{\epsilon}{4}-\frac{\eta}{2}\right]\langle\hat{\mathcal{R}}_c(0,\mathbf{k}_1)\hat{\mathcal{R}}_c^2(0,\mathbf{k}_2)\rangle + (1\leftrightarrow 2) \\
&= \left[\frac{\epsilon}{4}-\frac{\eta}{2}\right]\left\langle \int \frac{d^3p}{(2\pi)^{3/2}}\,[\hat{\mathcal{R}}_c(0,\mathbf{k}_1)\hat{\mathcal{R}}_c(0,\mathbf{p})\hat{\mathcal{R}}_c(0,\mathbf{k}_2-\mathbf{p}) + \hat{\mathcal{R}}_c(0,\mathbf{k}_2)\hat{\mathcal{R}}_c(0,\mathbf{p})\hat{\mathcal{R}}_c(0,\mathbf{k}_1-\mathbf{p})]\right\rangle \\
&= -i\left[\frac{\epsilon}{4}-\frac{\eta}{2}\right]\frac{H^3}{8m_p^3\epsilon^{3/2}}\int \frac{d^3p}{(2\pi)^{3/2}}\left\{\sqrt{\frac{1}{k_1^3p^3|\mathbf{k}_2-\mathbf{p}|^3}}\left\langle(\hat{a}_{\mathbf{k}_1}-\hat{a}^\dagger_{-\mathbf{k}_1})(\hat{a}_{\mathbf{p}}-\hat{a}^\dagger_{-\mathbf{p}})(\hat{a}_{\mathbf{k}_2-\mathbf{p}}-\hat{a}^\dagger_{\mathbf{p}-\mathbf{k}_2})\right\rangle\right. \\
&\left.+\sqrt{\frac{1}{k_2^3p^3|\mathbf{k}_1-\mathbf{p}|^3}}\left\langle(\hat{a}_{\mathbf{k}_2}-\hat{a}^\dagger_{-\mathbf{k}_2})(\hat{a}_{\mathbf{p}}-\hat{a}^\dagger_{-\mathbf{p}})(\hat{a}_{\mathbf{k}_1-\mathbf{p}}-\hat{a}^\dagger_{\mathbf{p}-\mathbf{k}_1})\right\rangle\right\}. \qquad \text{(C.8)}
\end{aligned}$$

The operators expectation values are

$$\begin{aligned}
&\left\langle(\hat{a}_{\mathbf{k}_1}-\hat{a}^\dagger_{-\mathbf{k}_1})(\hat{a}_{\mathbf{p}}-\hat{a}^\dagger_{-\mathbf{p}})(\hat{a}_{\mathbf{k}_2-\mathbf{p}}-\hat{a}^\dagger_{\mathbf{p}-\mathbf{k}_2})\right\rangle \\
&= \delta(\mathbf{k}_1+\mathbf{k}_2)\Big[\mathcal{H}_1(\mathbf{k}_1,\mathbf{p}) - \mathcal{H}_2(\mathbf{k}_1,\mathbf{p}) - \mathcal{H}_2(\mathbf{k}_1,\mathbf{k}_2-\mathbf{p}) + \mathcal{H}_2^*(-\mathbf{p},\mathbf{p}-\mathbf{k}_2) - \mathcal{H}_2(\mathbf{p},\mathbf{k}_2-\mathbf{p}) \\
&+\mathcal{H}_2^*(-\mathbf{k}_1,\mathbf{p}-\mathbf{k}_2) + \mathcal{H}_2^*(-\mathbf{k}_1,-\mathbf{p}) - \mathcal{H}_1^*(-\mathbf{k}_1,-\mathbf{p})\Big], \qquad \text{(C.9)}
\end{aligned}$$

and the same with $1\leftrightarrow 2$ for the second expectation value.

The second piece is

$$\begin{aligned}
&\frac{\epsilon}{2}\langle\hat{\mathcal{R}}_c(0,\mathbf{k}_1)\partial^{-2}[\hat{\mathcal{R}}_c\partial^2\hat{\mathcal{R}}_c](0,\mathbf{k}_2)\rangle + (1\leftrightarrow 2) \\
&= \frac{\epsilon}{2}\left\langle \int \frac{d^3p}{(2\pi)^{3/2}}\,\frac{|\mathbf{k}_2-\mathbf{p}|^2}{k_2^2}\hat{\mathcal{R}}_c(0,\mathbf{k}_1)\hat{\mathcal{R}}_c(0,\mathbf{p})\hat{\mathcal{R}}_c(0,\mathbf{k}_2-\mathbf{p}) + \frac{|\mathbf{k}_1-\mathbf{p}|^2}{k_1^2}\hat{\mathcal{R}}_c(0,\mathbf{k}_2)\right. \\
&\left.\hat{\mathcal{R}}_c(0,\mathbf{p})\hat{\mathcal{R}}_c(0,\mathbf{k}_1-\mathbf{p})\right\rangle = -i\frac{H^3}{16m_p^3\epsilon^{1/2}}\int \frac{d^3p}{(2\pi)^{3/2}}\left\{\frac{\sqrt{|\mathbf{k}_2-\mathbf{p}|}}{k_1^{3/2}p^{3/2}k_2^2}\left\langle(\hat{a}_{\mathbf{k}_1}-\hat{a}^\dagger_{-\mathbf{k}_1})(\hat{a}_{\mathbf{p}}-\hat{a}^\dagger_{-\mathbf{p}})\right.\right. \\
&\left.\left.(\hat{a}_{\mathbf{k}_2-\mathbf{p}}-\hat{a}^\dagger_{\mathbf{p}-\mathbf{k}_2})\right\rangle + \frac{\sqrt{|\mathbf{k}_1-\mathbf{p}|}}{k_2^{3/2}p^{3/2}k_1^2}\left\langle(\hat{a}_{\mathbf{k}_2}-\hat{a}^\dagger_{-\mathbf{k}_2})(\hat{a}_{\mathbf{p}}-\hat{a}^\dagger_{-\mathbf{p}})(\hat{a}_{\mathbf{k}_1-\mathbf{p}}-\hat{a}^\dagger_{\mathbf{p}-\mathbf{k}_1})\right\rangle\right\}. \qquad \text{(C.10)}
\end{aligned}$$

The operators expectation values are identical to the one given above.

Commutator contribution

The commutator contribution is given by

$$
\begin{aligned}
&-i\int_{\tau_i}^{0}d\tau'\left\langle\left[\hat{\mathcal{R}}_c(0,\mathbf{k}_1)\hat{\mathcal{R}}_c(0,\mathbf{k}_2),\hat{H}_{\rm int}\right]\right\rangle\\
&=\frac{H^3}{8\epsilon^{1/2}m_p^3k_1^{3/2}k_2^{3/2}}\int_{\tau_i}^{0}d\tau'\int\frac{d^3k'd^3p}{(2\pi)^{3/2}}\frac{\sqrt{k'p|\mathbf{p}-\mathbf{k}'|}}{p^2}\left\langle\Big[(\hat{a}_{\mathbf{k}_1}-\hat{a}^\dagger_{-\mathbf{k}_1})(\hat{a}_{\mathbf{k}_2}-\hat{a}^\dagger_{-\mathbf{k}_2})\right.\\
&\left.,(\hat{a}_{\mathbf{k}'}e^{-i\tau'k'}-\hat{a}^\dagger_{-\mathbf{k}'}e^{i\tau'k'})(\hat{a}_{\mathbf{p}-\mathbf{k}'}e^{-i\tau'|\mathbf{p}-\mathbf{k}'|}-\hat{a}^\dagger_{\mathbf{k}'-\mathbf{p}}e^{i\tau'|\mathbf{p}-\mathbf{k}'|})(\hat{a}_{-\mathbf{p}}e^{-i\tau'p}-\hat{a}^\dagger_{\mathbf{p}}e^{i\tau'p})\Big]\right\rangle.
\end{aligned}
\tag{C.11}
$$

The operators commutator becomes

$$
\begin{aligned}
&\delta(\mathbf{p}-\mathbf{k}_1)\Big\{\Big[e^{-i\tau'(k'+|\mathbf{p}-\mathbf{k}'|+p)}-e^{-i\tau'(k'+|\mathbf{p}-\mathbf{k}'|-p)}\Big]\langle\hat{a}_{\mathbf{k}_2}\hat{a}_{\mathbf{k}'}\hat{a}_{\mathbf{p}-\mathbf{k}'}-\hat{a}^\dagger_{-\mathbf{k}_2}\hat{a}_{\mathbf{p}-\mathbf{k}'}\hat{a}_{\mathbf{k}'}\rangle\\
&+\Big[e^{-i\tau'(k'-|\mathbf{p}-\mathbf{k}'|-p)}-e^{-i\tau'(k'-|\mathbf{p}-\mathbf{k}'|+p)}\Big]\langle\hat{a}^\dagger_{\mathbf{k}'-\mathbf{p}}\hat{a}_{\mathbf{k}_2}\hat{a}_{\mathbf{k}'}-\hat{a}^\dagger_{-\mathbf{k}_2}\hat{a}^\dagger_{\mathbf{k}'-\mathbf{p}}\hat{a}_{\mathbf{k}'}\rangle\\
&+\Big[e^{-i\tau'(-k'+|\mathbf{p}-\mathbf{k}'|-p)}-e^{-i\tau'(-k'+|\mathbf{p}-\mathbf{k}'|+p)}\Big]\langle\hat{a}^\dagger_{-\mathbf{k}'}\hat{a}_{\mathbf{k}_2}\hat{a}_{\mathbf{p}-\mathbf{k}'}-\hat{a}^\dagger_{-\mathbf{k}_2}\hat{a}^\dagger_{-\mathbf{k}'}\hat{a}_{\mathbf{p}-\mathbf{k}'}\rangle\\
&+\Big[e^{-i\tau'(-k'-|\mathbf{p}-\mathbf{k}'|+p)}-e^{-i\tau'(-k'-|\mathbf{p}-\mathbf{k}'|-p)}\Big]\langle\hat{a}^\dagger_{-\mathbf{k}'}\hat{a}^\dagger_{\mathbf{k}'-\mathbf{p}}\hat{a}_{\mathbf{k}_2}-\hat{a}^\dagger_{-\mathbf{k}_2}\hat{a}^\dagger_{-\mathbf{k}'}\hat{a}^\dagger_{\mathbf{k}'-\mathbf{p}}\rangle\Big\}\\
&+(\mathbf{p}\leftrightarrow-\mathbf{k}')+(\mathbf{p}\leftrightarrow\mathbf{k}'-\mathbf{p})
\end{aligned}
\tag{C.12}
$$

as well as the same result with $\mathbf{k}_1$ being replaced by $\mathbf{k}_2$. The result given in Eq. (3.5.12) then follows easily from above.

Explicit calculation of the bispectrum

We fix values of $\mathbf{k}_1$, $\mathbf{k}_2$ and $\mathbf{k}_3$, all of which we assume to be nonvanishing. As discussed previously, there are two contributions to the bispectrum, one from the field redefinition (3.3.3) and one from the interaction Hamiltonian. We now compute these two contributions in turn.

Field redefinition contribution

We substitute the field redefinition (3.3.3) into the first term on the right hand side of Eq. (3.6.2). We neglect the second term in Eq. (3.6.2) for the moment; that term will be treated in Sec. C below.

The result is

$$\begin{aligned}\langle\hat{\mathcal{R}}(0,\mathbf{k}_1)\hat{\mathcal{R}}(0,\mathbf{k}_2)\hat{\mathcal{R}}(0,\mathbf{k}_3)\rangle &= \langle\hat{\mathcal{R}}_c(0,\mathbf{k}_1)\hat{\mathcal{R}}_c(0,\mathbf{k}_2)\hat{\mathcal{R}}_c(0,\mathbf{k}_3)\rangle \\ &+\left(\frac{1}{2}\frac{\ddot{\dot{\phi}}}{\dot{\phi}H}+\frac{1}{8m_p^2}\frac{\dot{\phi}^2}{H^2}\right)\Bigg(\int\frac{d^3p}{(2\pi)^{3/2}}\langle\hat{\mathcal{R}}_c(0,\mathbf{k}_1)\hat{\mathcal{R}}_c(0,\mathbf{k}_2)\hat{\mathcal{R}}_c(0,\mathbf{p}) \\ &\hat{\mathcal{R}}_c(0,\mathbf{k}_3-\mathbf{p})\rangle+\text{cyclic perms.}\Bigg)+\frac{\dot{\phi}^2}{4m_p^2H^2}\Bigg(\frac{1}{k_3^2}\int\frac{d^3p}{(2\pi)^{3/2}}\langle\hat{\mathcal{R}}_c(0,\mathbf{k}_1) \\ &\hat{\mathcal{R}}_c(0,\mathbf{k}_2)\hat{\mathcal{R}}_c(0,\mathbf{p})\hat{\mathcal{R}}_c(0,\mathbf{k}_3-\mathbf{p})\rangle(\mathbf{k}_3-\mathbf{p})^2+\text{cyclic perms.}\Bigg). \quad \text{(C.13)}\end{aligned}$$

Using the mode expansion (3.3.4), Eq. (2.5.2) for the superhorizon mode functions, the decomposition (3.4.10) of the function $\mathcal{H}_1$, and noting that 1-point functions vanish because of our homogeneity assumption, we find for the first term in Eq. (C.13)

$$\langle\hat{\mathcal{R}}_c(0,\mathbf{k}_1)\hat{\mathcal{R}}_c(0,\mathbf{k}_2)\hat{\mathcal{R}}_c(0,\mathbf{k}_3)\rangle=\delta(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\left[\mathcal{B}_{III}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)+\mathcal{B}_{III}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)^{0,\mathrm{dr}}\right]. \quad \text{(C.14)}$$

Here

$$\begin{aligned}\mathcal{B}_{III}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) &= -\frac{iH^3}{8m_p^3\epsilon^{\frac{3}{2}}}\frac{1}{(k_1k_2k_3)^{\frac{3}{2}}}\Big[\hat{\mathcal{H}}_1(\mathbf{k}_1,\mathbf{k}_2)-\hat{\mathcal{H}}_1^*(-\mathbf{k}_1,-\mathbf{k}_2)-\mathcal{H}_2(\mathbf{k}_1,\mathbf{k}_2) \\ &+\mathcal{H}_2^*(-\mathbf{k}_1,-\mathbf{k}_2)-\mathcal{H}_2(\mathbf{k}_1,\mathbf{k}_3)+\mathcal{H}_2^*(-\mathbf{k}_1,-\mathbf{k}_3)-\mathcal{H}_2(\mathbf{k}_2,\mathbf{k}_3) \\ &+\mathcal{H}_2^*(-\mathbf{k}_2,-\mathbf{k}_3)\Big] \quad \text{(C.15)}\end{aligned}$$

and

$$\begin{aligned}\mathcal{B}_{III}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)^{0,\mathrm{dr}} &= -\frac{iH^3}{8m_p^3\epsilon^{\frac{3}{2}}}\frac{1}{(k_1k_2k_3)^{\frac{3}{2}}}\left[\mathcal{H}_1^{0,\mathrm{dr}}(\mathbf{k}_1,\mathbf{k}_2)-\mathcal{H}_1^{0,\mathrm{dr}*}(-\mathbf{k}_1,-\mathbf{k}_2)\right] \\ &= \frac{H^4}{4m_p^4\epsilon(2\pi)^{3/2}}\frac{1}{k_1k_2k_3}\sum_{i=0}^{3}\frac{1}{k_i^2k_t}\cos(\tau_ik_t), \quad \text{(C.16)}\end{aligned}$$

where we used the expression (C.6) for $\mathcal{H}_1^{0,\mathrm{dr}}$.

We can similarly evaluate the products of four field operators that appear in Eq. (C.13). We

find that $\langle\hat{\mathcal{R}}(0,\mathbf{k}_1)\hat{\mathcal{R}}(0,\mathbf{k}_2)\hat{\mathcal{R}}(0,\mathbf{p})\hat{\mathcal{R}}(0,\mathbf{k}_3-\mathbf{p})\rangle$ evaluates to

$$\begin{aligned}
&\delta(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\frac{H^4}{16m_p^4\epsilon^2}\frac{1}{(k_1k_2p|\mathbf{k}_3-\mathbf{p}|)^{\frac{3}{2}}}\Bigg[\mathcal{G}_1(\mathbf{k}_1,\mathbf{k}_2,\mathbf{p})-\mathcal{G}_2(\mathbf{k}_3-\mathbf{p},\mathbf{k}_1,\mathbf{k}_2)-\mathcal{G}_2(\mathbf{p},\mathbf{k}_1,\mathbf{k}_2)\\
&-\mathcal{G}_2(\mathbf{k}_2,\mathbf{k}_1,\mathbf{p})-\mathcal{G}_2^*(-\mathbf{k}_1,-\mathbf{k}_2,-\mathbf{p})-\mathcal{G}_2(\mathbf{k}_1,\mathbf{k}_2,\mathbf{p})-\mathcal{G}_2^*(-\mathbf{k}_2,-\mathbf{k}_1,-\mathbf{p})-\mathcal{G}_2^*(-\mathbf{p},-\mathbf{k}_1,-\mathbf{k}_2)\\
&-\mathcal{G}_2^*(\mathbf{p}-\mathbf{k}_3,-\mathbf{k}_1,-\mathbf{k}_2)+\mathcal{G}_3(\mathbf{p},\mathbf{k}_3-\mathbf{p},\mathbf{k}_1)+\mathcal{G}_3(\mathbf{k}_2,\mathbf{k}_3-\mathbf{p},\mathbf{k}_1)+\mathcal{G}_3(\mathbf{k}_2,\mathbf{p},\mathbf{k}_1)+\mathcal{G}_3(\mathbf{k}_1,\mathbf{k}_3-\mathbf{p},\mathbf{k}_2)\\
&+\mathcal{G}_3(\mathbf{k}_1,\mathbf{p},\mathbf{k}_2)+\mathcal{G}_3(\mathbf{k}_1,\mathbf{k}_2,\mathbf{p})+\mathcal{G}_1^*(-\mathbf{k}_1,-\mathbf{k}_2,-\mathbf{p})\\
&+\delta(\mathbf{p}+\mathbf{k}_1)\Big\{-2\mathrm{Re}\mathcal{F}_1(\mathbf{k}_1)-2\mathrm{Re}\mathcal{F}_1(\mathbf{k}_2)+\mathcal{F}_2(-\mathbf{k}_2)+\mathcal{F}_2(-\mathbf{k}_1)+\mathcal{F}_2(\mathbf{k}_2)+\mathcal{F}_2(\mathbf{k}_1)\Big\}\\
&+\delta(\mathbf{p}+\mathbf{k}_2)\Big\{-2\mathrm{Re}\mathcal{F}_1(\mathbf{k}_1)-2\mathrm{Re}\mathcal{F}_1(\mathbf{k}_2)+\mathcal{F}_2(-\mathbf{k}_2)+\mathcal{F}_2(-\mathbf{k}_1)+\mathcal{F}_2(\mathbf{k}_2)+\mathcal{F}_2(\mathbf{k}_1)\Big\}\\
&+\delta(\mathbf{p}+\mathbf{k}_1)+\delta(\mathbf{p}+\mathbf{k}_2)\Bigg].
\end{aligned}\tag{C.17}$$

We now substitute this expression together with the expression (C.14) back into Eq. (C.13), and then into Eq. (3.6.2), and then simplify using the decompositions (3.4.8). This yields the following pieces of our expression (3.6.3) for the bispectrum: a part of the vacuum term $\mathcal{B}_{0,\mathrm{dr}}$, the term $\mathcal{B}_{III}$, the pieces $\mathcal{A}_{II}$ and $\tilde{\mathcal{A}}_{II}$ of the term $\mathcal{B}_{II}$, and the pieces $\mathcal{A}_{IV}$ and $\tilde{\mathcal{A}}_{IV}$ of the term $\mathcal{B}_{IV}$. It also generates the additional term (C.16) which will be canceled by a term obtained in the next subsection.

