

Copyright
by
Raphael Manfred Flauger
2009

The Dissertation Committee for Raphael Manfred Flauger
certifies that this is the approved version of the following dissertation:

Constraining Fundamental Physics with Cosmology

Committee:

Steven Weinberg, Supervisor

Arno Bohm

Willy Fischler

Dan Freed

Sonia Paban

Constraining Fundamental Physics with Cosmology

by

Raphael Manfred Flauger, M.A.; M.Sc.

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2009

To my parents

Acknowledgments

First and foremost, I would like to express my sincere gratitude to Steven Weinberg for his supervision during these past years and thank him for his words of advice and guidance whenever they were needed. His immense knowledge as well as his dedication to theoretical physics have been a great inspiration to me.

I owe great debt also to the other professors in the Theory Group. I greatly enjoyed the many insightful and entertaining hours spent discussing physics together with Willy Fischler on the blackboard in his office, and I am grateful to Sonia Paban for listening to me patiently whenever I had come across something I found confusing. This thesis would most certainly also not have been possible without a course on string theory taught by Jacques Distler, and Vadim Kaplunovsky's meticulously detailed lectures on supersymmetry. It is a pleasure to thank Dan Freed for his crystal clear lectures on Riemannian Geometry and the Atiyah-Singer index theorem.

I am indebted to my collaborators Daniel Baumann, Elena Caceres, Sergei Dubovsky, Matthias Ihl, Mark Jackson, Liam McAllister, Enrico Pajer, Igor Tkachev, Eva Silverstein, Alexei Starobinsky, Alexander Westphal, Timm Wrase, Gang Xu for many enlightening discussions and would like to thank them for the fruitful collaborations.

I would also like to thank Jan Duffy for taking care of bureaucratic matters so efficiently, and Terrence Riley for keeping the computer network running smoothly at all times.

Last but not least, I would like to thank my parents for their support.

Constraining Fundamental Physics with Cosmology

Publication No. _____

Raphael Manfred Flauger, Ph.D.
The University of Texas at Austin, 2009

Supervisor: Steven Weinberg

It is shown in three examples that future cosmological data may allow us to constrain fundamental physics in interesting ways.

The first example illustrates that correlations in the polarization of the cosmic microwave background may allow us to put the strongest limit yet on the mass of a particle, the graviton, at a level of $m \lesssim 10^{-30}$ eV.

In the second example, it is shown that observations of the correlations of temperature anisotropies and polarization of the cosmic microwave background may reveal hints for the realization of a class of string theoretic inflationary models that go by the name of axion monodromy inflation, or, rule them out. If the evidence for inflation strengthens substantially, just the requirement that inflation occurred may be used to constrain models of fundamental physics. The third example shows that a class of string compactifications that are commonly used in the context of string phenomenology cannot support inflation and might thus be ruled out by cosmology.

For completeness, a review of the physics underlying the cosmic microwave background radiation is included and some analytical results for the signatures of primordial gravitational waves in the cosmic microwave background are given.

Table of Contents

Acknowledgments	v
Abstract	vi
Chapter 1. Introduction	1
Chapter 2. A Brief Review of CMB Physics	7
2.1 Observables – Intensity Matrix and Stokes Parameters	11
2.2 Comparison with Theory – The Multipole Coefficients	18
2.3 Derivation of the Multipole Coefficients	24
2.3.1 Linear perturbation theory in a flat FRW universe	26
2.3.2 The Boltzmann equation for photons	36
2.3.2.1 Scalar modes	41
2.3.2.2 Tensor modes	47
2.3.3 The Boltzmann equation for neutrinos	50
2.3.3.1 Scalar modes	52
2.3.3.2 Tensor modes	54
2.3.4 The Multipole Coefficients	55
2.3.4.1 Scalar modes	56
2.3.4.2 Tensor modes	64
2.4 Inflation and the Primordial Power Spectrum	69
2.4.1 The Background Evolution	74
2.4.2 Primordial spectrum for scalar modes	77
2.4.3 Primordial spectrum for tensor modes	86
Chapter 3. Tensor Microwave Background Fluctuations for Large Multipole Order	94
3.1 The Large- ℓ Approximation	97
3.2 Parameter-Dependence of the EE and BB Correlations	103