Commutator contribution

We now evaluate the second term in Eq. (3.6.2). For this term the effect of the field redefinition is of subleading order and can be neglected. Using the interaction Hamiltonian (3.3.2), the mode expansion (3.3.4), and the asymptotic form Eq. (2.5.2) of the mode functions we find that this term evaluates to

$$\begin{aligned}
&-\int_{\tau_i}^0 d\tau'\frac{iH^4}{16\epsilon m_p^4}\Big(\frac{1}{k_1k_2k_3}\Big)^{\frac{3}{2}}\int\frac{d^3p\;d^3k'}{(2\pi)^{3/2}}\frac{\sqrt{pk'|\mathbf{p}-\mathbf{k}'|}}{p^2}\Big\langle\Big[\Big(\hat{a}_{\mathbf{k}_1}-\hat{a}^\dagger_{-\mathbf{k}_1}\Big)\Big(\hat{a}_{\mathbf{k}_2}-\hat{a}^\dagger_{-\mathbf{k}_2}\Big)\\
&\times\Big(\hat{a}_{\mathbf{k}_3}-\hat{a}^\dagger_{-\mathbf{k}_3}\Big),\Big(\hat{a}_{\mathbf{k}'}e^{-i\tau'k'}-\hat{a}^\dagger_{-\mathbf{k}'}e^{i\tau'k'}\Big)\Big(\hat{a}_{\mathbf{p}-\mathbf{k}'}e^{-i\tau'|\mathbf{p}-\mathbf{k}'|}-\hat{a}^\dagger_{\mathbf{k}'-\mathbf{p}}e^{i\tau'|\mathbf{p}-\mathbf{k}'|}\Big)\\
&\times\Big(\hat{a}_{-\mathbf{p}}e^{-i\tau'p}-\hat{a}^\dagger_{\mathbf{p}}e^{i\tau'p}\Big)\Big]\Big\rangle.
\end{aligned}\tag{C.18}$$

Using the definitions (3.4.1) and (3.4.3) of the free functions $\mathcal{F}_i$ and $\mathcal{G}_i$, the expression inside the expectation value brackets $\langle \ldots \rangle$ in Eq. (C.18) evaluates to

$$
\begin{aligned}
&\delta(\mathbf{p}-\mathbf{k}_3)\delta(\mathbf{k}_1+\mathbf{k}_2+\mathbf{p})\Bigg[(e^{-i\tau'(k'+|\mathbf{p}-\mathbf{k}'|+p)}-e^{-i\tau'(k'+|\mathbf{p}-\mathbf{k}'|-p)})\big[\mathcal{G}_1(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}')+\mathcal{G}_3(\mathbf{k}_2,\mathbf{k}_1,\mathbf{k}')\big]\\
&+(e^{-i\tau'(k'-|\mathbf{p}-\mathbf{k}'|-p)}-e^{-i\tau'(k'-|\mathbf{p}-\mathbf{k}'|+p)})\big[\mathcal{G}_2(\mathbf{p}-\mathbf{k}',\mathbf{k}_1,\mathbf{k}_2)+\mathcal{G}_2^*(-\mathbf{k}',-\mathbf{k}_1,-\mathbf{k}_2)\big]\\
&+(e^{-i\tau'(-k'+|\mathbf{p}-\mathbf{k}'|-p)}-e^{-i\tau'(-k'+|\mathbf{p}-\mathbf{k}'|+p)})\big[\mathcal{G}_2(\mathbf{k}',\mathbf{k}_1,\mathbf{k}_2)+\mathcal{G}_2^*(\mathbf{k}'-\mathbf{p},-\mathbf{k}_1,-\mathbf{k}_2)\big]\\
&+(e^{-i\tau'(-k'-|\mathbf{p}-\mathbf{k}'|+p)}-e^{-i\tau'(-k'-|\mathbf{p}-\mathbf{k}'|-p)})\big[\mathcal{G}_3(\mathbf{p}-\mathbf{k}',\mathbf{k}',\mathbf{k}_1)+\mathcal{G}_1^*(-\mathbf{k}_1,-\mathbf{k}_2,-\mathbf{p})\big]\\
&+(e^{-i\tau'(k'+|\mathbf{p}-\mathbf{k}'|-p)}-e^{-i\tau'(k'+|\mathbf{p}-\mathbf{k}'|+p)})\big[\mathcal{G}_2(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}')+\mathcal{G}_2(\mathbf{k}_2,\mathbf{k}_1,\mathbf{k}')\big]\\
&+(e^{-i\tau'(k'-|\mathbf{p}-\mathbf{k}'|+p)}-e^{-i\tau'(k'-|\mathbf{p}-\mathbf{k}'|-p)})\big[\mathcal{G}_3(\mathbf{p}-\mathbf{k}',\mathbf{k}_1,\mathbf{k}_2)+\mathcal{G}_3(\mathbf{p}-\mathbf{k}',\mathbf{k}_2,\mathbf{k}_1)\big]\\
&+(e^{-i\tau'(-k'+|\mathbf{p}-\mathbf{k}'|+p)}-e^{-i\tau'(-k'+|\mathbf{p}-\mathbf{k}'|-p)})\big[\mathcal{G}_3(\mathbf{k}',\mathbf{k}_1,\mathbf{k}_2)+\mathcal{G}_3(\mathbf{k}',\mathbf{k}_2,\mathbf{k}_1)\big]\\
&+(e^{-i\tau'(-k'-|\mathbf{p}-\mathbf{k}'|-p)}-e^{-i\tau'(-k'-|\mathbf{p}-\mathbf{k}'|+p)})\big[\mathcal{G}_2^*(-\mathbf{k}_1,-\mathbf{k}_2,-\mathbf{k}')+\mathcal{G}_2^*(-\mathbf{k}_2,-\mathbf{k}_1,-\mathbf{k}')\big]\Bigg]\\
&+\delta(\mathbf{p}-\mathbf{k}_3)\big[\delta(\mathbf{k}'+\mathbf{k}_1)\delta(\mathbf{k}_2+\mathbf{p}-\mathbf{k}')+\delta(\mathbf{k}_2+\mathbf{k}')\delta(\mathbf{k}_1+\mathbf{p}-\mathbf{k}')\big]\\
&\times\Bigg[(e^{-i\tau'(-k'-|\mathbf{p}-\mathbf{k}'|+p)}-e^{-i\tau'(k'+|\mathbf{p}-\mathbf{k}'|+p)})\mathcal{F}_1(-\mathbf{p})+(e^{-i\tau'(k'+|\mathbf{p}-\mathbf{k}'|-p)}-e^{-i\tau'(-k'-|\mathbf{p}-\mathbf{k}'|-p)})\mathcal{F}_2(-\mathbf{p})\\
&+(e^{-i\tau'(k'+|\mathbf{p}-\mathbf{k}'|+p)}-e^{-i\tau'(-k'-|\mathbf{p}-\mathbf{k}'|+p)})\mathcal{F}_2^*(\mathbf{p})+(e^{-i\tau'(-k'-|\mathbf{p}-\mathbf{k}'|-p)}-e^{-i\tau'(k'+|\mathbf{p}-\mathbf{k}'|-p)})\mathcal{F}_1^*(\mathbf{p})\Bigg]\\
&-2i\delta(\mathbf{p}-\mathbf{k}_3)\big[\delta(\mathbf{k}'+\mathbf{k}_1)\delta(\mathbf{k}_2+\mathbf{p}-\mathbf{k}')+\delta(\mathbf{k}_2+\mathbf{k}')\delta(\mathbf{k}_1+\mathbf{p}-\mathbf{k}')\big]\sin\tau'(k'+|\mathbf{p}-\mathbf{k}|+p)\\
&+(\mathbf{p}\leftrightarrow-\mathbf{k}')+(\mathbf{p}\leftrightarrow\mathbf{k}'-\mathbf{p}).\\
&+\text{cyclic perms.}
\end{aligned}
\tag{C.19}
$$

We now substitute the expression (C.19) into Eq. (C.18) and then into Eq. (3.6.2), and then simplify using the decompositions (3.4.8). This yields the remaining pieces of our expression (3.6.3) for the bispectrum: the remaining part of the vacuum term $\mathcal{B}_{0,\mathrm{dr}}$, the piece $\hat{\mathcal{A}}_{II}$ of the term $\mathcal{B}_{II}$, and the piece $\hat{\mathcal{A}}_{IV}$ of the term $\mathcal{B}_{IV}$. It also generates an additional term, from the time integral of the third last line of Eq. (C.19), which cancels the contribution (C.16).

Appendix D

Appendices to chapter 4

Perturbing the action for a perfect isentropic fluid

As mentioned in Sec. 4.3.I, there are various formulations of the action principle for a perfect isentropic fluid, most of which are reviewed in [120]. The formulation that we have used is the one that appears in [40], which in turn is based on an older formulation given in [140]. This fluid action is given by

$$S_{\rm fluid} = -\int \sqrt{-g}\; d^4x\; \rho_{\rm f}(|\vec{J}|), \tag{D.1}$$

where $\rho_{\rm f}$ is the fluid's energy density and $\vec{J} = n\vec{u}$ is the fluid's number density current that satisfies $\nabla_a J^a = 0$. Note that $|\vec{J}| = n$ by virtue of $\vec{u}$ being a unit 4-vector. The fluid's energy density is in turn defined as [140]

$$\rho_{\rm f} = n\left[m - \frac{p_{\rm f}}{n} + \int \frac{dp_{\rm f}}{dn}\frac{dn}{n}\right], \tag{D.2}$$

where m is the mass of each fluid particle and $p_{\rm f}$ is the fluid's pressure. Note that the definition of $p_{\rm f}$ given in Eq. (4.3.4) is consistent with Eq. (D.2).

In order to expand the action (D.1) order by order in perturbations, one should first expand $\rho_{\rm f}$ in the number density perturbations δn. Up to second order in perturbations, we write

$$\rho_{\rm f} = \bar{\rho}_{\rm f} + \delta_1\rho_{\rm f} + \delta_2\rho_{\rm f}, \tag{D.3}$$

where

$$\delta_1\rho_{\rm f} = \frac{\bar{\rho}_{\rm f} + \bar{p}_{\rm f}}{\bar{n}}\delta_1 n, \qquad \delta_2\rho_{\rm f} = \frac{1}{2}c_s^2(\bar{\rho}_{\rm f} + \bar{p}_{\rm f})\frac{(\delta_1 n)^2}{\bar{n}^2} + \frac{\bar{\rho}_{\rm f} + \bar{p}_{\rm f}}{\bar{n}}\delta_2 n, \tag{D.4}$$

where $c_s^2 \equiv \bar{p}_{\rm f}/\bar{\rho}_{\rm f}$ and we used Eq. (D.2). One then needs to relate the number density perturbations to the perturbations ξ^a of the fluid's flow line and the metric perturbations via Eq. (4.3.2). The

final result for the first order and second order perturbations to the number density is [1] [40]

$$\begin{aligned}
&\delta_1 n = \bar{n}[3\Psi - \partial^2 E - \partial^2 \xi], \\
&\delta_2 n = \bar{n}\Big[- \partial B.\partial \xi' - \frac{1}{2}(\partial \xi')^2 - \frac{1}{2}(\partial B)^2 + \frac{15}{2}\Psi^2 + \frac{1}{2}(\partial^2 E)^2 + \partial_i \partial_j E \partial^i \partial^j E \\
&-5\Psi \partial^2 E - (3\Psi - \partial^2 E)\partial^2 \xi + \partial_i \Big(\xi' \partial^i \xi' + \frac{1}{2}\partial^i \partial_j \xi \partial^j \xi + \frac{1}{2}\partial^i \xi \partial^2 \xi \Big) \Big],
\end{aligned} \tag{D.5}$$

where $\partial^2 \equiv \delta^{ij}\partial_i \partial_j$ and $'$ denotes differentiation with respect to the conformal time τ. Eqs. (D.4) and (D.5) are then used to compute the second order scalar perturbations to the fluid action (D.1). The final result appears in Eq. (4.3.7).

Quantization scheme for the scalar perturbations

In this section we show how to properly quantize the scalar perturbations of our fluid-inflaton model. Our starting point is the scalar perturbation action

$$\delta_2 S = \int d\tau \int d^3x \, \frac{1}{2}\Big[\mathbf{\Pi}'^{\mathrm{T}} \mathbf{\Pi}' - \partial_i \mathbf{\Pi}^{\mathrm{T}} \boldsymbol{c}_s^2 \partial^i \mathbf{\Pi} + \mathbf{\Pi}^{\mathrm{T}} \mathbf{\Gamma} \mathbf{\Pi} + \mathbf{\Pi}'^{\mathrm{T}} \boldsymbol{\lambda} \mathbf{\Pi} \Big], \tag{D.6}$$

where the matrices are defined as

$$\begin{aligned}
&\mathbf{\Pi} \equiv \begin{pmatrix} \pi \\ \nu \end{pmatrix}, \qquad \boldsymbol{c}_s^2 \equiv \begin{pmatrix} 1 & 0 \\ 0 & c_s^2 \end{pmatrix}, \qquad \boldsymbol{\lambda} \equiv \begin{pmatrix} 0 & \sqrt{2}\frac{B}{\zeta} \\ -\sqrt{2}\frac{B}{\zeta} & 0 \end{pmatrix}, \\
&\mathbf{\Gamma} \equiv \begin{pmatrix} \frac{\zeta''}{\zeta} + \frac{2}{\zeta^2}\big[4z^2 A - (zB)'\big] & 4\sqrt{2}A\frac{z}{\zeta} - \frac{\sqrt{2}}{2}B\frac{\zeta'}{\zeta^2} - \frac{\sqrt{2}}{2}\frac{B'}{\zeta} \\ 4\sqrt{2}A\frac{z}{\zeta} - \frac{\sqrt{2}}{2}B\frac{\zeta'}{\zeta^2} - \frac{\sqrt{2}}{2}\frac{B'}{\zeta} & \frac{z''}{z} + 4A \end{pmatrix}.
\end{aligned} \tag{D.7}$$

We have suppressed dependences on τ for all background functions above. The Euler-Lagrange equations of motion for this action are

$$\mathbf{\Pi}'' - \big[\boldsymbol{c}_s^2 \partial^2 + \mathbf{\Gamma} - \frac{1}{2}\boldsymbol{\lambda}'\big]\mathbf{\Pi} + \boldsymbol{\lambda}\mathbf{\Pi}' = 0. \tag{D.8}$$

Next we compute the Hamiltonian for this system. First note that the canonical momentum $\mathbf{P}$ in this model is

$$\mathbf{P} \equiv \frac{\delta \mathrm{L}_{\mathrm{scalar}}}{\delta \mathbf{\Pi}'^{\mathrm{T}}} = \mathbf{\Pi}' + \frac{1}{2}\boldsymbol{\lambda}\mathbf{\Pi}, \tag{D.9}$$

[1] Considering scalar perturbations only. Recall that $\xi_i = \partial_i \xi$.

where L is the Lagrangian appearing in the action (D.6). Employing the Legendre transformation, the Hamiltonian becomes

$$\begin{aligned}\mathrm{H}_{\mathrm{scalar}} &= \int dx^3 \ \left[\mathbf{P}^{\mathrm{T}}\boldsymbol{\Pi}' - \mathcal{L}\right] \\ &= \int dx^3 \ \frac{1}{2}\Big[\mathbf{P}^{\mathrm{T}}\mathbf{P} + \partial_i\boldsymbol{\Pi}^{\mathrm{T}}\boldsymbol{c}_s^2\partial^i\boldsymbol{\Pi} - \mathbf{P}^{\mathrm{T}}\boldsymbol{\lambda}\boldsymbol{\Pi} + \boldsymbol{\Pi}^{\mathrm{T}}\big(\frac{1}{4}\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{\Gamma}\big)\boldsymbol{\Pi}\Big],\end{aligned} \tag{D.10}$$

which in Fourier space reduces to

$$\mathrm{H}_{\mathrm{scalar}} = \int dk^3 \ \frac{1}{2}\big[\mathbf{P}^{\mathrm{T}}(\tau,\boldsymbol{k})\mathbf{P}(\tau,-\boldsymbol{k}) + \boldsymbol{\Pi}^{\mathrm{T}}(\tau,\boldsymbol{k})\boldsymbol{\Sigma}\boldsymbol{\Pi}(\tau,-\boldsymbol{k}) - \mathbf{P}^{\mathrm{T}}(\tau,\boldsymbol{k})\boldsymbol{\lambda}\boldsymbol{\Pi}(\tau,-\boldsymbol{k})\big]. \tag{D.11}$$

Here we define

$$\boldsymbol{\Sigma} \equiv \boldsymbol{c}_s^2 k^2 - \boldsymbol{\Gamma} + \frac{1}{4}\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\lambda}. \tag{D.12}$$

The Hamilton's equations of motion for each mode are

$$\begin{aligned}\mathbf{P}'(\tau,\boldsymbol{k}) &= -\frac{\delta \mathrm{H}}{\delta \boldsymbol{\Pi}^{\mathrm{T}}(\tau,-\boldsymbol{k})} = -\boldsymbol{\Sigma}\boldsymbol{\Pi}(\tau,\boldsymbol{k}) - \frac{1}{2}\boldsymbol{\lambda}\mathbf{P}(\tau,\boldsymbol{k}), \\ \boldsymbol{\Pi}'(\tau,\boldsymbol{k}) &= \frac{\delta \mathrm{H}}{\delta \mathbf{P}^{\mathrm{T}}(\tau,-\boldsymbol{k})} = \mathbf{P}(\tau,\boldsymbol{k}) - \frac{1}{2}\boldsymbol{\lambda}\boldsymbol{\Pi}(\tau,\boldsymbol{k}).\end{aligned} \tag{D.13}$$

The Hamiltonian (D.11) is the starting point for quantization. After promoting $\mathbf{P}(\tau,\boldsymbol{k})$ and $\boldsymbol{\Pi}(\tau,\boldsymbol{k})$ to quantum operators, we impose the following equal-time quantization relations,

$$\begin{aligned}&\big[\hat{\Pi}_i(\tau,\boldsymbol{k}),\hat{\Pi}_j(\tau,\boldsymbol{k}')\big] = 0, \qquad\qquad \big[\hat{\mathrm{P}}_i(\tau,\boldsymbol{k}),\hat{\mathrm{P}}_j(\tau,\boldsymbol{k}')\big] = 0, \\ &\big[\hat{\Pi}_i(\tau,\boldsymbol{k}),\hat{\mathrm{P}}_j(\tau,\boldsymbol{k}')\big] = i\delta_{ij}\delta(\boldsymbol{k}+\boldsymbol{k}').\end{aligned} \tag{D.14}$$

Quantum operators such as $\hat{\mathbf{P}}$ and $\hat{\boldsymbol{\Pi}}$ are often conveniently expanded in terms of the creation and annihilation operators and the mode functions that are solutions to the equations of motion (D.8) or (D.13). Note that the space of solutions of (D.8) is four-dimensional. Therefore, we can pick any two linearly independent solutions $\boldsymbol{\Pi}_A$ with $A = 1, 2$ together with their complex conjugates to span the space of (Lagrangian) solutions. This leads to the expansion of the quantum operator $\hat{\boldsymbol{\Pi}}$ as

$$\hat{\boldsymbol{\Pi}}(\tau,\boldsymbol{k}) = \sum_{A=1,2} \boldsymbol{\Pi}_A(\tau,k)\hat{a}_{A,\boldsymbol{k}} + \boldsymbol{\Pi}_A^*(\tau,k)\hat{a}^\dagger_{A,-\boldsymbol{k}}, \tag{D.15}$$

where $\hat{a}$ and $\hat{a}^\dagger$ are the creation and annihilation operators, and summation over the index A is assumed.

It remains to give a proper definition for the creation and annihilation operators $\hat{a}$ and $\hat{a}^\dagger$. To do this, we define the inner product on the space of solutions of (D.8) as

$$\langle \mathbf{U}, \mathbf{V} \rangle \equiv i\big[\mathbf{U}^+\mathbf{V}' - \mathbf{U}'^+\mathbf{V} + \mathbf{U}^+\boldsymbol{\lambda}\mathbf{V}\big], \tag{D.16}$$

where + denotes the Hermitian conjugate of a matrix. Using the equations of motion (D.8) one can show this inner product to be conserved, i.e. $\partial_\tau\langle \boldsymbol{U}, \boldsymbol{V}\rangle = 0$, for all solutions, making it an appropriate definition of inner product on the space of Lagrangian solutions. We then require the solutions $\boldsymbol{\Pi}_A(\tau, k)$ appearing in (D.15) to be normalized in the sense of

$$\big\langle \boldsymbol{\Pi}_A(\tau,k), \boldsymbol{\Pi}_B(\tau,k)\big\rangle = \delta_{AB}. \tag{D.17}$$

Conventionally we pick

$$\boldsymbol{\Pi}_1(k,\tau) = \begin{pmatrix} \frac{e^{-ik\tau}}{\sqrt{2k}} \\ 0 \end{pmatrix}, \qquad \boldsymbol{\Pi}_2(k,\tau) = \begin{pmatrix} 0 \\ \frac{e^{-ic_sk\tau}}{\sqrt{2c_sk}} \end{pmatrix}, \tag{D.18}$$

as the solutions in the deep subhorizon limit. These two solutions are linearly independent, normalized in the sense of (D.17), and together with their complex conjugates form a basis for the space of solutions of (D.8) at early times, when the fluid and inflaton scalar perturbations are both decoupled and adiabatic. These solutions correspond to the fluid and inflaton perturbation modes being in some vacuum state at early times [40].

Finally we define the creation and annihilation operators as

$$\hat{a}_{A,\boldsymbol{k}} \equiv \big\langle \boldsymbol{\Pi}_A(\tau,k), \hat{\boldsymbol{\Pi}}(\tau,\boldsymbol{k})\big\rangle, \qquad \hat{a}^\dagger_{A,\boldsymbol{k}} \equiv -\big\langle \boldsymbol{\Pi}^*_A(\tau,k), \hat{\boldsymbol{\Pi}}(\tau,-\boldsymbol{k})\big\rangle. \tag{D.19}$$

It is easy to check that

$$[\hat{a}_{A,\boldsymbol{k}}, \hat{a}^\dagger_{B,\boldsymbol{k}'}] = \delta_{AB}\delta^3(\boldsymbol{k}-\boldsymbol{k}'), \qquad [\hat{a}_{A,\boldsymbol{k}}, \hat{a}_{B,\boldsymbol{k}'}] = [\hat{a}^\dagger_{A,\boldsymbol{k}}, \hat{a}^\dagger_{B,\boldsymbol{k}'}] = 0. \tag{D.20}$$

The operators $\hat{a}_{A\boldsymbol{k}}$ annihilate the ground states $|\Omega_A\rangle$ defined as

$$|\Omega_A\rangle \equiv \bigotimes_{\boldsymbol{k}} |0_{A\boldsymbol{k}}\rangle, \tag{D.21}$$

where $|0_{A\boldsymbol{k}}\rangle$ denotes the vacuum state corresponding to the mode with momentum $\boldsymbol{k}$. Existence of a natural choice of vacuum at early times is the consequence of the adiabaticity of fluid and inflaton perturbation modes discussed in Sec. 4.4. Indeed the vacuum $|\Omega_{1(2)}\rangle$ is the analog of the

Bunch-Davies state for the inflaton (fluid) perturbations, corresponding to mode functions with initial conditions (D.18). The vacuum state for the scalar perturbations is then defined as

$$|\mathcal{I}\rangle \equiv \bigotimes_A |\Omega_A\rangle. \tag{D.22}$$

Part II

Entropy-Mass-Timescale Relation In Gravitational Physics

Chapter 5

Introduction To Black Hole Thermodynamics

In this chapter we will begin by reviewing the basic principles of black hole thermodynamics with an emphasis on the concept of black hole entropy. We will discuss two well known upper bounds for entropies of self gravitating objects in general relativity. We will then introduce a conjectured entropy-mass-timescale relationship in general relativity and discuss its validity in the weak field regime. Throughout this part, we will be working in the units of $k_B = c = G = \hbar = 1$ unless stated otherwise.