Chapter 4. Signatures of a Graviton Mass in the Cosmic Microwave Background	113
4.1 Tensor Contribution to the CMB for a Massive Graviton . . .	118
4.1.1 Analytic results for low multipole coefficients	121
4.1.2 Numerical results	132
4.2 Summary	138
Chapter 5. Oscillations in the CMB from Axion Monodromy Inflation	141
5.1 Review of axion monodromy inflation	144
5.2 Background Evolution	147
5.3 Spectrum of Scalar Perturbations	150
5.3.1 Analytic solution of the Mukhanov-Sasaki equation . . .	152
5.3.2 Saddle-point approximation	156
5.3.3 Bispectrum of scalar perturbations	160
5.4 Observational Constraints	163
5.5 Microphysics of Axion Monodromy Inflation	169
5.5.1 Axions in string theory	170
5.5.2 Dimensional reduction and moduli stabilization	171
5.5.2.1 Four-dimensional data of O3-O7 orientifolds . .	171
5.5.2.2 Nonperturbative stabilization of the Kähler moduli	173
5.5.2.3 Nonperturbative breaking of axionic shift sym-	
metries	175
5.5.3 Axion decay constants in string theory	179
5.5.3.1 Decay constants in terms of $\mathcal{N} = 1$ data	179
5.5.3.2 An example: a complex plane of fixed points . .	181
5.6 Microscopic Constraints	182
5.6.1 Constraints from computability	183
5.6.2 Constraints from backreaction on the geometry	186
5.6.3 Constraints from higher-derivative terms	189
5.6.4 Constraints on the axion decay constant	192
5.6.5 Constraints on the amplitude of the modulations	194
5.7 Combined Theoretical and Observational Constraints	199
5.8 Summary	205

Chapter 6. On Slow-Roll Moduli Inflation in Massive IIA Supergravity with Metric Fluxes	209
6.1 Low Energy Theory	215
6.1.1 Metric and non-geometric fluxes	215
6.1.2 Effective potential	218
6.1.3 Slow-roll parameters	222
6.1.4 Scalar potential with metric fluxes	224
6.2 No-Go Theorems	224
6.2.1 General manifolds	225
6.2.2 Factorization in the Kähler sector	226
6.2.3 Factorization in the complex structure sector	230
6.2.4 Factorization in both sectors	232
6.3 Toroidal Orientifolds	233
6.3.1 Classification of orientifolds	233
6.3.2 Turning on NSNS fluxes	238
6.3.3 Residual symmetries	243
6.4 Application of the No-Go Theorems to Toroidal Orientifolds	244
6.4.1 Case I	246
6.4.2 Case II	247
6.4.3 Case III	247
6.4.4 Case IV	248
6.4.5 Case V	248
6.4.6 Case VI	249
6.4.7 Case VII	250
6.4.8 Case VIII	250
6.4.9 Case IX	251
6.4.10 Case X	252
6.4.11 Case XI	253
6.5 Examples with Small ϵ	254
6.5.1 Standard orientifold of $\mathbb{Z}_2 \times \mathbb{Z}_2$	255
6.5.2 Non-standard orientifold of $\mathbb{Z}_2 \times \mathbb{Z}_2$	257
6.6 Summary	259

Chapter 7. Conclusions	262
Appendices	264
Appendix A. Conventions and Notations	265
A.1 Conventions for IIB	265
A.2 Conventions for IIA	266
Appendix B. Phase Conventions for Spherical Harmonics	270
Appendix C. Induced Shift of the Four-Cycle Volume	274
C.0.1 A simple illustration of the suppression mechanism . . .	277
C.0.2 The conifold and its resolution	280
C.0.2.1 The λUY embedding	282
C.0.3 The shift of the four-cycle volume	283
Appendix D. The Kaluza-Klein Spectrum	287
D.1 The effective theory	287
D.2 Effects of the light Kaluza-Klein modes	289
Appendix E. Numerical Examples	292
E.1 Intersection numbers: set I	292
E.2 Intersection numbers: set II	293
Bibliography	294
Vita	321

Chapter 1

Introduction

It might almost be viewed as one of the traits that make us human that we try to understand where we, and the world around us, came from. While the degree of sophistication may vary, most likely everyone has wondered about it at some point. Contemplating about this does not take up a significant amount of time for most people. Theoretical high energy physicists, however, spend a fair amount of their time thinking about these and related issues and call the study of the origin and the evolution of the universe cosmology.¹ Until rather recently, very little experimental data had been available, and cosmology was considered by many to be closer to philosophy than to science. The situation has changed quite dramatically over the course of the past twenty years, and the measurement of the temperature of the cosmic microwave background radiation by a satellite called the COsmic Background Explorer, or COBE, is often regarded as the milestone marking the beginning of what is sometimes called the era of precision cosmology. Since then a large number of different experiments such as observations of distant supernovae,² more detailed measurements of the cosmic microwave background radiation, as well as galaxy surveys have been performed and have led us to a coherent

¹It is amusing to note that even though the origin of the word is Greek, it seems to have been used first by a German mathematician and philosopher, Christian Wolff, in his *Cosmologia Generalis* in 1730.

²A supernova is the explosion of a star.

picture of the early universe that is usually referred to as the standard or concordance model of cosmology.

It seems fair to say that we still know nothing about the very beginning, and we could not even disprove that it all began with a void,³ as Hesiod writes in his *Theogony*. We are beginning, however, to understand the subsequent evolution with some confidence, and it turns out to involve much less violence and bloodshed than Hesiod's version. But it is certainly no less interesting.

It now seems likely that the universe one way or another entered a period of accelerated expansion called inflation. During this period the universe grew or rather exploded in size by at least some thirty orders of magnitude in as little as 10^{-34} seconds, explaining among other things why the universe we see around us is so big. Quantum fluctuations during this era are believed to be the seeds for the structures in the universe we see today such as stars and galaxies. This period of inflation ends with a process called reheating during which the standard model particles such as quarks, leptons, and gauge bosons, as well as possibly others that we have not observed yet, were first created. By the time this process was finished, the universe was filled with a very hot and dense plasma. As the universe expanded, this plasma of elementary particles, sometimes also referred to as the primordial soup, cooled down and every time the temperature dropped below the mass of some species of particles, these particles annihilated with their antiparticles and disappeared from the plasma. Sometime after reheating but still before the end of an event known as the electroweak phase transition, a slight asymmetry between matter and anti-matter must have been produced leaving a tiny bit of matter behind when a species

³Hesiod talks about *Χάος*, which is often translated as Chaos. The word originally merely stands for a void or a chasm and does not imply any disorder or confusion.

of particles annihilated, explaining why we, as well as stars and galaxies are around. This is commonly referred to as baryogenesis. When the temperature dropped below some $10^{12} K$, which happened some $100 \mu s$ after the big bang, protons and neutrons made their first appearance in the universe, and as the universe continued to expand and cool, the energies dropped below the binding energies of nuclei, and the light elements were created in the universe. This era is referred to as nucleosynthesis and ended about three minutes after the big bang. After its completion, the universe was filled with a plasma containing protons, helium nuclei, small traces of deuterium and lithium, electrons, photons, neutrinos, and a component called dark matter as well as one called dark energy must also have been present. The protons, helium nuclei, electrons, and photons still interacted very efficiently with each other until the temperature was low enough for atoms to form. At this point, which occurred about 380,000 years after the big bang, the universe was filled by a neutral gas of hydrogen and helium. The photons decoupled and have been traveling freely throughout the universe since then. These photons, known as the cosmic microwave background, can be detected and will play a central role in this thesis. Some time not too long after this, the first stars must have formed out of the gas of hydrogen and helium. Gravitational interactions caused the original density fluctuations to grow, leading to regions that decoupled from the Hubble expansion called minihaloes. At the centers of these minihaloes the densities became high enough for fusion to set in, leading to the first stars. These are thought to be very massive and short lived, and are thought to collapse to a black hole or lead to a violent supernova explosion depending on their masses. The explosions of the first generation of stars enriched the medium with the heavier elements that were generated in their interiors, and from this medium a second and third generation of stars formed. These are

the stars we see in the night-sky today.

While many details especially regarding star and galaxy formation still need to be better understood, and it remains to unravel what the microphysical origin of dark matter and dark energy are, most of the evolution of the universe as described above that happened after nucleosynthesis is reasonably well supported by experiment. The evolution of the universe before nucleosynthesis is more speculative and is obtained by extrapolating our theories backwards in time using information we have obtained from experiments at particle accelerator experiments.