5.1 Basics of black hole thermodynamics

That black holes are thermal objects was firmly proven following the discovery of Hawking radiation [141, 142]. Hawking's computations showed that once the backreaction effects of quantum fields on an otherwise fixed stationary black hole background are incorporated, quanta of energy are carried off to future null infinity. It was further shown that these quanta are thermally distributed, with a temperature that is proportional to the black hole surface gravity. Prior to these quantum calculations, however, speculations regarding the thermal nature of black holes had arisen as a result of the laws of black hole mechanics, which we now review.

I Laws of black hole mechanics

The laws of black hole mechanics were formulated in the early 1970s by Bardeen, Carter, and Hawking [143]. These are a set of four laws, all of which have been rigorously proven under fairly

reasonable assumptions by various methods.

The zeroth law of black hole mechanics states that the surface gravity of a black hole event horizon is constant across the event horizon. Indeed, assuming that black hole event horizon is a Killing horizon [1] with $\vec{\chi}$ as its normal Killing field, the surface gravity

$$\kappa^2 \equiv -\frac{1}{2}\nabla_a \chi_b \nabla^a \chi^b \tag{5.1.1}$$

is readily seen to be conserved along each null generator of the event horizon. If the Killing horizon is further assumed to be a bifurcate Killing horizon, then a simple proof can be given that κ is conserved on the entire event horizon. Absent a bifurcation surface, it can still be proven that κ is constant on the event horizon as long as the matter content satisfies the dominant energy condition [2] [143].

To state the first law of black hole mechanics, let us restrict attention to asymptotically flat, stationary, and axisymmetric black holes [3]. The first law then relates the infinitesimal change δM in the ADM mass of a black hole to the infinitesimal change δA in the area of its event horizon and the infinitesimal change δJ in its asymptotic angular momentum via

$$\delta M = \Omega \delta J + \frac{\kappa}{8\pi}\delta A, \tag{5.1.2}$$

where Ω is the angular velocity of the black hole as measured by observers at spacelike or null infinity [4]. A number of derivations have been provided for Eq. (5.1.2). Aside from the original derivation given in [143], a significantly simpler and yet stronger derivation was provided using the Hamiltonian formulation of general relativity [146][5] [6].

[1] This has been proven for asymptotically flat stationary black holes with "suitable" matter sources [144].

[2] i.e. the local energy density is larger than the absolute value of the spatial eigenvalues of the stress energy tensor as measured by all timelike observers. See [1] for a definition of different energy conditions.

[3] Generalizations to charged black holes can be found in many references, e.g. [145].

[4] Recall that J vanishes for static black holes. For non-static black holes, the Killing field $\vec{\chi} = \vec{\partial}_t + \Omega\vec{\partial}_\phi$ is normal to the event horizon, where $\vec{\partial}_t$ is timelike at large distances from the black hole and it generates asymptotic time-translations. $\vec{\partial}_\phi$ is spacelike everywhere and it has close orbits that are usually normalized to be 2π.

[5] Unlike the original derivation which assumes that the final state black hole is stationary and axisymmetric, the derivation given in [146] proves Eq. (5.1.2) for arbitrary regular asymptotically flat perturbations to black hole as long as the perturbation functions satisfy the linearized Einstein constraint equations.

[6] A simple but less rigorous proof for the first law can be given as follows [53]: Consider perturbing a stationary axisymmetric black hole by throwing some matter distribution that is described by a stress energy tensor $\mathcal{T}_{ab}$ across its event horizon. Suppose further that the black hole settles down to a stationary axisymmetric state once the matter crosses the event horizon. The changes to the mass and angular momentum

The second law of black hole mechanics includes as part of it the black hole area theorem. This theorem asserts that for an asymptotically predictable spacetime [7] that satisfies the null energy condition [8], the area of the event horizon does not decrease with time [9]. More generally, the second law requires that for two coalescing black holes with horizon areas A_1 and A_2, the final state black hole has an event horizon area $A > A_1 + A_2$.

Finally, we should mention that the third law of black hole mechanics has been expressed in a number of different ways. Most commonly, it is expressed as the impossibility of reducing the black hole surface gravity κ to zero by a finite number of steps. A more precise formulation for this law has been provided and proven in [147].

II Black hole entropy

Hawking's subsequent quantum calculations revealed that observers at null infinity receive thermal radiation with a temperature

$$T = \frac{\kappa}{2\pi}. \tag{5.1.3}$$

For a Schwarzschild black hole, we have $\kappa = 1/4M$. Thus, the above temperature can be written as (after restoring the fundamental constants)[53]

$$T = \frac{\hbar c^3}{8\pi k_B G M} = 6 \times 10^{-8} \left(\frac{M_\odot}{M}\right)^\circ K, \tag{5.1.4}$$

where $M_\odot$ is the mass of the Sun. As can be seen from this relation, the smaller the black hole mass is the hotter the temperature of thermal radiation received at null infinity becomes.

This revelation grants the first law of black hole mechanics a robust thermodynamical meaning. It states that the changes δM in the overall gravitational energy of an isolated black hole must be

of the black hole are given by

$$\delta M = \int_{\rm EH} \mathcal{T}_{ab} (\partial_t)^a k^b, \qquad \delta J = -\int_{\rm EH} \mathcal{T}_{ab} (\partial_\phi)^a k^b,$$

where EH stands for event horizon and $\vec{k} \equiv \vec{\chi}/(\kappa\lambda)$ with λ being the affine parameter for the null geodesics that generate the event horizon. Eq. (5.1.2) then follows without much difficulty from integrating $\kappa\lambda\theta_{,\lambda}$ with $\theta \equiv \nabla_a k^a$ along the event horizon and using the Rauchaudhuri equation with $\mathcal{T}_{ab}$ as its matter source.

[7] An asymptotically predictable spacetime M is an asymptotically flat spacetime for which there exists a Cauchy surface Σ with a compact interior, the domain of dependence of which includes the event horizon [1].

[8] i.e. $T_{ab} k^a k^b \geq 0$ for all null $\vec{k}$.

[9] A proof for this theorem can be found in [144] or [53].

related to the product of its temperature and the changes in its entropy. Using Eqs. (5.1.2) and (5.1.3), one can then define black hole entropy as

$$S = \frac{A}{4}. \tag{5.1.5}$$

This is the celebrated Bekenstein-Hawking formula for black hole entropy in general relativity. For a Schwarzschild black hole, this formula becomes (after restoring the fundamental constants)

$$S = \frac{4\pi k_B G}{c\hbar} M^2. \tag{5.1.6}$$

Black hole entropy can be quite large. For a stellar mass black hole $S/k_B \sim 10^{77}$, which is about twenty orders of magnitude larger than the entropy of the Sun.

III Black hole entropy in more general theories of gravity

The derivation of Eq. (5.1.2) makes use of the Einstein field equations. In different theories of gravity with black hole solutions, the definition given in Eq. (5.1.5) for black hole entropy is no longer valid. Nevertheless, a rigorous prescription [10] for computing black hole entropy for any diffeomorphism invariant theory derivable from a Lagrangian has been given in terms of the integral of the Noether charge associated with the Killing vector field $\vec{\chi}$ that is normal to the event horizon [150, 151]. For a Lagrangian of the form $L(g_{ab}, R_{abcd}, \nabla_e R_{abcd}, ...; \psi, \nabla_a \psi, ...)$ with ψ being an arbitrary matter field, the entropy is given by

$$S = -2\pi \int_\sigma E^{abcd} n_{ab} n_{cd}, \tag{5.1.7}$$

where σ is a section along the event horizon of the black hole [11], n_{ab} is the binormal form to σ, and

$$E^{abcd} \equiv \frac{\delta L}{\delta R_{abcd}}. \tag{5.1.8}$$

For the case of general relativity with $L = R/(16\pi)$, it follows without much difficulty that the entropy given in Eq. (5.1.7) results in $S = A/4$ as expected.

[10] Other methods have been proposed for computing black hole entropy as well. One technique uses a specific redefinition of metric to cast a complicated theory of matter and gravity into a simpler theory in which the method for computing black hole entropy is known [148]. An older approach to black hole entropy computation was given via the Euclidean path integral in [149].

[11] In the original derivation of [150], σ was assumed to be the bifurcation surface for the event horizon, which was required to be a bifurcate Killing horizon. It was later proved in [148] that σ can indeed be an arbitrary surface along the event horizon of the black hole.

IV Making sense of black hole entropy

Ever since the discovery of the thermal nature of black holes, many attempts have been made to provide an interpretation for their entropies in the context of statistical mechanics. Indeed, if black holes correspond to thermal equilibrium states of quantum gravitational degrees of freedom, their entropies should be a measure of the number of distinct quantum states that these degrees of freedom have [12]. It is presently not understood what the quantum gravitational degrees of freedom responsible for black hole entropy are. Nonetheless, several proposal have been suggested that aim to explain the microscopic origin of black hole entropy. We briefly review a number of these proposals below [13]:

- **Formation Degeneracy:** An old idea, going back to Bekenstein, is that black hole entropy corresponds to the logarithm of the number of ways that a black hole with a given set of macroscopic parameters can be formed [141, 153]. The motivation for this proposal is that the number of states of a system is not expected to change if the system undergoes a unitary quantum mechanical evolution. Thus, one expects that the number N of distinct initial data sets that can collapse to form a given black hole must match the number of its internal states, and correspondingly match the Bekenstein-Hawking entropy, $\ln N \sim M^2$, where M is the black hole mass. Following this idea, Thorne and Zurek calculated the number of ways in which a black hole of mass M can be formed from the gradual accretion of a dilute gas of radiation quanta by a small black hole with a mass $m \ll M$ [154]. Their calculations confirmed Bekenstein's hypothesis for this particular scenario of gravitational collapse. We will have more to say about this process in Sec. 5.3.

- **Thermal Entropy of the Exterior Radiation Bath / Entanglement Entropy:** Two essentially identical proposals attribute black hole entropy to the entropy of the thermal bath of radiation as seen by the stationary observers in the exterior region of a black hole [155, 156, 157, 158]. In the entanglement entropy approach, one obtains a density matrix for the quantum field degrees of freedom that are outside of the black hole event horizon by

[12] Seen in the framework of statistical mechanics, the entropy $A/4$ of a black hole corresponds to an ensemble of roughly $e^{A/4}$ black holes, each having the same macroscopic characteristics (i.e. the same ADM mass, etc.) while being in distinct microscopic states.

[13] See [145, 152] for more discussions.

tracing over the quantum states in the interior region. The reduced density matrix for the exterior region has an expected non-vanishing entanglement entropy. This entropy is just the thermodynamic entropy since the natural state for the quantum fields in the exterior region is the thermal state as seen by the stationary observers, with a temperate given by $T/|\vec{\chi}|$ [14]. Therefore, this calculation reduces to that of the entropy of a thermalized gas of radiation in the exterior region of a black hole. As the temperature of the radiation bath diverges at the horizon, so does its entropy density $s \sim T^3$. It turns out that if one only considers the radiation quanta that are a proper distance on the order of Planck length away from the horizon, then one arrives at $S_{\text{bath}} \sim A$. The exact constant of proportionality, however, is cutoff dependent.

This interpretation, though remarkable in reproducing the entropy scaling with area, has been criticized from different perspectives. One main criticism is that no where in this derivation one makes use of the Einstein equations. This is potentially problematic since black hole entropy for more general theories of gravity is no longer the area of the event horizon [15]. Additionally, in the case of entanglement entropy, a similarly divergent value is derived for the reduced density matrix across any null surface. Thus, it is not clear why the event horizon should play a unique role in this calculation.

- **Entropy of Quantum Geometric Degrees of Freedom:** In the context of loop quantum gravity, the number of states associated with certain quantized geometric operators along the event horizon of a black hole has been calculated [159, 160]. The entropy of these states is shown to be proportional to the area of the event horizon, with a constant of proportionality that depends on the specific quantization scheme. Unlike the previous proposal, the degrees of freedom responsible for black hole entropy in this case are particular to the black hole event horizon.

- **Entropy of the Weak Coupling States in String Theory:** In the weak coupling limit of string theory, certain quantum states can be identified with black hole solutions. The entropy of these states can be computed using the standard procedures of statistical mechanics. There are several examples where the entropy of the weak coupling states corresponding to certain

[14] Recall that $\vec{\chi}$ is the Killing vector field that is normal to the event horizon. For a Schwarzschild black hole, $\vec{\chi} = \vec{\partial}_t$.

[15] See Eq. (5.1.7).

types of black holes are exactly the same as the Bekenstein-Hawking entropy. See [161] for one of the first string theory calculations of black hole entropy.

5.2 The Bekenstein-Hawking entropy bound

It is a remarkable fact that shortly after the formulation of black hole laws of mechanics and prior to the discovery of Hawking radiation, Bekenstein posited that the area of a black hole event horizon is proportional to its entropy. Since the destruction of an object's entropy following its fall into a black hole singularity poses a challenge for the second law of thermodynamics, Bekenstein argued for the existence of the Generalized Second Law of thermodynamics [162]. This law asserts that the sum of the entropy of a black hole together with the entropy of matter outside of its event horizon never decreases with time, i.e.

$$\Delta S_{\text{total}} = \Delta S_{\text{bh}} + \Delta S_{\text{matter}} \geq 0. \tag{5.2.1}$$

This proposal was found to be accurate for many scenarios, including the case of an evaporating black hole for which the entropy of the thermal radiation that escapes to null infinity more than compensates for the decrease in the black hole entropy.

Despite the many examples bearing evidence to the validity of relation (5.2.1), a few examples have been discussed for which the Generalized Second Law could be violated. Most famous among those examples is the case of a box with energy E [16] and entropy S that is quasi-statically lowered down to vicinity of a black hole event horizon and then released. Since the conserved energy of the box is proportional to a redshift factor that vanishes on the horizon, by bringing the box to sufficiently near horizon distances we can arrange for its conserved energy to become arbitrarily small. Thus, with little energy left, the fallen box leaves the area of the black hole event horizon unchanged. It follows from a short calculation that Eq. (5.2.1) is violated in this scenario if the box is lowered to a proper distance D smaller than $S/(2\pi E)$ [17]. Motivated by this example, Bekenstein proposed the following entropy bound for an object [163]

$$S \leq 2\pi ER, \tag{5.2.2}$$

[16] As measured asymptotically.

[17] This apparent violation of Eq. (5.2.1) has since been resolved. See [53] for a discussion.

where E is the locally measured energy of an object[18] and R is the maximum length scale associated with the size of the object. Many examples have been given for the validity of this bound, including several calculations in flat spacetime [164, 165]. It should also be noted that this bound is saturated for black holes. Nevertheless, this bound cannot be generalized to arbitrary curved spacetimes without sufficient symmetry as there may be no preferred or clear definitions for E or R.

I The emergence of the Covariant Entropy Bound

The aforementioned difficulties with defining energy and length in the Bekenstein-Hawking bound (5.2.2) for arbitrary curved spacetimes have been circumvented in a relatively new entropy bound. Motivated by principles of holography, Bousso proposed the so-called Covariant Entropy Bound in 1999 [166] [19]. The Covariant Entropy Bound can be stated as follows: For a closed spacelike 2-surface $\mathcal{M}$, consider the congruence of non-expanding null geodesics that are normal to it. Then the entropy flux S of matter crossing this null hypersurface [20] is bounded above by a quarter of the area of $\mathcal{M}$ [21]. This bound is widely expected to be true. In fact, semi-classical proofs have been provided for it [168, 169]. We will not have much to say about this bound in the remainder of this part.

5.3 Entropy-Mass-Timescale relation for gravitational collapse

It is well known that for black holes that are formed in astrophysical collapses the entropies S of the pre-collapse configurations do not saturate the Bekenstein-Hawking value. When gravity is weak, the maximum entropy that can be contained in a region of size R using an energy E is that of thermal radiation, $S \sim (ER)^{3/4}$, and for the final stages of collapse taking $E \sim M$ and $R \sim M$ gives $S \sim M^{3/2}$, far smaller than the Bekenstein-Hawking value. Although this estimate is based on the assumption of weak gravity which is violated during the late stages of collapse, one expects it to be a robust upper bound for systems which start in a weak gravity regime. The final stages of such

[18] Ignoring negative energy states, such as the Casimir effect.

[19] See [167] for a detailed review.

[20] Commonly known as light sheet.

[21] A straightforward generalization of this bound can be stated as follows: In addition to $\mathcal{M}$ consider another spacelike 2-surface $\mathcal{N}$ that intersects the non-expanding null congruence and is normal to it. Then the entropy flux crossing the null sheet between $\mathcal{M}$ and $\mathcal{N}$ is bounded above by $|A_{\mathcal{N}} - A_{\mathcal{M}}|/4$.

collapses occur on a timescale $\sim M$ [22].

A different formation scenario can saturate[23] the Bekenstein-Hawking entropy. As first pointed out by Zurek and Thorne [154], small black holes of initial mass m much smaller than the final mass M can accrete quanta of energies $E \sim 1/m$ one-by-one, each taking a time of order m. The black hole then grows in an approximate time reverse of the Hawking evaporation process, yielding a final mass M after accreting $\sim M^2$ quanta. The number N of such initial states is then given by $\ln N \sim S \sim M^2$. However, this formation process occurs on a much longer timescale than astrophysical collapse, of order $\tau \sim M^3$ [24].

One can also construct scenarios which are intermediate between the two extremes of dynamical collapse and the Thorne-Zurek scenario. We consider a black hole accreting quanta of a quantum field where each sector l, m for $l \leq l_{\max}$ is an independent $1+1$ dimensional thermal gas with quanta of energy $E \sim l_{\max}/M$ and size τ. The total number of quanta per sector is $\tau E \sim \tau l_{\max}/M$, and the total number of quanta is given by multiplying by the number $l_{\max}^2$ of sectors, giving $S \sim \tau l_{\max}^3/M$. Similarly the total energy per sector is $\tau E^2 \sim \tau l_{\max}^2/M^2$, and the total energy is then $\tau l_{\max}^4/M^2$. Requiring this to be less than M and eliminating $l_{\max}$ yields a relation between entropy S, black hole mass M and formation timescale τ which interpolates between the two extreme examples:

$$S \lesssim M^{5/4}\tau^{1/4}. \tag{5.3.1}$$

The examples discussed above, and specifically the relation (5.3.1), all support a conjecture that generic black hole microstates must be formed by collapse processes on timescales much longer than M. In particular, one might conjecture that the relation (5.3.1) is generic, valid in order of magnitudes for all collapse scenarios with physically realistic matter that begin in weak gravity regime [25]. The purpose of the next chapter is to find evidence against this relation.

[22] See Sec. 6.4.III for a definition of collapse timescale.

[23] In the weak sense that the logarithm of the number of states scales like M^2, not in the stronger sense that it is $M^2/4$ plus subleading corrections.

[24] Also see [170], which shares some similarities with the Thorne Zurek construction.

[25] We conjecture that (5.3.1) is valid for all distributions of matter that satisfy the dominant energy condition. A weaker version of the conjecture in which matter merely satisfies the weak energy condition is probably incorrect, as we argue in the next section. We also require that black holes eventually settle down to stationary and asymptotically flat states.

Chapter 6

Robustness Of The Entropy-Mass-Timescale Relation

6.1 Overview

In this chapter we will examine the robustness of the entropy-mass-timescale relation given in Eq. (5.3.1). There have been a number of indirect attempts to challenge this relation. Most famous among these attempts are the initial data sets first constructed by Sorkin *et.al.* in [171] which later became known as "monsters"[172] [1]. A monster is a time symmetric and asymptotically flat spacelike Cauchy surface that is foliated by nearly marginally trapped 2-spheres. The stress energy tensor is everywhere vacuum except for a compact spherical region which is filled with a thermalized gas of radiation. Monsters are expected to form black holes within a timescale on the order of M [2]. A large entropy on the order of M^2 is achieved for monsters due to the enhanced effect of the intrinsic curvature on the Cauchy surface. Specifically, the large intrinsic curvature allows for a relatively large number of quanta to be packed together in a spatial region with a radius on the order of the Schwarzschild radius [3].

Despite the fact that monsters saturate the Bekenstein-Hawking entropy bound as initial data sets, they may not give rise to valid gravitational collapse processes. In fact, it is far from clear if their full Cauchy development is pathology free. This question is particularly important since a monster, by virtue of being in close vicinity of the future apparent horizon, is expected to be

[1] Also see [171, 173] for the so-called bags of gold initial data sets. Similar to monsters, the full Cauchy development of these initial data sets involve both a white hole and a black hole singularity.

[2] The definition for timescale that is given in Sec. 6.4.III may not be applicable for time symmetric spacetimes. Nonetheless, monsters are fully inside their event horizons within a timescale on the order of M as measured by the freely falling observers on their external boundaries.

[3] In this case a spherical region with an area of $\sim M^2$ can have a volume $\sim M^4$, significantly larger than what it would be if the intrinsic curvature were negligible.

largely within the future event horizon. Thus, one would need to ascertain that the future event horizon is everywhere regular. Moreover, monsters have a time symmetric Cauchy development. This implies that they are endowed with two distinct curvature singularities, a black hole and a white hole. White holes possess classical and quantum instabilities that render them unphysical [174, 175, 176]. These issues suggest that monster initial data sets are not physically realistic or relevant counterexamples to the conjectured relation (5.3.1).

Here we will investigate the global structure of the spacetimes that can arise from monster-like initial conditions. We shall focus on spacetimes with a dust dominated collapsing core. We will review the essential geometric theory in Secs. 6.2, 6.3.I, and 6.3.II. In Sec. 6.3.III we will see how such a spacetime can saturate the Bekenstein-Hawking entropy bound and maintains an everywhere regular future event horizon. We will then discuss a formalism in Sec. 6.4 for modifying the spacetime geometry in the past limit in order to eliminate the white hole singularity. We will show that the spacetimes modified using this prescription satisfy the weak energy condition. Finally, in Sec. 6.4.III we will define the black hole formation timescale for spacetimes without white holes and show that for the particular example that we constructed in Sec. 6.3.III and then modified in Sec. 6.4.II, this timescale is on the order of M^2. Our construction thus invalidates the conjectured entropy-mass-timescale relation by a factor of $M^{1/4}$. However, the constructed spacetime appears to be finely tuned and may not represent a generic scenario of gravitational collapse.

6.2 The hydrodynamic approximation for dust fluids

In this section we briefly review the basics of dust fluids and discuss the conditions for the validity of the hydrodynamic approximation in arbitrary spacetimes.

A pressureless dust fluid is characterized by a four-velocity vector field $\vec{u}$ and a number density n, satisfying $u^a \nabla_a u^b = 0$ and $\nabla_a (n u^a) = 0$. The stress energy tensor for the dust fluid is given by

$$T_{ab} = \rho \; u_a u_b = mn \; u_a u_b, \tag{6.2.1}$$

where $\rho = mn$ is the energy density measured in the comoving frame of the fluid, and m is the particle mass. Here we allow m to depend on the spatial coordinates.

The entropy S of a dust fluid is roughly equal to the number of its constituent particles. Therefore,

one can associate a covariantly conserved entropy current $s^a = nu^a$ to the dust fluid [4], and then compute the total entropy on any achronal spacetime slice Σ by

$$S = \int_\Sigma s^a \, d\Sigma_a = \int_\Sigma \frac{\rho}{m} u^a \, d\Sigma_a = \int_\Sigma \, nu^a \, d\Sigma_a. \tag{6.2.2}$$

The validity of the hydrodynamic approximation underlying the dust fluid model hinges on the following conditions[5]:

$$\begin{aligned} &(8\pi m) \; n^{-1/3} \gtrsim 1, \\ &(8\pi m)^{-1} \gtrsim 1, \\ &\Big(\frac{16\pi}{3}\Big)^{1/3} \frac{m}{|\mathbf{D}m|} \gtrsim n^{-1/3}, \\ &\mathcal{L} \gg n^{-1/3}, \end{aligned} \tag{6.2.3}$$

where $\mathbf{D}$ denotes differentiation with respect to the spatial coordinates, and $\mathcal{L}$ is the radius of curvature defined as $\mathcal{L}^{-4} \equiv R_{abcd}R^{abcd}$ for the spacetime Riemann curvature tensor R_{abcd}. To have a valid model of gravitational collapse, one needs to ensure that these relations are valid everywhere to the past of the event horizon of the ensuing black hole.

The above conditions can be understood as follows. The first condition is the non-degeneracy condition, asserting that the degeneracy pressure can be ignored as long as the dust particles remain non-overlapping. The violation of this condition undermines the pressureless dust description of the fluid [6]. The second condition asserts that the Compton wavelength of each particle must be above the Planck length [7] for quantum gravity effects to be ignored. The third and fourth conditions require the radius of curvature and the length scales over which the particle mass varies to be larger than the intra-particle spacing. This last condition is equivalent to saying that most of the contribution to the total entropy comes from the modes with wavelengths small compared to the spacetime radius of curvature.

[4] the entropy current defined here is generally correct upto some order unity prefactor.

[5] These relations are not precise inequalities. They are only defined upto factors of order unity. The numerical factors are added for convenience when we later discuss saturating these relations.

[6] As the intra-particle spacing approaches the Compton wavelength of particles, the quantum fluctuations in the stress energy tensor of the dust fluid become comparable to the classical stress energy tensor.

[7] The Planck length is defined as $\sqrt{\hbar G/c^3}$, which is equal to unity in our choice of units.

6.3 Dust objects in spherical symmetry

I Lemaître-Tolman-Bondi dust models

Consider a spherically symmetric compact object that is composed of non-interacting dust particles. For an arbitrary density profile [8], the interior geometry of the object is described by the Lemaître-Tolman-Bondi (LTB) metric [177, 178, 179]

$$ds^2 = -dt^2 + \frac{R'(t,r)^2 dr^2}{1 - r^2 k(r)} + R(t,r)^2 d\Omega^2, \tag{6.3.1}$$

where prime denotes differentiation with respect to the coordinate r, $R(t,r)$ is the areal radius, $k(r)$ determines the intrinsic curvature on constant t slices [9], and $d\Omega^2 = d\theta^2 + \sin\theta^2 d\phi^2$. Since the metric (6.3.1) is describing the interior of a compact region, the coordinate radius r is only defined out to some $r_{\rm out}$. Furthermore, $R'(t,r) > 0$ and $k(r) < 1/r^2$ as the spherical shells are not allowed to cross one another.

Given the metric (6.3.1), the stress energy tensor for the pressureless dust fluid is given by (6.2.1) with

$$\vec{u} = \vec{\partial}_t, \quad \rho(t,r) = \frac{\bar{\rho} r^2}{R(t,r)^2 \; R'(t,r)}, \tag{6.3.2}$$

for some positive constant $\bar{\rho}$. The metric function $R(t,r)$ can be solved for analytically from the $G_{rr} = 8\pi T_{rr} = 0$ component of the Einstein equations. Here we restrict attention to the $k(r) > 0$ case. In this case we have

$$G_{rr} = 0 \Rightarrow \dot{R}(t,r)^2 + 2R(t,r)\ddot{R}(t,r) + r^2 k(r) = 0, \tag{6.3.3}$$

where dot denotes differentiation with respect to t. The closed form solution to Eq. (6.3.3) is given by the parametric equations

$$R(t,r) = \frac{4\pi\bar{\rho} r}{3k(r)}(1 - \cos u), \quad t - t_0(r) = \frac{4\pi\bar{\rho}}{3k(r)^{3/2}}(u - \sin u), \tag{6.3.4}$$

where $0 \leq u \leq 2\pi$. Evidently, there are two curvature singularities in this spacetime. The function $t_0(r)$ is the "bang function" specifying the coordinate time at which a spherical shell at a coordinate radius r departs from one of the curvature singularities, only to arrive at the other one at the

[8] For homogeneous density profiles, the metric is the Friedmann-Robertson-Walker (FRW) metric (1.1.1). The areal radius $R(t,r)$ in the metric (6.3.1) then reduces to $ra(t)$ in the homogeneous case.

[9] The intrinsic scalar curvature of the constant t slices is ${}^{(3)}R = 6k(r) + 2rk'(r)$.

coordinate time $t = t_0(r) + 4\pi^2\bar{\rho}/[3k(r)^{3/2}]$. Note that the function $t_0(r)$ is chosen such that the inner spherical shells arrive at the curvature singularities prior to the outer spherical shells.

II Dust objects with vacuum exterior

We now construct models of collapsing dust objects by gluing LTB interior solutions to Schwarzschild exterior solutions.

As we mentioned in Sec. 6.3.I, the LTB metric describes the interior geometry of a compact spherical region. The boundary of the interior region is a timelike three dimensional hypersurface that consists of points with the LTB coordinates $\{t, r_{\rm out}\}$.

The exterior geometry is given by the Schwarzschild metric

$$ds^2 = -\omega(\bar{r})d\bar{t}^2 + \omega(\bar{r})^{-1}d\bar{r}^2 + \bar{r}^2 d\Omega^2, \quad \omega(\bar{r}) = 1 - \frac{2M}{\bar{r}}, \tag{6.3.5}$$

where M is the gravitational mass associated with the collapsing object. The boundary of the compact inner region in the Schwarzschild coordinates is given by $\big\{t, R(t, r_{\rm out})\big\}$, where t is the proper time of the freely falling dust particles on the boundary of the inner region. Note that t coincides with the LTB time coordinate.

To smoothly join inner LTB solutions to outer Schwarzschild solutions one must impose the Israel junction conditions on the boundary. The details of this process is essentially identical to the Oppenheimer-Snyder collapse and it can be found in many references [10]. It follows from imposing the Israel junction conditions on the boundary region that

$$\begin{aligned} M &= \frac{R(t, r_{\rm out})}{2}\left[\dot{R}(t, r_{\rm out})^2 + r_{\rm out}^2 k(r_{\rm out})\right] \\ &= \frac{4\pi\bar{\rho} r_{\rm out}^3}{3}, \end{aligned} \tag{6.3.6}$$

where the second equality above follows from Eqs.(6.3.3) and (6.3.4). Actually, a more general statement follows from Eq.(6.3.3); that the Misner-Sharp mass enclosed by each spherical shell at the LTB coordinate radius r is

$$M(r) = \frac{R(t, r)}{2}\left[\dot{R}(t, r)^2 + r^2 k(r)\right] = \frac{4\pi\bar{\rho} r^3}{3}. \tag{6.3.7}$$

[10] We review the junction conditions for this case in App. E.

The future (and past) apparent horizon is then defined as the surface for which [11] [12]

$$1 - \frac{2M(r)}{R(t,r)} = 1 - r^2 k(r) - \dot{R}(t,r)^2 = 0 \tag{6.3.8}$$

for all r. All spherical shells are said to be marginally trapped on the apparent horizon. Once the apparent horizon is crossed, i.e.

$$1 - \frac{2M(r)}{R(t,r)} = 1 - r^2 k(r) - \dot{R}(t,r)^2 < 0 \tag{6.3.9}$$

for some spherical shell at r, then the shell is said to have become trapped.

III Dust objects with order M^2 entropy

In this section we show that one can construct collapsing dust objects using the LTB geometries with $k(r) > 0$ that come close to saturating the Bekenstein-Hawking entropy bound of $S_{\rm BH} = 4\pi M^2$ for non-black hole objects [13].

Using Eqs. (6.2.2) [14] (6.3.1) (for $k(r) > 0$) and (6.3.2) we find the entropy of a collapsing dust object on an achronal but otherwise arbitrary slice to be

$$S = 4\pi \int_0^{r_{\rm out}} dr \frac{\bar{\rho} r^2}{m(r)\sqrt{1 - r^2 k(r)}}. \tag{6.3.10}$$

Clearly, not all choices of the metric and particle mass functions result in valid models that can saturate the Bekenstein-Hawking entropy bound. In addition to requiring the hydrodynamic relations (6.2.3) to remain valid everywhere to the past of the future event horizon [15], we must also

[11] This can also be derived by finding the necessary condition for which the congruences of the outgoing future and past directed null geodesics that are orthogonal to a 2-sphere at the coordinate radius r have zero expansions.

[12] It is proven in [144] that for strongly predictable spacetimes satisfying the weak or strong energy conditions the future apparent horizon is inside the future event horizon. The LTB-Schwarzschild spacetimes for which the future event horizon is regular everywhere are strongly predictable and satisfy both the weak and strong energy conditions.

[13] This cannot be done for homogeneous dust objects in the Oppenheimer-Snyder collapse. The entropy (6.2.2) in these models is roughly M/m. However, for the fluid to obey the first and fourth of the conditions given in (6.2.3) we must have $m^{-1} \lesssim n^{-1/3} \ll M$. Similar conclusions hold for the LTB geometries with $k(r) = 0$ or $k(r) < 0$.

[14] Note that the LTB metric (6.3.1) together with Eq. (6.3.4) guarantee $\nabla_a(\rho u^a/m) = 0$ for m being an arbitrary function of r.

[15] For some choices of the LTB metric functions, it is possible for the future event horizon to encounter the past curvature singularity at some $r > 0$ [see App. E]. It is then implied that the hydrodynamics conditions (6.2.3) are violated somewhere on the future event horizon. Additionally, it follows that there are no spacetime slices that fall entirely outside of the future event horizon. Thus, the integral (6.3.10) does not account for the entropy of a non-black hole object.

require that a sufficiently large portion of the future apparent horizon consist of 2-spheres for which $|\dot{R}(t,r)| \ll 1$. This is a revision of a similar requirement that was originally set forth in [171]. It can be seen from Eq. (6.3.8) that the future (or past) apparent horizon cannot be entirely foliated by 2-spheres with $\dot{R} = 0$. In fact that would require $k(r) = 1/r^2$ for all r, for which the metric (6.3.1) would not be defined. Nonetheless, one might wonder if there are any constraints on how small $|\dot{R}|$ is for the 2-spheres that foliate the apparent horizon. Without any constraints, the integral (6.3.10) can be made arbitrarily large for choices of $k(r)$ that are arbitrarily close to $1/r^2$.

A physically realistic choice of constraint on the smallness of $|\dot{R}|$ along the apparent horizon can be explained as follows. Consider the 3-dimensional achronal surface Σ parametrized by $t_\Sigma(r)$ that is foliated by the 2-spheres with $\dot{R} = 0$. Now consider a constant t surface that intersects Σ at $\{t_\Sigma(\check{r}), \check{r}\}$. Define a coordinate radius $\tilde{r}$ on this constant t surface using the following relation,

$$R[t_\Sigma(\check{r}), \tilde{r}] = 2M(\check{r}). \tag{6.3.11}$$

The proper distance between the shell at $\check{r}$ and the shell at $\tilde{r}$ on this constant t surface is given by

$$\int_{\tilde{r}}^{\check{r}} \sqrt{g_{rr}} dr. \tag{6.3.12}$$

This distance corresponds to the spatial distance that the shell at $\{t_\Sigma(\check{r}), \check{r}\}$ would have to travel in order to arrive at the apparent horizon. We require this distance to be larger than the particle spacing on the shell at $\check{r}$, i.e.

$$\int_{\tilde{r}}^{\check{r}} \sqrt{g_{rr}} dr \geq 4n^{-1/3}\big[t_\Sigma(\check{r}), \check{r}\big], \tag{6.3.13}$$

[16]. According to the this constraint, Σ cannot be arbitrarily close to the apparent horizon since that would require the number density n to diverge on Σ, which takes the dust fluid outside of its regime of validity. We shall impose this constraint in the example that we discuss below.

We now construct an example as follows. We set $8\pi\bar{\rho} = 1$ and assume that $0 \leq r \leq 99/10$ for the interior region. It follows from Eq. (6.3.6) that the collapsing object has a gravitational mass $M = 161.7$ and a Bekenstein-Hawking entropy $S_{\rm BH} = 3.3 \times 10^5$ for the ensuing black hole in our choice of units. We then require the spacetime to admit a null surface Σ which is foliated by the

[16]The factor 4 is added so that the equality case for the relation (6.3.13) results in the entropy (6.3.10) being bounded by $S_{\rm BH}$.

spheres with $\dot{R} = 0$ in the interior region, i.e. $u = \pi$ everywhere on the interior part of Σ. Evaluating Eq. (6.3.4) at $u = \pi$ we find

$$R\big[t(r), r\big] = \frac{r}{3k(r)}, \qquad t(r) = t_0(r) + \frac{\pi}{6k(r)^{3/2}}, \tag{6.3.14}$$

the latter giving the defining equation for the surface Σ in the interior region. We can then use the fact that Σ is a null surface together with Eqs. (6.3.1), (6.3.4) and (6.3.14) to find a differential equation for the bang function $t_0(r)$:

$$\begin{aligned}
&\frac{dt}{dr} = \frac{R'(t,r)}{\sqrt{1-r^2k(r)}}\bigg|_{\Sigma} \\
&\Rightarrow t_0'(r) = \frac{1}{3k(r)}\left[\frac{1-\frac{rk'(r)}{k(r)}}{\sqrt{1-r^2k(r)}} + \frac{3\pi}{4}\frac{k'(r)}{k(r)^{3/2}}\right].
\end{aligned} \tag{6.3.15}$$

Thus, the knowledge of $k(r)$ is sufficient to fully specify the interior geometry. Once we have the interior metric, we can find the location of the future event horizon by first solving Eq. (6.3.4) for r_{out} to find the time t_* when the outermost spherical shell becomes trapped, and then solve the following equation for the null geodesic in the interior region with $\{t_*, r_{\text{out}}\}$ as its end point:

$$\begin{aligned}
&\frac{dt}{dr} = \frac{R'(t,r)}{\sqrt{1-r^2k(r)}} \\
&= \frac{\partial_r\left[\frac{r}{6k(r)}\Big(1-\cos\big[\mathcal{F}\big\{6k(r)^{3/2}(t-t_0(r)\big\}\big]\Big)\right]}{\sqrt{1-r^2k(r)}},
\end{aligned} \tag{6.3.16}$$

where we defined $\mathcal{F}(u - \sin u) \equiv u$ and used Eq. (6.3.4). We could then confirm that the surface Σ is indeed located outside of the future event horizon.

We proceed by finding a differential equation for $k(r)$ by saturating both the non-degeneracy condition given in (6.2.3) and the shell physical distance condition given in (6.3.13) for some spherical shells on Σ. On a given constant t surface intersecting Σ at $\{t_\Sigma(\check{r}), \check{r}\}$, we have

$$\begin{aligned}
&\int_{\tilde{r}}^{\check{r}} \sqrt{g_{rr}}dr = \int_{\tilde{r}}^{\check{r}} \frac{R'\big[t(r),r\big]dr}{\sqrt{1-r^2k(r)}}dr \\
&\approx \frac{1}{\sqrt{1-\check{r}^2k(\check{r})}}\int_{\tilde{r}}^{\check{r}} R'\big[t(r),r\big]dr \\
&= \frac{R\big[t(\check{r}),\check{r}\big] - 2M[\check{r}]}{\sqrt{1-\check{r}^2k(\check{r})}} = R\big[t(\check{r}),\check{r}\big]\sqrt{1-\check{r}^2k(\check{r})} \\
&= 4\times\big(8\pi\rho[t(\check{r}),\check{r}]\big)^{-1/4},
\end{aligned} \tag{6.3.17}$$

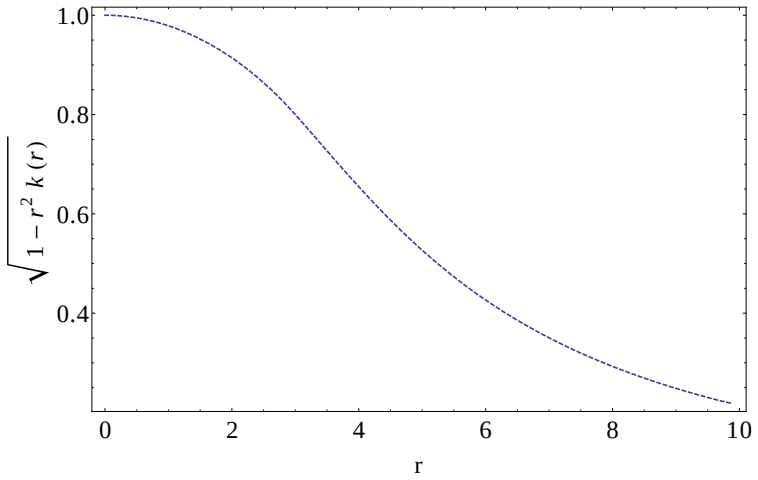


Figure 22: This is the plot of $\sqrt{1-r^2k(r)}$ versus r. As expected, $1-r^2k(r) \ll 1$ for $7 \leq r \leq 99/10$.

where we approximated the integral by assuming that the shell at $\{t(\check{r}), \check{r}\}$ is nearly trapped, i.e. $R\big[t(\check{r}), \check{r}\big] - 2M(\check{r}) = 1 - \check{r}^2 k(\check{r}) \ll 1$. We construct $k(r)$ by solving Eq. (6.3.17) for $7 \leq r \leq 99/10$ and then continuously joining that solution to a slowly varying function of r for $0 \leq r \leq 7$ [17]. We plot the resulting $\sqrt{1-r^2k(r)}$ in Fig.22. Note that $1-r^2k(r) \ll 1$ for $7 \leq r \leq 99/10$ is consistent with the assumption $R\big[t(r), r\big] - 2M(r) \ll 1$ for these shells used in Eq. (6.3.17).

With $k(r)$ in hand, we can now compute the integral (6.3.10) to find the total entropy

$$S = 1.8 \times 10^5 \approx 0.55\ S_{\rm BH}, \tag{6.3.18}$$

which turns out to be quite close to saturating the Bekenstein-Hawking entropy bound. See Fig. 23

Upon solving Eqs. (6.3.15) and (6.3.16), we find the location t_Σ of the slice Σ as well as the location $t_{\rm H}$ of the future event horizon. We plot $t_{\rm H} - t_\Sigma$ in Fig. 24. Note that the surface Σ is completely outside of the future event horizon. Additionally, as expected, the future apparent horizon is inside the future event horizon. See Fig. 25.

Finally, in Fig. 26 we plot the four relevant fluid micro-length scales together with the radius of curvature, $\{n^{-1/3}, (8\pi m)^{-1}, (16\pi/3)^{1/3} m/|\mathbf{D}m|, \rho^{-1/4}, \mathcal{L}\}$, defined in Sec. 6.2 along the future event horizon. Notice that all micro-length scales are larger than the particle spacing $n^{-1/3}$ everywhere along the future event horizon. Notice that the third condition given in (6.2.3) is close to becoming saturated. This indicates that the length scales over which the particle mass function $m(r)$ varies is

[17] The resulting $k(r)$ need not necessarily be C^∞ but it should be at least C^2.

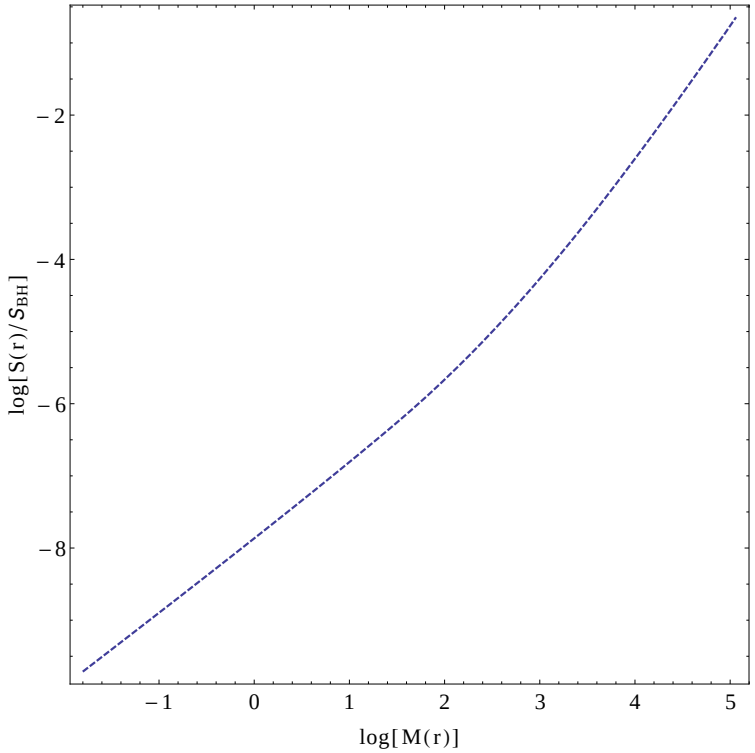


Figure 23: This is the plot of $\log\left[S(r)/S_{\rm BH}\right]$ versus $\log\left[M(r)\right]$ for the example we have constructed here. Notice that the contribution of the spherical shells at $r \geq 7$ brings the entropy given in Eq. (6.3.10) close to the Bekenstein-Hawking entropy $S_{\rm BH} = 3.3 \times 10^5$.

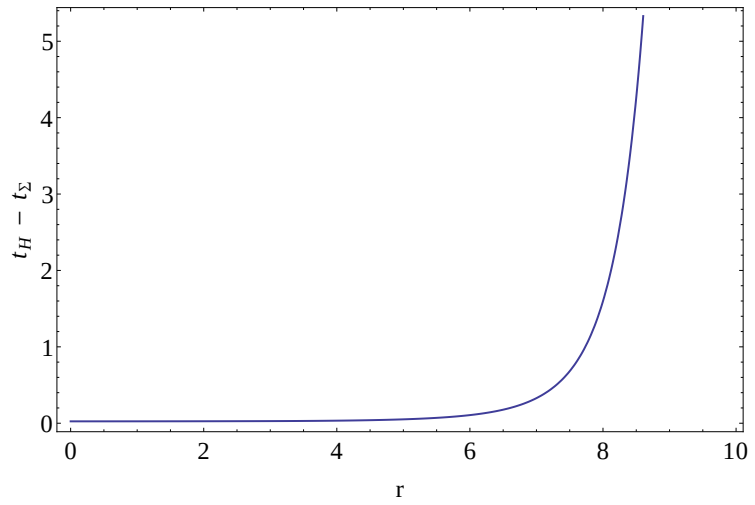


Figure 24: This is the plot of $t_{\rm H} - t_\Sigma$ versus radius r. The future event horizon is indeed to the future of the slice Σ

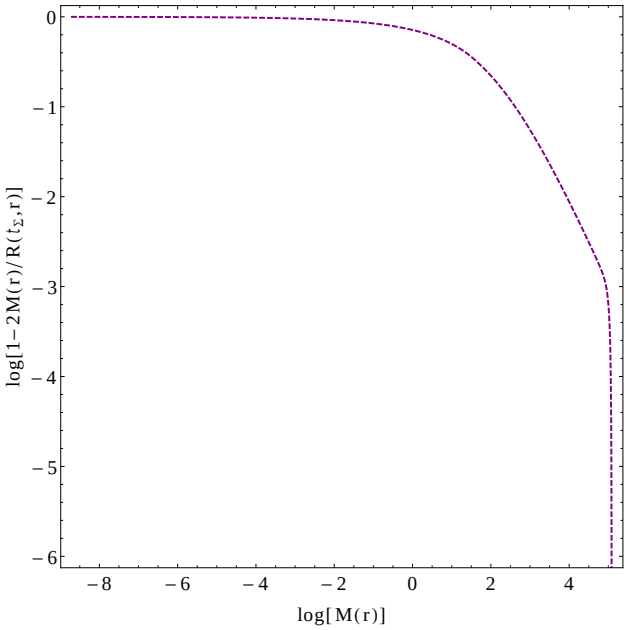


Figure 25: Here we plot $\log[1-2M(r)/R(t_{\rm H},r)]$ verses $\log[M(r)]$ along the future event horizon. Notice that none of the spherical shells are trapped, except for the outermost one for which $\log[1-2M(r_{\rm out})/R(t_{\rm H},r_{\rm out})]$ diverges.

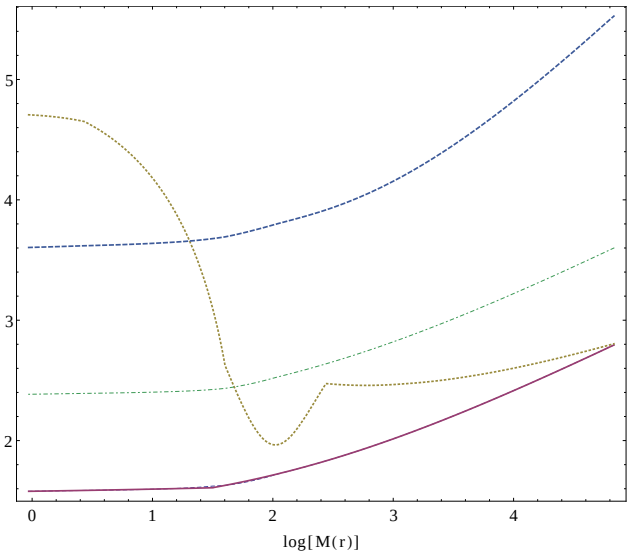


Figure 26: This is the plot of $\log\left[n^{-1/3}\right]$ (the solid blue line), $\log\left[(8\pi m)^{-1}\right]$ (the solid violet line which is collinear with the solid blue line), $\log\left[(16\pi/3)^{1/3}m/|\mathbf{D}m|\right]$ (the dotted brown line), $\log\left[\rho^{-1/4}\right]$ (the green dotted-dashed line), and $\log\left[\mathcal{L}\right]$ (the blue dashed line) versus $\log\left[M(r)\right]$ along the future event horizon.

becoming comparable to the particle spacing $n^{-1/3}$ along the future event horizon. Also, one should note that all five length scales are above the Planck length.

6.4 Attempts to find past geometries with no initial singularity

In Sec. (6.3.III) we constructed a collapsing dust object that saturates the Bekenstein-Hawking entropy bound. However, as we mentioned in the overview, the global time symmetry of the resulting spacetime renders it pathological due to the presence of a white hole. We find it worthwhile to attempt to construct spacetimes that have the same large entropy but not the white hole singularity in the past.

In the remainder of this chapter we report on a formalism that we have explored in order to circumvent the past white hole singularity. We find it convenient to initially lay out the formalism for the closed FRW geometries, and then apply it to the LTB geometries with $k(r) > 0$. As we show below, the numerical investigations appear to be indicative of the success of this formalism in both avoiding the white hole singularity and keeping the interior modified spacetime's stress energy tensor from violating the dominant energy condition [18]. However, we show in Sec. 6.4.II that a boundary stress energy tensor is needed to match the interior metric to the exterior Schwarzschild geometry. Given that this boundary stress energy tensor only satisfies the weak energy condition, the overall spacetime only satisfies the weak energy condition as well.

I The modified closed Friedman-Robertson-Walker geometry

In this section we show that one can modify the closed FRW geometry in the past in order to avoid the white hole singularity. By construction, the stress energy tensor of the modified spacetime satisfies the dominant energy conditions.

The metric for the modified spacetime is

$$ds^2 = a(\eta)^2\left[-d\eta^2 + d\chi^2 + \left(\frac{\sin\sqrt{\kappa(\eta)}\chi}{\sqrt{\kappa(\eta)}}\right)^2 d\Omega^2\right]. \tag{6.4.1}$$

where $\eta \equiv \int dt/a(t)$ is the conformal time and $\chi \equiv \int dr/\sqrt{1-\kappa r^2}$ [19]. Here, unlike the closed FRW geometry, the spatial curvature κ is a function of the conformal time η. We require $\kappa(\eta) > 0$ at all

[18] See [1] for a definition of the energy conditions.

[19] χ is defined for the FRW metric where κ is a constant.

times and $\kappa(\eta) = 1$ for $\eta \geq \eta_\Phi$, where η_Φ is a conformal time beyond which the spacetime geometry reduces to the closed FRW geometry. For times $-\infty < \eta \leq \eta_\Phi$, we define the metric functions $a(\eta)$ and $\kappa(\eta)$ by the following set of equations

$$\begin{aligned} &2\mathcal{H}' + \mathcal{H}^2 + \kappa\left(1 + \frac{\mathcal{L}}{\mathcal{H}}\right) = 0, \\ &\mathcal{L}' - \frac{1}{2}\mathcal{L}^2 + \mathcal{H}\mathcal{L} = 2g\mathcal{H}', \end{aligned} \tag{6.4.2}$$

with the initial conditions

$$\kappa(\eta_\Phi) = 1, \qquad \kappa'(\eta_\Phi) = 0, \qquad 0 < \mathcal{H}(\eta_\Phi) \ll 1^{20}. \tag{6.4.3}$$

The functions appearing in Eq. (6.4.2) are $\mathcal{L} \equiv \kappa'/\kappa$, $\mathcal{H} \equiv a'/a$, and $g(\eta) \equiv 1 - \exp\left[-(\eta_{\bar{\Phi}} - \eta)\right]$ for $\eta \leq \eta_{\bar{\Phi}}$ and zero otherwise. One might consider other choices of $g(\eta)$ depending on the smoothness properties that one wishes to demand of the stress energy tensor. For this particular choice of $g(\eta)$, the stress energy tensor will be continuous at $\eta = \eta_{\bar{\Phi}}$ and smooth everywhere else. Geometrically, we start from a point near the moment of time symmetry for the closed FRW geometry and evolve the spacetime backwards in time while diluting the spatial curvature on constant η slices. The objective is to push the moment of time symmetry off to $\eta \to -\infty$, thereby avoiding the white hole singularity.

Before discussing the properties of the modified FRW spacetime, we find it helpful to comment on the motivation behind the set of equations given in (6.4.2). The first equation can be regarded as the conservation of the homogenized Misner-Sharp mass. Indeed if we define $R(t, r) = ra(t)$ and require k to have time dependence in Eq. (6.3.7), we arrive at the first equation in (6.4.2) by setting $M_{,t} = 0$. The second equation in (6.4.2) is perhaps less motivated, merely designed to give a sufficiently slow evolution for κ in order to keep the modified spacetime from violating the dominant energy condition.

We now provide more details on the solutions to the dynamical equations given in (6.4.2). To see that the moment of time symmetry is avoided at any finite time, note that initially $\mathcal{H}'(\eta_\Phi) < 0$ and $\mathcal{H}(\eta_\Phi) > 0$ as we evolve the spacetime backwards in time. If $\mathcal{L} < 0$ and $\kappa > 0$ at all times $\eta < \eta_\Phi$, then $\mathcal{H}$ cannot go to zero since that would imply $\mathcal{H}' \to +\infty$, which is a contradiction.

[20]Note that we are evolving the metric functions from η_Φ to $-\infty$. Therefore, the sign of $\mathcal{H}(\eta_\Phi)$ is positive as the spacetime is expanding in this direction.

To see how $\mathcal{L}$ evolves in time, note that the second equation in (6.4.2) together with the initial conditions $\mathcal{L}(\eta_\Phi) = \mathcal{L}'(\eta_\Phi) = 0$ implies that $\mathcal{L}''(\eta_\Phi) < 0$. Thus, $\mathcal{L} < 0$ at times $\eta < \eta_\Phi$ and sufficiently close to η_Φ. Furthermore, we expect to have $\mathcal{L} > -\mathcal{H}$ as long as $\kappa > 0$. Indeed if $\mathcal{L} = -\mathcal{H}$ at some finite time $\eta_* < \eta_\Phi$, Eq. (6.4.2) implies $\mathcal{L}'(\eta_*) > |\mathcal{H}'(\eta_*)| = \mathcal{H}(\eta_*)^2/2$. Therefore, while $\mathcal{L}$ initially decreases below 0 we must have $\mathcal{L} > -\mathcal{H}$. The only way that $\mathcal{L}$ can become equal to $-\mathcal{H}$ at some time η_* is to have $0 \leq \mathcal{L}'(\eta_*) \leq |\mathcal{H}'(\eta_*)|$ which is inconsistent. Likewise, $\mathcal{L}(\eta_*) = 0$ is forbidden since it requires $\mathcal{L}' \geq 0$ in the vicinity of η_*, whereas Eq. (6.4.2) implies $\mathcal{L}' = 2g\mathcal{H}' < 0$ at η_*.

In fact, given our choice of initial conditions, it is not difficult to see that the dynamical system under study has an attractor solution $\mathcal{L} \to -\mathcal{H}$. Informally, we can see that this solution is a true attractor by writing $\mathcal{L}(\eta) = \mathcal{L}_0(\eta) + \delta(\eta)$, where $\mathcal{L}_0 = -\mathcal{H}$ and initially $|\delta| \ll |\mathcal{L}_0|$. We can then substitute this expression for $\mathcal{L}$ into the second equation in (6.4.2) and assume $|\eta - \eta_\Phi| \gg 1$ to find the following differential equation for δ

$$\delta_{,\eta} - \frac{1}{2}\delta^2 - 2\mathcal{L}_0\delta = 0. \tag{6.4.4}$$

Since for the attractor solution we have $\mathcal{H} \approx -\mathcal{L} \approx 2/|\eta|$, Eq. (6.4.4) gives $\delta \propto 1/\eta^4$, which falls off faster than $\mathcal{L}_0$ for large $|\eta|$ as expected. Thus, we have $a \propto \eta^2$ and $\kappa \propto 1/\eta^2$ as $\eta \to -\infty$. As a result, one can check that the components of the Riemann tensor for the constant η slices go to zero as $\eta \to -\infty$. This is equivalent to saying that the constant η slices become nearly intrinsically flat at large $|\eta|$.

For the dominant energy condition to be satisfied we must have

$$\rho \geq |p_{\hat{x}_1}|, \qquad \rho \geq |p_{\hat{\theta}}|, \tag{6.4.5}$$

at all times, where $\{\rho, p_{\hat{x}_1}, p_{\hat{\theta}}\}$ are the independent eigenvalues of the stress energy tensor in a locally orthonormal frame

$$ds^2 = -d\hat{x}_0^2 + d\hat{x}_1^2 + d\hat{\theta}^2 + d\hat{\phi}^2 \tag{6.4.6}$$

($p_{\hat{\theta}} = p_{\hat{\phi}}$ due to spherical symmetry). Given the metric (6.4.1), a routine calculation gives (η dependence is suppressed)

$$\begin{aligned}
\rho &= \frac{1}{16\pi a^2}\Big[G_{\eta\eta} - G_{\chi\chi} + \sqrt{(G_{\eta\eta} + G_{\chi\chi})^2 - 4G_{\eta\chi}^2}\Big], \\
p_{\hat{x}_1} &= \frac{1}{16\pi a^2}\Big[-G_{\eta\eta} + G_{\chi\chi} + \sqrt{(G_{\eta\eta} + G_{\chi\chi})^2 - 4G_{\eta\chi}^2}\Big], \\
p_{\hat{\theta}} &= \frac{\kappa}{8\pi a^2 \sin(\sqrt{\kappa}\chi)^2} G_{\theta\theta},
\end{aligned} \tag{6.4.7}$$

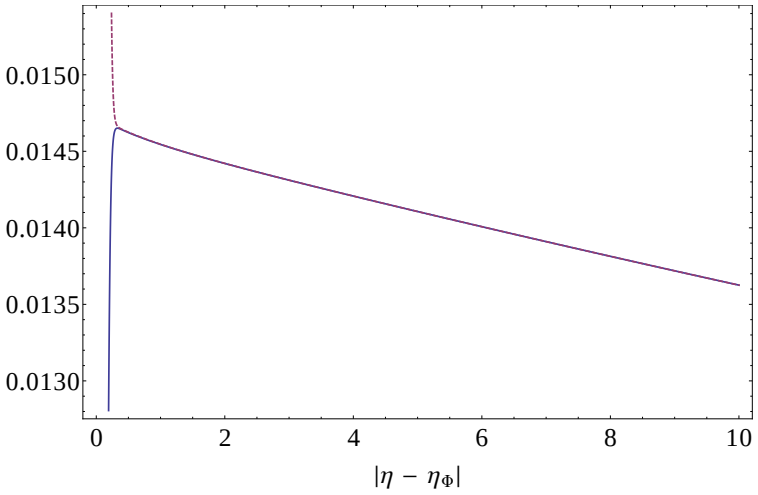


Figure 27: This is the plot of $|\mathcal{L}|$ (the solid blue line) and $\mathcal{H}$ (the dashed purple line) versus $|\eta - \eta_\Phi|$ as we evolve the spacetime backwards in time. Notice that $|\mathcal{L}|$ nearly converges to $\mathcal{H}$ very quickly.

where G_{ab} is the Einstein's tensor. We should first show that the quantity inside the square root in Eq. (6.4.7) remains positive semi-definite at all times. This condition is also necessary to ensure that the eigenvector associated with ρ is timelike. It is sufficient to have

$$\begin{aligned}
&G_{\eta\eta} + G_{\chi\chi} \geq 2|G_{\eta\chi}| \\
&\Rightarrow 2\mathcal{H}^2 - 2\mathcal{H}' + 2\kappa + \frac{1}{2}\mathcal{L}^2\big[\chi^2\kappa - f(\chi,\kappa)\big] + f(\chi,\kappa)\mathcal{L}' \\
&\geq -2\chi\kappa\mathcal{L} \\
&\Rightarrow 2\mathcal{H}^2 - 2\big[1 - gf(\chi,\kappa)\big]\mathcal{H}' + 2\kappa - f(\chi,\kappa)\mathcal{H}\mathcal{L} + \frac{1}{2}\chi^2\kappa\mathcal{L}^2 \\
&\geq -2\chi\kappa\mathcal{L}, \qquad (6.4.8)
\end{aligned}$$

where $f(\chi,\kappa) \equiv 1 - \chi\sqrt{\kappa}\cot(\chi\sqrt{\kappa})$ and we used the fact that $\mathcal{L} \leq 0$ at all times. We also used Eq. (6.4.2) in the last inequality. Since $0 < \kappa \leq 1$ and $0 \leq \chi \leq \pi/2$, we have $0 \leq f \leq 1$. Given that $|\mathcal{L}| < \mathcal{H}$ at all times, the inequality (6.4.8) comes down to

$$\begin{aligned}
&2\mathcal{H}^2 - 2\big[1 - gf(\chi,\kappa)\big]\mathcal{H}' + 2\kappa\big[1 - \frac{\pi}{2}\mathcal{H}\big] - f(\chi,\kappa)\mathcal{H}\mathcal{L} \\
&+\frac{1}{2}\chi^2\kappa\mathcal{L}^2 \geq 0 \qquad (6.4.9)
\end{aligned}$$

which is correct at times $\eta < \eta_\Phi$ because $\mathcal{H} \ll 1$, $\mathcal{L}\mathcal{H} < 0$, and $[1 - gf(\chi,\kappa)]\mathcal{H}' < 0$, with the latter following from Eq. (6.4.2) and $0 \leq g < 1$.

To show $\rho \geq |p_{\hat{x}_1}|$, it is sufficient to show that $G_{\eta\eta} - G_{\chi\chi} \geq 0$. Using Eq. (6.4.2), we have

$$\begin{aligned}
&G_{\eta\eta} - G_{\chi\chi} = 4\mathcal{H}^2 + 2[1 - gf(\chi,\kappa)]\mathcal{H}' - 3f(\chi,\kappa)\mathcal{H}\mathcal{L} + 4\kappa \\
&+\mathcal{L}^2\Big[1 - \frac{1}{2}f(\chi,\kappa) - \frac{3}{2}\chi\sqrt{\kappa}\cot(\chi\sqrt{\kappa}) \\
&+\frac{1}{2}\kappa\chi^2\big(-1 + \cot(\sqrt{\kappa}\chi)^2\big)\Big] \geq 4\mathcal{H}^2 + 2\mathcal{H}' - 3f(\chi,\kappa)\mathcal{H}\mathcal{L} \\
&+4\kappa - \mathcal{L}^2 \geq 2\mathcal{H}^2 + 3\kappa - \kappa\frac{\mathcal{L}}{\mathcal{H}} - 3\mathcal{H}\mathcal{L} \geq 0.
\end{aligned} \tag{6.4.10}$$

Finally, we ought to show that $\rho \geq |p_{\hat{\theta}}|$. As for $|p_{\hat{\theta}}|$, we have

$$\begin{aligned}
&8\pi a^2|p_{\hat{\theta}}| = \Big| -\kappa\frac{\mathcal{L}}{\mathcal{H}}\Big[1 + \frac{1}{2}gf(\chi,\kappa)\Big] - \frac{1}{2}gf(\chi,\kappa)[\kappa + \mathcal{H}^2] \\
&+\frac{1}{2}f(\chi,\kappa)\mathcal{H}\mathcal{L} + \frac{1}{4}\chi^2\kappa\mathcal{L}^2\Big| \leq -\frac{3}{2}\kappa\frac{\mathcal{L}}{\mathcal{H}} + \frac{1}{2}[\kappa + \mathcal{H}^2] - \frac{1}{2}\mathcal{H}\mathcal{L} \\
&+\frac{1}{4}\chi^2\kappa\mathcal{L}^2 \leq 2\kappa + \mathcal{H}^2\Big[1 + \frac{1}{4}\chi^2\kappa\Big] \leq 2[\kappa + \mathcal{H}^2].
\end{aligned} \tag{6.4.11}$$

Likewise we have for the locally measured energy density

$$\begin{aligned}
&8\pi a^2\rho = \frac{1}{2}\Big[G_{\eta\eta} - G_{\chi\chi} + \sqrt{(G_{\eta\eta} + G_{\chi\chi})^2 - 4G_{\eta\chi}^2}\Big] \\
&\geq G_{\eta\eta} - |G_{\eta\chi}| = 3[\kappa + \mathcal{H}^2] - 2f(\chi,\kappa)\mathcal{H}\mathcal{L} + \chi\kappa\mathcal{L} \\
&+\frac{1}{2}\mathcal{L}^2\Big[f(\chi,\kappa) - \frac{1}{2} + \frac{1}{2}\kappa\chi^2\cot(\chi\sqrt{\kappa})^2\Big] \\
&\geq 3[\kappa + \mathcal{H}^2] + \chi\kappa\mathcal{L} \geq \big[3 - \frac{\pi}{2}\mathcal{H}\big]\kappa + 3\mathcal{H}^2,
\end{aligned} \tag{6.4.12}$$

where we used

$$f(\chi,\kappa) - \frac{1}{2} + \frac{1}{2}\kappa\chi^2\cot(\chi\sqrt{\kappa})^2 > 0. \tag{6.4.13}$$

As can be seen from the last expressions in Eq. (6.4.11) and Eq. (6.4.12),for $\rho \geq |p_{\hat{\theta}}|$ it suffices to have $1 - (\pi/2)\mathcal{H} \geq 0$, which is true for $\eta < \eta_\Phi$ given that $\mathcal{H} \ll 1$ at such times. Therefore, we conclude that the modified FRW spacetime satisfies the dominant energy condition everywhere to the past of η_Φ.

II A numerical study of the modified Lemaître-Tolman-Bondi geometry

We now apply the formalism developed in Sec. 6.4.I to the LTB metric given in Eq. (6.3.1). The equations (6.4.2) in this case become

$$\begin{aligned}
&\dot{R}^2 + 2R\ddot{R} + r^2 k\Big(1 + \frac{\dot{k}R}{k\dot{R}}\Big) = 0, \\
&\dot{R}\frac{\dot{k}}{k} - \frac{1}{2}R\Big(\frac{\dot{k}}{k}\Big)^2 + \partial_t\Big(R\frac{\dot{k}}{k}\Big) = 2g\ddot{R},
\end{aligned} \tag{6.4.14}$$

where dot denotes differentiation with respect to the LTB time coordinate t and $g = 1 - \exp\left[-\{t_\Phi(r) - t\}\right]$. We define $t_\Phi(r)$ to be the location of the spacelike slice between t_Σ and the future event horizon found in Sec. 6.3.III for which all spherical shells have

$$\begin{aligned}
\dot{R}[t_\Phi(r), r] &\approx \frac{\Big(\frac{\pi}{2} - 3k(r)^{3/2}[t_\Phi(r) - t_0(r))\Big]}{2} r\sqrt{k(r)} \\
&= -\frac{10^{-6}}{2} r\sqrt{k(r)},
\end{aligned} \tag{6.4.15}$$

where $k(r)$ is the spatial curvature that we constructed in Sec. 6.3.III. The relation given in the first line of Eq.(6.4.15) comes from the Taylor expansion of $\dot{R}$ near t_Σ for which $\dot{R} = 0$ everywhere. The numerical factor 10^{-6} in the second line of Eq. (6.4.15) is chosen so that all spherical shells are located between t_Σ and the future event horizon.

We numerically solve the coupled system of ordinary differential equations (6.4.14) for $-\infty < t < t_\Phi(r)$ and $0 \leq r \leq 99/10$ using the following initial conditions [21]

$$\begin{aligned}
&k[t_\Phi(r), r] = k(r), \qquad \dot{k}[t_\Phi(r), r] = 0, \\
&\dot{R}[t_\Phi(r), r] = \frac{10^{-6}}{2} r\sqrt{k(r)}.
\end{aligned} \tag{6.4.16}$$

We then examine the locally measured energy density and principal pressures of the modified spacetime, which are given by

$$\begin{aligned}
\rho &= \frac{1}{16\pi}\Big[G_{tt} - G^r{}_r + \sqrt{(G_{tt} + G^r{}_r)^2 - 4G_{tr}G^r{}_t}\Big], \\
p_{\hat{x}_1} &= \frac{1}{16\pi}\Big[-G_{tt} + G^r{}_r + \sqrt{(G_{tt} + G^r{}_r)^2 - 4G_{tr}G^r{}_t}\Big], \\
p_{\hat{\theta}} &= \frac{1}{8\pi}G^\theta{}_\theta.
\end{aligned} \tag{6.4.17}$$

The modified LTB spacetime shares very similar qualitative features to its homogeneous counterpart. In particular, $\dot{k}/k \to -\dot{R}/R$, or equivalently $R \propto t^{2/3}$ and $k \propto t^{-2/3}$, at times $t \ll t_\Phi(r)$ for all

[21] Recall that we are integrating backwards in time, so $\dot{R} > 0$.

spherical shells. Similar to the case of the modified FRW metric, it is not difficult to check that the constant t slices become nearly intrinsically flat as $t \to -\infty$. Despite the fact that the eigenvalues of the modified stress energy tensor no longer have the same functional form as the ones obtained in the previous section as they now contain terms with radial derivatives, our extensive numerical investigations provide no evidence for the violation of the dominant energy condition. Therefore, we are convinced that the modified LTB spacetime constructed using (6.4.14) avoids the white hole singularity while satisfying the dominant energy condition. In Fig. 28 we plot the eigenvalues of the modified stress energy tensor in a local orthonormal frame for spherical shells at a few different radii.

Finally, note that there is also a non-vanishing boundary stress energy tensor for $t < t_\Phi(r_{\text{out}})$. This is due to the discontinuity of the time-time component of the boundary extrinsic curvature in this region of spacetime. Indeed, $\vec{u} = \vec{\partial}_t$ is no longer a geodesic in the exterior Schwarzschild geometry for $t < t_\Phi(r_{\text{out}})$, while remaining a geodesic in the interior geometry at all times. The independent eigenvalues of the boundary stress energy tensor in the orthonormal $\{t, \hat{\theta}, \hat{\phi}\}$ frame are given by [22]

$$\bar{\rho} = 0, \qquad \bar{p} = -\frac{1}{16\pi} \frac{r^2 k}{R\sqrt{1 - \frac{2M}{R} + \dot{R}^2}} \frac{\dot{k}R}{k\dot{R}}, \tag{6.4.18}$$

which are evidently positive semi-definite at all times due to the fact that $(\dot{k}R)/(k\dot{R}) \leq 0$ [See Fig. 27]. Therefore, the boundary stress energy tensor satisfies the weak energy condition.

III Black hole formation timescale

We find it appropriate to define the black hole formation timescale to be the interval along $\mathcal{I}_-$ between the two ingoing null geodesics that intersect the beginning of the event horizon and the outer boundary of the collapsing object as it crosses the event horizon [See Fig. 29]. Intuitively, this corresponds to the time required for the collapsing object to cross its own event horizon as measured by the asymptotic null observers at $\mathcal{I}_-$.

We now show that this timescale is on the order of M^2 for the modified LTB collapse models constructed in Sec. 6.4.II. To begin, notice that the modified LTB geometry evolves very slowly

[22] The boundary stress energy tensor comes from a standard calculation for the exterior Schwarzschild geometry which can be found in [180]. We derive this result in App. E.

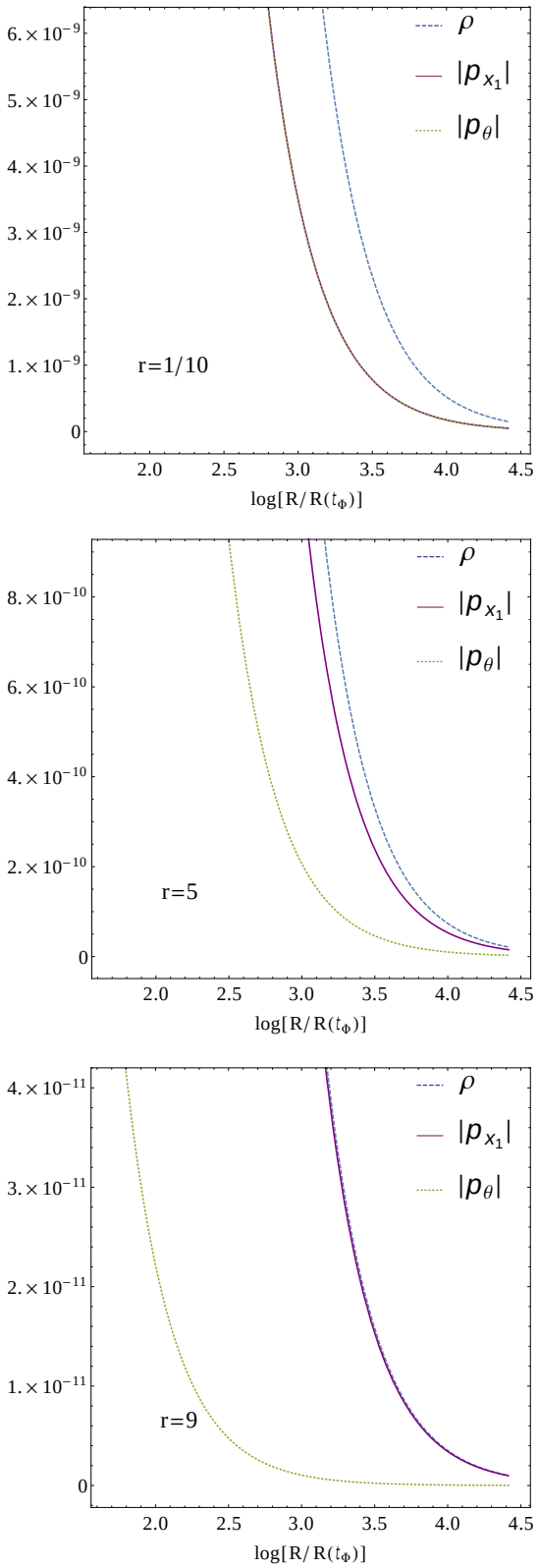


Figure 28: This is the plot of the eigenvalues of the modified stress energy tensor versus $\log[R(t,r)/R(t_\Phi,r)]$ for spherical shells at three different radii.

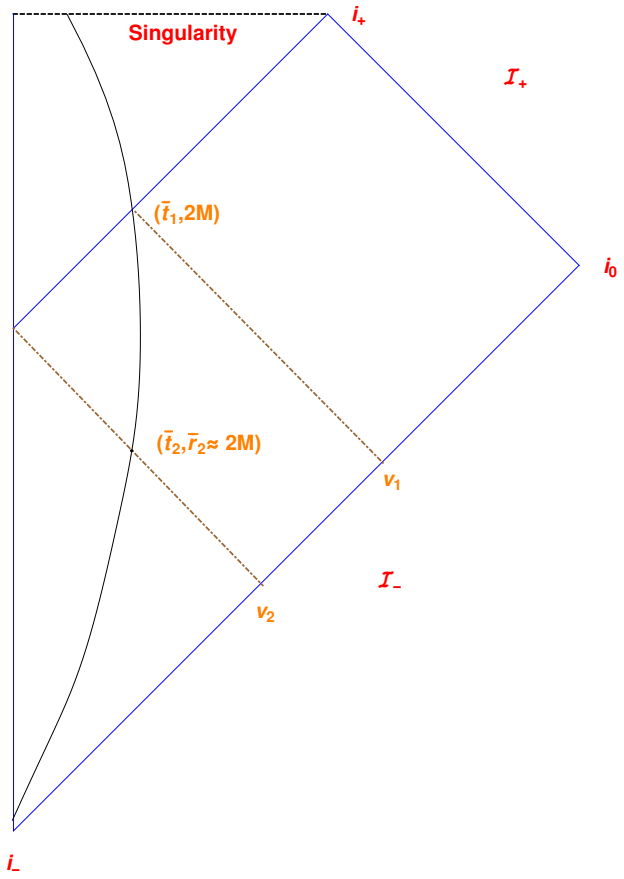


Figure 29: This is the Penrose diagram of a collapsing object embedded in a Schwarzschild exterior. The ingoing null rays with advanced Eddington-Finkelstein coordinates v_1 and v_2 intersect the collapsing object at $(\bar{t}_1, 2M)$ and $(\bar{t}_2, \bar{r}_2 \approx 2M)$ respectively.

compared to the LTB geometry. In fact, to a good approximation we have

$$\Delta v = v_1 - v_2 \approx \tilde{v}_1 - v_2 = \Delta\bar{t} + \Delta\bar{r}_* \approx \Delta\bar{t}, \tag{6.4.19}$$

where $\bar{r}_* \equiv \bar{r} + 2M \log|\bar{r}/(2M) - 1|$ is the Schwarzschild's tortoise coordinate and $\tilde{v}_1$ is the advanced Eddington-Finkelstein coordinate associated with the boundary 2-sphere on the surface Σ [23]. In other words, the proper observer on the boundary is nearly stationary between the times $\bar{t}_1$ and $\bar{t}_2$. As a result, we have

$$\Delta\bar{t} \approx \frac{\Delta t}{\sqrt{1 - \frac{2M}{R(t_1, r_{\rm out})}}}, \tag{6.4.20}$$

where Δt is the proper time interval between $\bar{t}_1$ and $\bar{t}_2$ measured by the proper observer on the boundary. Using the fact that the proper observer on the boundary remains quite close to its initial

[23] See Sec. 6.3.III. Recall that Σ is near the event horizon.

location where $\dot{R} \approx 0$, it follows from the Eqs. (6.3.15), and (6.3.17) that

$$\Delta\bar{t} \sim M\rho_{\rm out}^{1/4}\Delta t \sim M\rho_{\rm out}^{1/2}S \sim M^3\rho_{\rm out}^{1/2} \sim M^2, \tag{6.4.21}$$

where S is the entropy evaluated on Σ and $\rho_{\rm out} \equiv \rho[t_\Phi(r_{\rm out}), r_{\rm out}] \sim r_{\rm out}^{-6} \sim M^{-2}$, assuming that $k \sim r^{-2}$ for the outer shells [24]. To arrive at (6.4.21), we also used the slow evolution of the interior metric to argue that Δt is on the order of the temporal change along the future event horizon. We then approximated the temporal change along the event horizon using its value along the surface Σ. This is a reasonable approximation given the proximity of Σ to the event horizon.

[24] See Sec. 6.3.III

Summary Of Results And Future Outlook

In part II we introduced the entropy-mass-timescale conjecture for the large class black hole formation scenarios in which the final state black holes are stationary and asymptotically flat. We conjectured that this relationship is valid generically for all black hole precursors that satisfy the dominant energy condition, though it is likely incorrect [25] for the ones that only satisfy the weak energy condition.

This conjecture, if correct, can further illuminate the nature of generic quantum mechanical states of black hole interiors. A direct consequence of this conjecture is that the vast majority of black hole quantum states are results of the formation processes that take a time on the order of the Hawking evaporation process, while only a small portion of them result from astrophysical collapse scenarios that are known to occur over a much shorter time. In conjunction with recent advances in Holography [26], conjectures of this kind can provide valuable information about quantum gravitational degrees of freedom that are responsible for black hole entropy.

[25] In the sense that there are counter examples (not necessarily generic) as we constructed in chapter 6.
[26] e.g. the Covariant Entropy Bound [166] and the Quantum Focusing Conjecture [181].

Appendix E

Appendices to chapter 6

Israel junction conditions for an LTB interior glued to a Schwarzschild exterior

The calculation provided here is essentially identical to the FRW-Schwarzschild model which can be found in standard references such as [180] and [182]. To begin, recall that the LTB and the Schwarzschild metrics are

$$
\begin{aligned}
{}^{(1)}ds^2 &= -dt^2 + \frac{R'(t,r)^2}{1-r^2k(r)}dr^2 + R(t,r)^2 d\Omega^2, \\
{}^{(2)}ds^2 &= -\omega(\bar{r})d\bar{t}^2 + \omega(\bar{r})^{-1}d\bar{r}^2 + \bar{r}^2 d\Omega^2,
\end{aligned}
\tag{E.1}
$$

where $\omega(\bar{r}) \equiv 1 - 2M/\bar{r}$, prime denotes differentiation with respect to r, and we used superscripts (1) and (2) to label the LTB and Schwarzschild spacetimes respectively. The LTB metric is used to describe the interior geometry of a collapsing object. As a result, the radial coordinate r is within an interval $0 \leq r \leq r_{\text{out}}$.

The boundary surface between the two spacetimes (i.e. the world tube) is a timelike three dimensional surface. In the LTB coordinates, this surfaces is given by $\{t, r_{\text{out}}\}$, for which the induced metric is then $-dt^2 + R^2 d\Omega^2$. In this coordinate system, the unit tangent vector field to the world tube is given by ${}^{(1)}\vec{u} = \vec{\partial}_t$ and the unit normal vector field is given by ${}^{(1)}\vec{n} = \vec{\partial}_r / \sqrt{{}^{(1)}g_{rr}}$.

In the Schwarzschild coordinates, the unit tangent and the unit normal vector fields to the world tube can be parametrized as ${}^{(2)}\vec{u} = \dot{\bar{t}}\vec{\partial}_{\bar{t}} + \dot{\bar{r}}\vec{\partial}_{\bar{r}}$ and ${}^{(2)}\vec{n} = \dot{\bar{r}}/\omega\vec{\partial}_{\bar{t}} + \omega\dot{\bar{t}}\vec{\partial}_{\bar{r}}$ respectively. Here $\bar{t}$ and $\bar{r}$ are functions of the proper time for an observer on the world tube. We assume $\dot{\bar{t}} > 0$ and $\dot{\bar{r}} < 0$. The induced metric on the world tube in this coordinate system is then $-d\tau^2 + \bar{r}^2 d\Omega^2$ with τ being the proper time for an observer on the surface.

The first Israel junction condition asserts the continuity of the induced metric on the world tube surface. It follows that the LTB coordinate time t and the proper time τ coincide. Moreover, the radial coordinate describing the world tube surface in the Schwarzschild spacetime is given by $R(t, r_{\rm out})$. The second Israel junction condition relates the discontinuity in the components of the world tube's extrinsic curvature in the two spacetimes (if any) to the presence of a surface stress energy tensor,

$$S_{ab} = -\frac{1}{8\pi}\Big([K_{ab}] - [K]h_{ab}\Big), \tag{E.2}$$

where h_{ab} is the induced metric on the world tube, $K_{ab} \equiv h_a{}^c h_b{}^d \nabla_c n_d$ is the extrinsic curvature of the world tube, and $[X] \equiv {}^{(2)}X - {}^{(1)}X$. The non zero components of the extrinsic curvature in the LTB spacetime are

$$\begin{aligned}
&{}^{(1)}K^t{}_t = -{}^{(1)}u_a\,{}^{(1)}u^b\,{}^{(1)}\nabla_b\,{}^{(1)}n^a = {}^{(1)}n^a\,{}^{(1)}u^b\,{}^{(1)}\nabla_b\,{}^{(1)}u_a = {}^{(1)}n^a\,{}^{(1)}\alpha_a = 0,\\
&{}^{(1)}K^\theta{}_\theta = {}^{(1)}\nabla_\theta\,{}^{(1)}n^\theta = {}^{(1)}\Gamma^\theta_{\theta r}\,{}^{(1)}n^r = \left.\frac{\sqrt{1-r^2k(r)}}{R}\right|_{r=r_{\rm out}},\\
&{}^{(1)}K^\theta{}_\theta = {}^{(1)}K^\phi{}_\phi,
\end{aligned} \tag{E.3}$$

where α_a is the four-acceleration of the proper observer on the world tube which vanishes since ${}^{(1)}\vec{u} = \vec{\partial}_t$ and the LTB metric has a Gaussian normal form. Likewise, the non zero components of the extrinsic curvature in the Schwarzschild spacetime are

$$\begin{aligned}
&{}^{(2)}K^t{}_t = {}^{(2)}n^a\,{}^{(2)}\alpha_a = \left.\frac{\ddot{R} + \frac{M}{R^2}}{\sqrt{\omega(R)+\dot{R}^2}}\right|_{r=r_{\rm out}},\\
&{}^{(2)}K^\theta{}_\theta = {}^{(2)}\nabla_\theta\,{}^{(2)}n^\theta = {}^{(2)}\Gamma^\theta_{\theta t}\,{}^{(2)}n^t + {}^{(2)}\Gamma^\theta_{\theta r}\,{}^{(2)}n^r = \left.\frac{\sqrt{\omega(R)+\dot{R}^2}}{R}\right|_{r=r_{\rm out}},\\
&{}^{(2)}K^\theta{}_\theta = {}^{(2)}K^\phi{}_\phi.
\end{aligned} \tag{E.4}$$

The Misner-Sharp mass defined in Eq. (6.3.7) together with the LTB equation (6.3.3) assert that $\ddot{R} + M/R^2$ vanishes at all radii, including at $r_{\rm out}$. Similarly, it follows from Eq. (6.3.7) that $K^\theta{}_\theta$ is continuous across the world tube boundary. Thus, all components of the world tube's extrinsic curvature are continuous, which implies that the LTB interior can be smoothly glued to a Schwarzschild exterior without a need for a boundary stress energy tensor.

Now suppose we promote the LTB curvature function $k(r)$ to be a function of the LTB time t as well. The components of the world tube's extrinsic curvature for the modified interior spacetime are

essentially identical to the ones given in Eq. (E.3), except that now $k(r)$ is replaced with $k(t,r)$. In this case, $K^{\theta}{}_{\theta}$ is still continuous across the world tube's boundary by virtue of the first equation given in Eq. (6.4.14), which is simply $M_{,t}=0$. However, the tt component of the extrinsic curvature no longer vanishes for the Schwarzschild spacetime. It follows from Eq. (6.4.14) that

$$ {}^{(2)}K^{t}{}_{t} = \left.\frac{\ddot{R}+\frac{M}{R^2}}{\sqrt{\omega(R)+\dot{R}^2}}\right|_{r=r_{\rm out}} = \left.\frac{2R\ddot{R}+\dot{R}^2+r^2k}{2R\sqrt{\omega(R)+\dot{R}^2}}\right|_{r=r_{\rm out}} = \left.\frac{-r^2k\frac{\dot{k}R}{\dot{R}k}}{2R\sqrt{\omega(R)+\dot{R}^2}}\right|_{r=r_{\rm out}}. \tag{E.5} $$

We therefore have

$$ [K_{tt}] = {}^{(2)}K_{tt}, \qquad [K_{\theta\theta}]=[K_{\phi\phi}]=0, \qquad [K] = {}^{(2)}K^{t}{}_{t}. \tag{E.6} $$

The surface stress energy tensor then becomes

$$ \begin{aligned} S_{tt} &= -\frac{1}{8\pi}\Big({}^{(2)}K_{tt} + {}^{(2)}K^{t}{}_{t}\Big) = 0, \\ S^{\theta}{}_{\theta} = S^{\phi}{}_{\phi} &= \frac{1}{8\pi}{}^{(2)}K^{t}{}_{t} = -\frac{1}{16\pi}\left.\frac{r^2k\frac{\dot{k}R}{\dot{R}k}}{R\sqrt{\omega(R)+\dot{R}^2}}\right|_{r=r_{\rm out}}. \end{aligned} \tag{E.7} $$

Therefore, the boundary stress energy tensor has vanishing energy density and non vanishing pressure that is given by $S^{\theta}{}_{\theta}$. Given the fact that $\dot{k}R/(k\dot{R}) \leq 0$ at all times, all eigenvalues of the boundary stress energy tensor remain positive semi-definite at all times.

Lemaître-Tolman-Bondi geometry with event horizon intersecting the singularity

In Sec. 6.3.III we mentioned that it is possible for some LTB-Schwarzschild geometries not to have any achronal slices that are completely outside of the future event horizon. It is then implied that in these spacetimes the future event horizon intersects the past curvature singularity. As an example, let us assume that the interior LTB geometry has a radial coordinate r that covers from $r=0$ to $r=99/10\equiv r_{\rm out}$. Now consider a constant time slice Θ for which we set $t=0$ and that it is foliated by the maximal spheres, i.e. $\dot{R}(t_{\Theta},r)=0$ everywhere on Θ. After setting $8\pi\bar{\rho}=1$, it follows from Eq. (6.3.4) that $u=\pi$ and

$$ t_0(r) = -\frac{\pi}{6k(r)^{3/2}}. \tag{E.8} $$

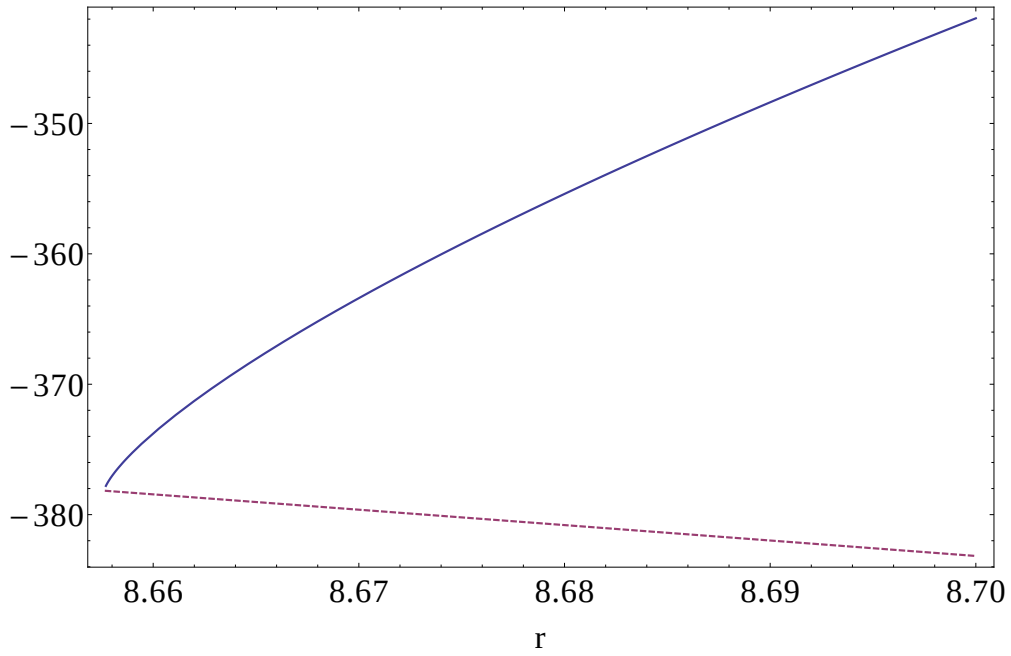


Figure 30: This is the plot of $t_{\rm H}$ (solid blue line) and $t_0(r)$ (dashed violet line) versus r. The future event horizon intersects the curvature singularity at $r \approx 8.657$.

As for $k(r)$, we take the same solution as in Sec. 6.3.III by assuming that the non-degeneracy condition (6.2.3) the shell-apparent horizon physical distance condition (6.3.13) are both saturated. We then solve Eq. (6.3.16) to find the location of the future event horizon subject to the appropriate boundary value for $r_{\rm out}$. It turns out that the future event horizon intersects the past curvature singularity at $r \approx 8.657$. See Fig. 30

On the validity of the Gaussian normal coordinates for the modified metrics

In Sec. 6.4 we introduced a formalism in which the radial-radial component of the FRW or LTB metric had an additional time dependence. We implicitly assumed that the Gaussian normal coordinates used to define the FRW and LTB metrics can also be used to define the modified metrics. Here we sketch an argument showing the validity of this assumption.

Recall that the Gaussian normal coordinates can be constructed for a spacetime if and only if there exists a family of $C^{n\geq 1}$ timelike geodesics in the spacetime for which every point is located on

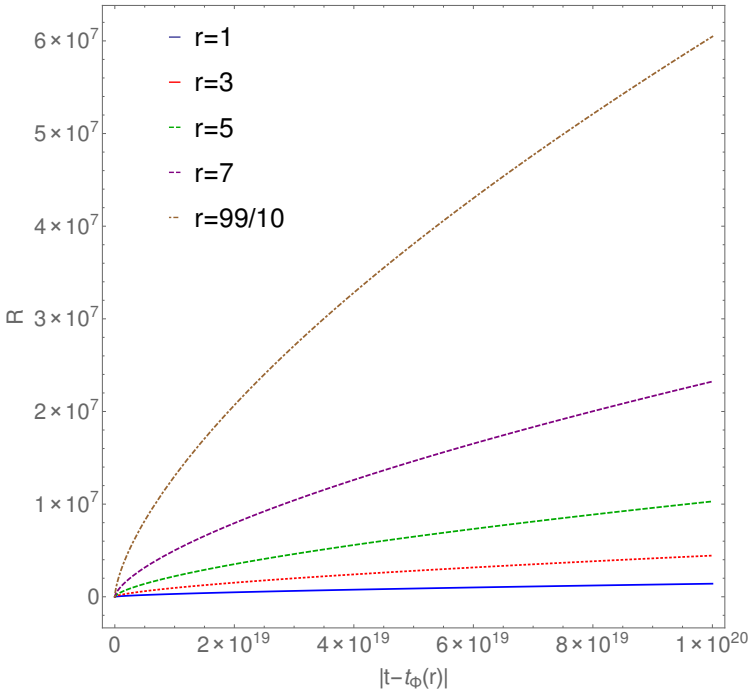


Figure 31: Here we have plotted the solution of Eq. (6.4.14) for a number of shells. As can be seen from the figure, shells expand without any crossing.

exactly one of these timelike geodesics. The coordinates are then constructed with $\vec{u} = \vec{\partial}_t$ being the unit tangent to each timelike geodesic. A trivial consequence of the existence of such a family of timelike geodesics is that there are no congruences of timelike geodesics with unit tangent $\vec{u} = \vec{\partial}_t$ that form a caustic anywhere in the spacetime [1].

The modified LTB metric is given by

$$ds^2 = -dt^2 + \frac{R'(t,r)^2 dr^2}{1 - r^2 k(t,r)} + R(t,r)^2 d\Omega^2, \tag{E.9}$$

where $r^2 k(t,r) < 1$ at all times and $\dot{k}(t,r) \neq 0$ for $-\infty < t < t_\Phi(r)$ [see Sec. 6.4.II]. Here we have implicitly assumed that the Gaussian normal coordinates used to define the LTB metric (6.3.1) can be used to define the modified LTB metric as well. To ensure that this can be done, we must show that the expansion $\theta \equiv {}^{(3)}g^{ab}\nabla_a u_b$ does not diverge for any congruence of timelike geodesics with

[1] Except at the curvature singularity, which is formally not part of the spacetime.

unit tangent $\vec{u} = \vec{\partial}_t$ [2]. A short calculation gives

$$\begin{aligned}\theta &= \frac{1}{2}{}^{(3)}g^{ab}\dot{g}_{ab} \\ &= \frac{\dot{R}'(t,r)}{R'(t,r)} + \frac{\dot{R}(t,r)}{R(t,r)} + \frac{r^2\dot{k}(t,r)}{2[1-r^2k(t,r)]}.\end{aligned} \tag{E.10}$$

Pursuant to our discussions in Sec. 6.4.I and 6.4.II with regards to the modified metric functions and their asymptotic behaviour, we do not expect any divergences to occur in (E.10) at any radii and time in the past. Therefore, we do not see a problem with the use of Gaussian normal coordinates for the modified FRW and LTB metrics.

Aerial radii do not coincide for distinct coordinate radii in the modified spacetime

If two aerial radii corresponding to two distinct coordinate radii coincide at some time to the past of $t_\Phi(r)$, the formalism of Sec. 6.4 would be invalidated. Numerically we can confirm that this does not happen for the solutions to the system of equations (6.4.14). See Fig. 31

[2] ${}^{(3)}g^{ab}$ is the inverse metric for constant t slices.

References

[1] Robert M Wald. *General relativity*. Chicago Univ. Press, Chicago, IL, 1984.

[2] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, 2016.

[3] M. Kowalski et al. Improved Cosmological Constraints from New, Old and Combined Supernova Datasets. *Astrophys. J.*, 686:749–778, 2008.

[4] Scott Burles, Kenneth M. Nollett, and Michael S. Turner. Big bang nucleosynthesis predictions for precision cosmology. *The Astrophysical Journal Letters*, 552(1):L1, 2001.

[5] N. A. Bahcall, R. Cen, R. Davé, J. P. Ostriker, and Q. Yu. The Mass-to-Light Function: Antibias and Ω_m. *apj*, 541:1–9, September 2000.

[6] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *apj*, 142:419–421, July 1965.

[7] P. J. E. Peebles. Primordial Helium Abundance and the Primordial Fireball. II. *apj*, 146:542, November 1966.

[8] YA.B. ZELDOVICH. Survey of modern cosmology. volume 3 of *Advances in Astronomy and Astrophysics*, pages 241 – 379. Elsevier, 1965.

[9] R. V. Wagoner, W. A. Fowler, and F. Hoyle. On the Synthesis of Elements at Very High Temperatures. *apj*, 148:3, April 1967.

[10] Steven Weinberg. The cosmological constant problem. *Rev. Mod. Phys.*, 61:1–23, Jan 1989.

[11] Daniel Baumann. Inflation. In *Physics of the large and the small, TASI 09*, pages 523–686, 2011.

[12] E. Komatsu et al. Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *Astrophys. J. Suppl.*, 180:330–376, 2009.

[13] R. Penrose. Difficulties with inflationary cosmology. *Annals of the New York Academy of Sciences*, 571:249–264, December 1989.

[14] S. Weinberg. *Cosmology*. Cosmology. OUP Oxford, 2008.

[15] G.D. Coughlan, W. Fischler, Edward W. Kolb, S. Raby, and G.G. Ross. Cosmological problems for the polonyi potential. *Physics Letters B*, 131(1):59 – 64, 1983.

[16] John P. Preskill. Cosmological production of superheavy magnetic monopoles. *Phys. Rev. Lett.*, 43:1365–1368, Nov 1979.

[17] J. Preskill. Magnetic Monopoles. *Annual Review of Nuclear and Particle Science*, 34:461–530, December 1984.

[18] Ya.B. Zeldovich and M.Yu. Khlopov. On the concentration of relic magnetic monopoles in the universe. *Physics Letters B*, 79(3):239 – 241, 1978.

[19] Alexei A. Starobinsky. A New Type of Isotropic Cosmological Models Without Singularity. *Phys. Lett.*, B91:99–102, 1980.

[20] Alan H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356, Jan 1981.

[21] A.D. Linde. A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters B*, 108(6):389 – 393, 1982.

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

[23] Sidney R. Coleman. The Fate of the False Vacuum. 1. Semiclassical Theory. *Phys. Rev.*, D15:2929–2936, 1977. [Erratum: Phys. Rev.D16,1248(1977)].

[24] Curtis G. Callan, Jr. and Sidney R. Coleman. The Fate of the False Vacuum. 2. First Quantum Corrections. *Phys. Rev.*, D16:1762–1768, 1977.

[25] Sidney Coleman and Frank De Luccia. Gravitational effects on and of vacuum decay. *Phys. Rev. D*, 21:3305–3315, Jun 1980.

[26] A.D. Linde. Chaotic inflation. *Physics Letters B*, 129(34):177 – 181, 1983.

[27] C. Armendariz-Picon, T. Damour, and Viatcheslav F. Mukhanov. k - inflation. *Phys.Lett.*, B458:209–218, 1999.

[28] David H. Lyth. Constraints on TeV scale hybrid inflation and comments on nonhybrid alternatives. *Phys. Lett.*, B466:85–94, 1999.

[29] P. A. R. Ade et al. Planck 2015 results. XX. Constraints on inflation. *Astron. Astrophys.*, 594:A20, 2016.

[30] Courtney M. Peterson and Max Tegmark. Testing multifield inflation: A geometric approach. *Phys. Rev.*, D87(10):103507, 2013.

[31] Justin Khoury, Burt A. Ovrut, Paul J. Steinhardt, and Neil Turok. The Ekpyrotic universe: Colliding branes and the origin of the hot big bang. *Phys. Rev.*, D64:123522, 2001.

[32] Abhay Ashtekar and Parampreet Singh. Loop Quantum Cosmology: A Status Report. *Class. Quant. Grav.*, 28:213001, 2011.

[33] Ben Freivogel, Matthew Kleban, Maria Rodriguez Martinez, and Leonard Susskind. Observational consequences of a landscape. *JHEP*, 03:039, 2006.

[34] John March-Russell and Francesco Riva. Signals of Inflation in a Friendly String Landscape. *JHEP*, 07:033, 2006.

[35] Daniel Baumann and Liam McAllister. *Inflation and String Theory*. Cambridge University Press, 2015.

[36] David H. Lyth and Antonio Riotto. Particle physics models of inflation and the cosmological density perturbation. *Phys. Rept.*, 314:1–146, 1999.

[37] L. Senatore. TASI 2012 Lectures on Inflation. In K. Matchev and et al., editors, *Searching for New Physics at Small and Large Scales (TASI 2012) - Proceedings of the 2012 Theoretical Advanced Study Institute in Elementary Particle Physics. Edited by Schmaltz Martin amp Pierpaoli Elena. Published by World Scientific Publishing Co. Pte. Ltd., 2013. ISBN 9789814525220, pp. 221-302*, pages 221–302, August 2013.

[38] Julien Lesgourgues. *Inflationary cosmology*.

[39] Andrei D. Linde. Recent progress in inflationary cosmology. *Lect. Notes Phys.*, 455:363–372, 1995. [,72(1994)].

[40] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger. Theory of cosmological perturbations. *physrep*, 215:203–333, June 1992.

[41] Viatcheslav Mukhanov. *Physical Foundations of Cosmology*. Cambridge Univ. Press, Cambridge, 2005.

[42] Andrew R. Liddle, Paul Parsons, and John D. Barrow. Formalizing the slow-roll approximation in inflation. *Phys. Rev. D*, 50:7222–7232, Dec 1994.

[43] Clifford Cheung, Paolo Creminelli, A. Liam Fitzpatrick, Jared Kaplan, and Leonardo Senatore. The Effective Field Theory of Inflation. *JHEP*, 03:014, 2008.

[44] Leonardo Senatore and Matias Zaldarriaga. The Effective Field Theory of Multifield Inflation. *JHEP*, 04:024, 2012.

[45] Steven Weinberg. Effective Field Theory for Inflation. *Phys. Rev.*, D77:123541, 2008.

[46] Edmund J. Copeland, Andrew R. Liddle, David H. Lyth, Ewan D. Stewart, and David Wands. False vacuum inflation with Einstein gravity. *Phys. Rev.*, D49:6410–6433, 1994.

[47] Liam McAllister, Eva Silverstein, and Alexander Westphal. Gravity Waves and Linear Inflation from Axion Monodromy. *Phys. Rev.*, D82:046003, 2010.

[48] Misao Sasaki. Gauge Invariant Scalar Perturbations in the New Inflationary Universe. *Prog.Theor.Phys.*, 70:394, 1983.

[49] Hideo Kodama and Misao Sasaki. Cosmological perturbation theory. *Progress of Theoretical Physics Supplement*, 78:1–166, 1984.

[50] Viatcheslav F. Mukhanov, L.A. Kofman, and D. Yu. Pogosian. Cosmological Perturbations in the Inflationary Universe. *Phys.Lett.*, B193:427–432, 1987.

[51] Steven Weinberg. Adiabatic modes in cosmology. *Phys. Rev. D*, 67:123504, Jun 2003.

[52] Steven Weinberg. Damping of tensor modes in cosmology. *Phys. Rev. D*, 69:023503, Jan 2004.

[53] R.M. Wald. *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. Chicago Lectures in Physics. University of Chicago Press, 1994.

[54] Viatcheslav F Mukhanov and Sergei Winitzki. *Introduction to quantum effects in gravity*. Cambridge Univ. Press, Cambridge, 2007.

[55] T. S. Bunch and P. C. W. Davies. Quantum field theory in de sitter space: Renormalization by point-splitting. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 360(1700):117–134, 1978.

[56] Robert H. Brandenberger and Jerome Martin. Trans-Planckian Issues for Inflationary Cosmology. *Class. Quant. Grav.*, 30:113001, 2013.

[57] Elisa G. M. Ferreira and Robert Brandenberger. The Trans-Planckian Problem in the Healthy Extension of Horava-Lifshitz Gravity. *Phys. Rev.*, D86:043514, 2012.

[58] Robert H. Brandenberger and Jerome Martin. The Robustness of inflation to changes in superPlanck scale physics. *Mod. Phys. Lett.*, A16:999–1006, 2001.

[59] Michael E Peskin and Daniel V Schroeder. *An Introduction to Quantum Field Theory; 1995 ed.* Westview, Boulder, CO, 1995. Includes exercises.

[60] Daniel Babich, Paolo Creminelli, and Matias Zaldarriaga. The Shape of non-Gaussianities. *JCAP*, 0408:009, 2004.

[61] P. A. R. Ade et al. Planck 2015 results. XVII. Constraints on primordial non-Gaussianity. *Astron. Astrophys.*, 594:A17, 2016.

[62] Juan Martin Maldacena. Non-Gaussian features of primordial fluctuations in single field inflationary models. *JHEP*, 05:013, 2003.

[63] Laura E. Allen, Sujata Gupta, and David Wands. Non-gaussian perturbations from multi-field inflation. *JCAP*, 0601:006, 2006.

[64] David H. Lyth. What would we learn by detecting a gravitational wave signal in the cosmic microwave background anisotropy? *Phys. Rev. Lett.*, 78:1861–1863, 1997.

[65] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *apj*, 142:419–421, July 1965.

[66] A. A. Penzias and R. W. Wilson. Measurement of the Flux Density of CAS a at 4080 Mc/s. *apj*, 142:1149, October 1965.

[67] G. Luzzi, R.T. Gnova-Santos, C.J.A.P. Martins, M. De Petris, and L. Lamagna. Constraining the evolution of the cmb temperature with sz measurements from planck data. *Journal of Cosmology and Astroparticle Physics*, 2015(09):011, 2015.

[68] P. J. E. Peebles and J. T. Yu. Primeval Adiabatic Perturbation in an Expanding Universe. *apj*, 162:815, December 1970.

[69] E. Dwek, R. G. Arendt, M. G. Hauser, D. Fixsen, T. Kelsall, D. Leisawitz, Y. C. Pei, E. L. Wright, J. C. Mather, S. H. Moseley, N. Odegard, R. Shafer, R. F. Silverberg, and J. L. Weiland. The COBE Diffuse Infrared Background Experiment Search for the Cosmic Infrared Background. IV. Cosmological Implications. *apj*, 508:106–122, November 1998.

[70] R. Adam et al. Planck intermediate results. XXX. The angular power spectrum of polarized dust emission at intermediate and high Galactic latitudes. *Astron. Astrophys.*, 586:A133, 2016.

[71] Scott Dodelson. *Modern cosmology.* Academic Press, San Diego, CA, 2003.

[72] Martin Bucher. Physics of the cosmic microwave background anisotropy. *Int. J. Mod. Phys.*, D24(02):1530004, 2015.

[73] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro. Large scale structure of the universe and cosmological perturbation theory. *Phys. Rept.*, 367:1–248, 2002.

[74] R. Hlozek, J. Dunkley, G. Addison, J. W. Appel, J. R. Bond, C. Sofia Carvalho, S. Das, M. J. Devlin, R. Dünner, T. Essinger-Hileman, J. W. Fowler, P. Gallardo, A. Hajian, M. Halpern, M. Hasselfield, M. Hilton, A. D. Hincks, J. P. Hughes, K. D. Irwin, J. Klein, A. Kosowsky, T. A. Marriage, D. Marsden, F. Menanteau, K. Moodley, M. D. Niemack, M. R. Nolta, L. A. Page, L. Parker, B. Partridge, F. Rojas, N. Sehgal, B. Sherwin, J. Sievers, D. N. Spergel, S. T. Staggs, D. S. Swetz, E. R. Switzer, R. Thornton, and E. Wollack. The Atacama Cosmology Telescope: A Measurement of the Primordial Power Spectrum. *apj*, 749:90, April 2012.

[75] P. A. R. Ade et al. Joint Analysis of BICEP2/*KeckArray* and *Planck* Data. *Phys. Rev. Lett.*, 114:101301, 2015.

[76] J. R. Pritchard and A. Loeb. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901, August 2012.

[77] J. P. Filippini, P. A. R. Ade, M. Amiri, S. J. Benton, R. Bihary, J. J. Bock, J. R. Bond, J. A. Bonetti, S. A. Bryan, B. Burger, H. C. Chiang, C. R. Contaldi, B. P. Crill, O. Doré, M. Farhang, L. M. Fissel, N. N. Gandilo, S. R. Golwala, J. E. Gudmundsson, M. Halpern, M. Hasselfield, G. Hilton, W. Holmes, V. V. Hristov, K. D. Irwin, W. C. Jones, C. L. Kuo, C. J. MacTavish, P. V. Mason, T. E. Montroy, T. A. Morford, C. B. Netterfield, D. T. O'Dea, A. S. Rahlin, C. D. Reintsema, J. E. Ruhl, M. C. Runyan, M. A. Schenker, J. A. Shariff, J. D. Soler, A. Trangsrud, C. Tucker, R. S. Tucker, and A. D. Turner. SPIDER: a balloon-borne CMB polarimeter for large angular scales. In *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy V*, volume 7741 of *procspie*, page 77411N, July 2010.

[78] M. D. Niemack, P. A. R. Ade, J. Aguirre, F. Barrientos, J. A. Beall, J. R. Bond, J. Britton, H. M. Cho, S. Das, M. J. Devlin, S. Dicker, J. Dunkley, R. Dünner, J. W. Fowler, A. Hajian, M. Halpern, M. Hasselfield, G. C. Hilton, M. Hilton, J. Hubmayr, J. P. Hughes, L. Infante, K. D. Irwin, N. Jarosik, J. Klein, A. Kosowsky, T. A. Marriage, J. McMahon, F. Menanteau,

K. Moodley, J. P. Nibarger, M. R. Nolta, L. A. Page, B. Partridge, E. D. Reese, J. Sievers, D. N. Spergel, S. T. Staggs, R. Thornton, C. Tucker, E. Wollack, and K. W. Yoon. ACTPol: a polarization-sensitive receiver for the Atacama Cosmology Telescope. In *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy V*, volume 7741 of *procspie*, page 77411S, July 2010.

[79] B. A. Benson, P. A. R. Ade, Z. Ahmed, S. W. Allen, K. Arnold, J. E. Austermann, A. N. Bender, L. E. Bleem, J. E. Carlstrom, C. L. Chang, H. M. Cho, J. F. Cliche, T. M. Crawford, A. Cukierman, T. de Haan, M. A. Dobbs, D. Dutcher, W. Everett, A. Gilbert, N. W. Halverson, D. Hanson, N. L. Harrington, K. Hattori, J. W. Henning, G. C. Hilton, G. P. Holder, W. L. Holzapfel, K. D. Irwin, R. Keisler, L. Knox, D. Kubik, C. L. Kuo, A. T. Lee, E. M. Leitch, D. Li, M. McDonald, S. S. Meyer, J. Montgomery, M. Myers, T. Natoli, H. Nguyen, V. Novosad, S. Padin, Z. Pan, J. Pearson, C. Reichardt, J. E. Ruhl, B. R. Saliwanchik, G. Simard, G. Smecher, J. T. Sayre, E. Shirokoff, A. A. Stark, K. Story, A. Suzuki, K. L. Thompson, C. Tucker, K. Vanderlinde, J. D. Vieira, A. Vikhlinin, G. Wang, V. Yefremenko, and K. W. Yoon. SPT-3G: a next-generation cosmic microwave background polarization experiment on the South Pole telescope. In *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII*, volume 9153 of *procspie*, page 91531P, July 2014.

[80] Luca Amendola et al. Cosmology and fundamental physics with the Euclid satellite. *Living Rev. Rel.*, 16:6, 2013.

[81] Amir Aghamousa et al. The DESI Experiment Part I: Science,Targeting, and Survey Design. 2016.

[82] Kevork N. Abazajian et al. CMB-S4 Science Book, First Edition. 2016.

[83] Jan Niklas Grieb et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological implications of the Fourier space wedges of the final sample. *Mon. Not. Roy. Astron. Soc.*, 2016.

[84] Qing-Guo Huang, Sai Wang, and Wen Zhao. Forecasting sensitivity on tilt of power spectrum of primordial gravitational waves after Planck satellite. *JCAP*, 1510(10):035, 2015.

[85] Ivan Agullo and Leonard Parker. Non-gaussianities and the Stimulated creation of quanta in the inflationary universe. *Phys.Rev.*, D83:063526, 2011.

[86] Nishant Agarwal, R. Holman, Andrew J. Tolley, and Jennifer Lin. Effective field theory and non-Gaussianity from general inflationary states. 2012.

[87] F. Nitti, M. Porrati, and J.-W. Rombouts. Naturalness in cosmological initial conditions. *Phys.Rev.*, D72:063503, 2005.

[88] M. Porrati. Effective field theory approach to cosmological initial conditions: Self-consistency bounds and non-Gaussianities. 2004.

[89] R. Holman and Andrew J. Tolley. Enhanced Non-Gaussianity from Excited Initial States. *JCAP*, 0805:001, 2008.

[90] Pieter Daniel Meerburg, Jan Pieter van der Schaar, and Pier Stefano Corasaniti. Signatures of Initial State Modifications on Bispectrum Statistics. *JCAP*, 0905:018, 2009.

[91] P. Daniel Meerburg, Jan Pieter van der Schaar, and Mark G. Jackson. Bispectrum signatures of a modified vacuum in single field inflation with a small speed of sound. *JCAP*, 1002:001, 2010.

[92] Jonathan Ganc. Calculating the local-type fNL for slow-roll inflation with a non-vacuum initial state. *Phys.Rev.*, D84:063514, 2011.

[93] Ivan Agullo and Sarah Shandera. Large non-Gaussian Halo Bias from Single Field Inflation. *JCAP*, 1209:007, 2012.

[94] Amjad Ashoorioon, Konstantinos Dimopoulos, M.M. Sheikh-Jabbari, and Gary Shiu. Reconciliation of High Energy Scale Models of Inflation with Planck. 2013.

[95] Aditya Aravind, Dustin Lorshbough, and Sonia Paban. Non-Gaussianity from Excited Initial Inflationary States. *JHEP*, 1307:076, 2013.

[96] Sandipan Kundu. Inflation with General Initial Conditions for Scalar Perturbations. *JCAP*, 1202:005, 2012.

[97] Suratna Das and Subhendra Mohanty. Non-Gaussianity as a signature of thermal initial condition of inflation. *Phys.Rev.*, D80:123537, 2009.

[98] Amjad Ashoorioon and Gary Shiu. A Note on Calm Excited States of Inflation. *JCAP*, 1103:025, 2011.

[99] Kevin Goldstein and David A. Lowe. Initial state effects on the cosmic microwave background and transPlanckian physics. *Phys.Rev.*, D67:063502, 2003.

[100] Hael Collins and R. Holman. Trans-Planckian enhancements of the primordial non-Gaussianities. *Phys.Rev.*, D80:043524, 2009.

[101] Richard Easther, Brian R. Greene, William H. Kinney, and Gary Shiu. Inflation as a probe of short distance physics. *Phys.Rev.*, D64:103502, 2001.

[102] Richard Easther, Brian R. Greene, William H. Kinney, and Gary Shiu. Imprints of short distance physics on inflationary cosmology. *Phys.Rev.*, D67:063508, 2003.

[103] Robert H. Brandenberger and Jerome Martin. On signatures of short distance physics in the cosmic microwave background. *Int.J.Mod.Phys.*, A17:3663–3680, 2002.

[104] Robert Brandenberger and Pei-Ming Ho. Noncommutative space-time, stringy space-time uncertainty principle, and density fluctuations. *Phys.Rev.*, D66:023517, 2002.

[105] Fedele Lizzi, Gianpiero Mangano, Gennaro Miele, and Marco Peloso. Cosmological perturbations and short distance physics from noncommutative geometry. *JHEP*, 0206:049, 2002.

[106] Richard Easther, Brian R. Greene, William H. Kinney, and Gary Shiu. A Generic estimate of transPlanckian modifications to the primordial power spectrum in inflation. *Phys.Rev.*, D66:023518, 2002.

[107] Ulf H. Danielsson. Inflation, holography, and the choice of vacuum in de Sitter space. *JHEP*, 0207:040, 2002.

[108] Nemanja Kaloper, Matthew Kleban, Albion E. Lawrence, and Stephen Shenker. Signatures of short distance physics in the cosmic microwave background. *Phys.Rev.*, D66:123510, 2002.

[109] Raphael Flauger, Daniel Green, and Rafael A. Porto. On squeezed limits in single-field inflation. Part I. *JCAP*, 1308:032, 2013.

[110] P. R. Anderson, C. Molina-París, and E. Mottola. Short distance and initial state effects in inflation: Stress tensor and decoherence. *Phys. Rev. D*, 72(4):043515, August 2005.

[111] A. Dey, E. Kovetz, and S. Paban. Non-Gaussianities in the cosmological perturbation spectrum due to primordial anisotropy II. *JCAP*, 10:55, October 2012.

[112] Mark G. Jackson. Integrating out Heavy Fields in Inflation. 2012.

[113] Katelin Schutz, Evangelos I Sfakianakis, and David I Kaiser. Multifield inflation after planck: Isocurvature modes from nonminimal couplings. *arXiv preprint arXiv:1310.8285*, 2013.

[114] Carlo R. Contaldi, Marco Peloso, Lev Kofman, and Andrei D. Linde. Suppressing the lower multipoles in the CMB anisotropies. *JCAP*, 0307:002, 2003.

[115] Ben Freivogel, Matthew Kleban, Maria Rodriguez Martinez, and Leonard Susskind. Observational consequences of a landscape: Epilogue. *arXiv preprint arXiv:1404.2274*, 2014.

[116] Raphael Bousso, Daniel Harlow, and Leonardo Senatore. Inflation After False Vacuum Decay: New Evidence from BICEP2. 2014.

[117] Alexei A. Starobinsky. Dynamics of Phase Transition in the New Inflationary Universe Scenario and Generation of Perturbations. *Phys. Lett.*, B117:175–178, 1982.

[118] Alexander Vilenkin. Gravitational Effects in Guth Cosmology. *Phys. Lett.*, B115:91, 1982.

[119] Alexander Vilenkin and L. H. Ford. Gravitational effects upon cosmological phase transitions. *Phys. Rev. D*, 26:1231–1241, Sep 1982.

[120] J. David Brown. Action functionals for relativistic perfect fluids. *Class.Quant.Grav.*, 10:1579–1606, 1993.

[121] Richard Easther, Brian R. Greene, William H. Kinney, and Gary Shiu. Inflation as a probe of short distance physics. *Phys.Rev.*, D64:103502, 2001.

[122] Ulf H. Danielsson. A Note on inflation and transPlanckian physics. *Phys.Rev.*, D66:023511, 2002.

[123] Jerome Martin and Robert Brandenberger. On the dependence of the spectra of fluctuations in inflationary cosmology on transPlanckian physics. *Phys.Rev.*, D68:063513, 2003.

[124] V. Bozza, Massimo Giovannini, and G. Veneziano. Cosmological perturbations from a new physics hypersurface. *JCAP*, 0305:001, 2003.

[125] Raphael Flauger, Liam McAllister, Enrico Pajer, Alexander Westphal, and Gang Xu. Oscillations in the CMB from Axion Monodromy Inflation. *JCAP*, 1006:009, 2010.

[126] Alexei A. Starobinsky. Spectrum of adiabatic perturbations in the universe when there are singularities in the inflation potential. *JETP Lett.*, 55:489–494, 1992.

[127] Paul Hunt and Subir Sarkar. Multiple inflation and the WMAP 'glitches'. *Phys.Rev.*, D70:103518, 2004.

[128] Daniel J.H. Chung, Edward W. Kolb, Antonio Riotto, and Igor I. Tkachev. Probing Planckian physics: Resonant production of particles during inflation and features in the primordial power spectrum. *Phys.Rev.*, D62:043508, 2000.

[129] Jennifer A. Adams, Bevan Cresswell, and Richard Easther. Inflationary perturbations from a potential with a step. *Phys.Rev.*, D64:123514, 2001.

[130] Amjad Ashoorioon and Axel Krause. Power Spectrum and Signatures for Cascade Inflation. 2006.

[131] Amjad Ashoorioon, Axel Krause, and Krzysztof Turzynski. Energy Transfer in Multi Field Inflation and Cosmological Perturbations. *JCAP*, 0902:014, 2009.

[132] Ana Achucarro, Jinn-Ouk Gong, Sjoerd Hardeman, Gonzalo A. Palma, and Subodh P. Patil. Features of heavy physics in the CMB power spectrum. *JCAP*, 1101:030, 2011.

[133] Minjoon Park and Lorenzo Sorbo. Sudden variations in the speed of sound during inflation: features in the power spectrum and bispectrum. *Phys. Rev.*, D85:083520, 2012.

[134] Masahiro Nakashima, Ryo Saito, Yu-ichi Takamizu, and Jun'ichi Yokoyama. The effect of varying sound velocity on primordial curvature perturbations. *Prog. Theor. Phys.*, 125:1035–1052, 2011.

[135] Rita Sinha and Tarun Souradeep. Post-wmap assessment of infrared cutoff in the primordial spectrum from inflation. *Phys.Rev.*, D74:043518, 2006.

[136] Shiro Hirai and Tomoyuki Takami. Effect of the length of inflation on angular tt and te power spectra in power-law inflation. *Class. Quant. Grav.*, 23:2541–2558, 2006.

[137] Brian A. Powell and William H. Kinney. The pre-inflationary vacuum in the cosmic microwave background. *Phys. Rev.*, D76:063512, 2007.

[138] I-Chin Wang and Kin-Wang Ng. Effects of a pre-inflation radiation-dominated epoch to CMB anisotropy. *Phys. Rev.*, D77:083501, 2008.

[139] A. Ashoorioon, J. L. Hovdebo, and Robert B. Mann. Running of the spectral index and violation of the consistency relation between tensor and scalar spectra from trans-Planckian physics. *Nucl. Phys.*, B727:63–76, 2005.

[140] V. A. Fock. *Theory of space, time and gravitation.* Nauka, Moscow, 1955, 1961 (second ed.). In Russian. English translation: Pergamon Press, New York-London, 1959; German Translation: V. Fock, Theorie von Raum, Zeit und Gravitation, Akademie-Verlag, Berlin, 1960.

[141] S. W. Hawking. Particle creation by black holes. *Comm. Math. Phys.*, 43(3):199–220, 1975.

[142] S. W. Hawking. Black holes and thermodynamics. *Phys. Rev. D*, 13:191–197, Jan 1976.

[143] James M Bardeen, Brandon Carter, and Stephen W Hawking. The four laws of black hole mechanics. *Communications in Mathematical Physics*, 31(2):161–170, 1973.

[144] Stephen W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Space-Time (Cambridge Monographs on Mathematical Physics).* Cambridge University Press, 1975.

[145] Ted Jacobson. Introductory lectures on black hole thermodynamics.

[146] Robert M. Wald. The First law of black hole mechanics. In *Directions in General Relativity: An International Symposium in Honor of the 60th Birthdays of Dieter Brill and Charles Misner College Park, Maryland, May 27-29, 1993*, pages 358–366, 1993.

[147] Werner Israel. Third law of black-hole dynamics: a formulation and proof. *Physical review letters*, 57(4):397, 1986.

[148] Ted Jacobson, Gungwon Kang, and Robert C. Myers. On black hole entropy. *Phys. Rev.*, D49:6587–6598, 1994.

[149] G. W. Gibbons and S. W. Hawking. Action integrals and partition functions in quantum gravity. *Phys. Rev. D*, 15:2752–2756, May 1977.

[150] Robert M. Wald. Black hole entropy is the Noether charge. *Phys. Rev.*, D48(8):R3427–R3431, 1993.

[151] Vivek Iyer and Robert M. Wald. Some properties of Noether charge and a proposal for dynamical black hole entropy. *Phys. Rev.*, D50:846–864, 1994.

[152] Robert M. Wald. The thermodynamics of black holes. *Living Rev. Rel.*, 4:6, 2001.

[153] Jacob D. Bekenstein. Black holes and entropy. *Phys. Rev.*, D7:2333–2346, 1973.

[154] W. H. Zurek and Kip S. Thorne. Statistical mechanical origin of the entropy of a rotating, charged black hole. *Phys. Rev. Lett.*, 54:2171–2175, May 1985.

[155] Gerard 't Hooft. On the quantum structure of a black hole. *Nuclear Physics B*, 256:727 – 745, 1985.

[156] Curtis G. Callan, Jr. and Frank Wilczek. On geometric entropy. *Phys. Lett.*, B333:55–61, 1994.

[157] Christoph Holzhey, Finn Larsen, and Frank Wilczek. Geometric and renormalized entropy in conformal field theory. *Nucl. Phys.*, B424:443–467, 1994.

[158] Jacob D. Bekenstein. NonArchimedean character of quantum buoyancy and the generalized second law of thermodynamics. *Phys. Rev.*, D60:124010, 1999.

[159] A. Ashtekar, J. Baez, A. Corichi, and Kirill Krasnov. Quantum geometry and black hole entropy. *Phys. Rev. Lett.*, 80:904–907, 1998.

[160] Abhay Ashtekar and Kirill Krasnov. Quantum geometry and black holes. 1998.

[161] Andrew Strominger. Black hole entropy from near horizon microstates. *JHEP*, 02:009, 1998.

[162] Jacob D. Bekenstein. Generalized second law of thermodynamics in black-hole physics. *Phys. Rev. D*, 9:3292–3300, Jun 1974.

[163] Jacob D. Bekenstein. Universal upper bound on the entropy-to-energy ratio for bounded systems. *Phys. Rev. D*, 23:287–298, Jan 1981.

[164] Jacob D. Bekenstein. Entropy content and information flow in systems with limited energy. *Phys. Rev. D*, 30:1669–1679, Oct 1984.

[165] Jacob D. Bekenstein and Marcelo Schiffer. Quantum limitations on the storage and transmission of information. *Int. J. Mod. Phys.*, C1:355–422, 1990.

[166] Raphael Bousso. A Covariant entropy conjecture. *JHEP*, 07:004, 1999.

[167] Raphael Bousso. The Holographic principle. *Rev. Mod. Phys.*, 74:825–874, 2002.

[168] Raphael Bousso, Eanna E. Flanagan, and Donald Marolf. Simple sufficient conditions for the generalized covariant entropy bound. *Phys. Rev.*, D68:064001, 2003.

[169] Eanna E. Flanagan, Donald Marolf, and Robert M. Wald. Proof of classical versions of the Bousso entropy bound and of the generalized second law. *Phys. Rev.*, D62:084035, 2000.

[170] Gia Dvali and Cesar Gomez. Black Holes as Critical Point of Quantum Phase Transition. *Eur. Phys. J.*, C74:2752, 2014.

[171] Rafael D. Sorkin, Robert M. Wald, and Zhen Jiu Zhang. Entropy of selfgravitating radiation. *Gen. Rel. Grav.*, 13:1127–1146, 1981.

[172] Stephen D. H. Hsu and David Reeb. Black hole entropy, curved space and monsters. *Phys. Lett.*, B658:244–248, 2008.

[173] Stephen D. H. Hsu and David Reeb. Monsters, black holes and the statistical mechanics of gravity. *Mod. Phys. Lett.*, A24:1875–1887, 2009.

[174] Douglas M. Eardley. Death of white holes in the early universe. *Phys. Rev. Lett.*, 33:442–444, Aug 1974.

[175] Robert M. Wald and Sriram Ramaswamy. Particle production by white holes. *Phys. Rev. D*, 21:2736–2741, May 1980.

[176] Robert M. Wald. Quantum gravity and time reversibility. *Phys. Rev. D*, 21:2742–2755, May 1980.

[177] H. Bondi. Spherically symmetrical models in general relativity. *Monthly Notices of the Royal Astronomical Society*, 107(5-6):410–425, 1947.

[178] Marie-Noelle Celerier. Do we really see a cosmological constant in the supernovae data? *Astron. Astrophys.*, 353:63–71, 2000.

[179] R. Ali Vanderveld, Eanna E. Flanagan, and Ira Wasserman. Mimicking dark energy with Lemaitre-Tolman-Bondi models: Weak central singularities and critical points. *Phys. Rev.*, D74:023506, 2006.

[180] Eric Poisson. *A relativist's toolkit : the mathematics of black-hole mechanics*. Cambridge, UK: Cambridge University Press,, 2004.

[181] Raphael Bousso, Zachary Fisher, Stefan Leichenauer, and Aron C. Wall. Quantum focusing conjecture. *Phys. Rev.*, D93(6):064044, 2016.

[182] C.W. Misner, K.S. Thorne, and J.A. Wheeler. *Gravitation*. W.H. Freeman and Company, 1973.