In this work, the cosmic microwave background will play a central role. It will be shown on examples that current and future observations of this radiation may allow us to constrain fundamental physics in interesting ways. We do not assume that the reader is familiar with the physics of the cosmic microwave background radiation, and give a brief review containing all relevant formulas in Chapter 2.

It will become clear that a certain pattern in the polarization of the cosmic microwave background called B-mode polarization is of particular interest because, at linear order in perturbation theory, it can only be generated by gravitational waves. An observation of this pattern would thus be an indirect measurement of gravitational radiation in the early universe. As explained in Chapter 2, inflation predicts that a gravitational wave background should be present, and may be large enough to be observed. The cosmic microwave background may thus provide us with a chance to strengthen the evidence for inflation and allow us to discriminate between different inflationary models.

In Chapter 3, we give analytical results for the prediction made for this signal by inflation. To be specific, for now postponing a definition of the various

quantities to Chapter 2, we give approximate formulas for the tensor contributions to the multipole coefficients $C_{TT,\ell}$, $C_{TE,\ell}$, $C_{EE,\ell}$, and $C_{BB,\ell}$ for large multipole order ℓ . These expressions are helpful for an understanding of the dependence of these multipole coefficients on cosmological parameters.

In Chapter 4, it is shown that the observation of a B-mode signal consistent with the prediction of inflation in Einstein gravity would severely constrain modified gravity theories known as massive gravity. The result is that in the case of a detection of a signal compatible with the inflationary prediction, the graviton mass must be less than $10^{-30} eV$. It applies for any such modification that gives a mass to the transverse, traceless degrees of freedom of the metric. In the context of brane inflation it has been established that no observable B-mode signal could be generated. This was boldly extrapolated and claims were made that an observation of a B-mode signal would rule out string theory. The model treated in Chapter 5, axion monodromy inflation, is a counter example. This scenario leads to an observable tensor signal and may furthermore lead to periodic modulations in the temperature multipole coefficients $C_{TT,\ell}$. These modulations are constrained both by comparison to data and by studying the microphysical constraints in this scenario. It is found that this class of models may lead to periodic modulations observable by future experiments in both two- and three-point functions.

In Chapter 6, different from the other chapters, there is no direct contact with experiment. No-go theorems are derived that show that inflation cannot be realized in certain compactifications of massive type IIA string theory to four dimensions. If the evidence for inflation becomes stronger, this will be enough to rule out this class of solutions. This is just a small class of models within the vast landscape of solutions of string theory, and there is little doubt that some solutions of the ten-dimensional equations of motion exist that do inflate,

but it may still help us narrow down the search somewhat.

This introduction was kept rather brief and contains no technical details. The reader interested in those will find more detailed introductions to the various topics at the beginning of each chapter.

Chapter 2

A Brief Review of CMB Physics

The cosmic microwave background radiation is most likely our best piece of evidence that the universe was once in a very hot and dense state called the Big Bang. The existence of this radiation had been predicted in a series of papers in the late 1940s by Gamow, Alpher, and Herman [1], [2], [3], [4], [5], [6], [5], [7], [8], [9], [10] as a consequence of their model of the formation of the elements during a hot, dense phase of the universe. Presumably because technology at the time was not advanced enough to detect this radiation, because the model had difficulties explaining the creation of elements heavier than lithium, and because Gamow, Alpher, and Herman left the field, the prediction soon passed out of people's minds.

In 1965 Dicke, Peebles, Roll, and Wilkinson in Princeton finally began to look for this remnant of the Big Bang, but it was first discovered somewhat accidentally by Penzias and Wilson as noise in their antenna [11], [12]. Since these first experiments, which observed the radiation to have the same intensity everywhere in the sky, a two-digit number of experiments have been dedicated to the study of the cosmic microwave background radiation. The first indication of a dipole isotropy due to the solar system's motion with respect to the cosmic microwave background was found in an experiment in the late 1960s by Conklin [13] and confirmed in the early 1970s by Henry [14]. Anisotropies beyond the dipole pattern were expected to be present as they were necessary to

understand how structures like galaxies and clusters of galaxies formed. This expectation was confirmed beautifully by the COsmic Background Explorer (COBE) [15].

Soon after the discovery of the cosmic microwave background radiation it was also predicted that this radiation should to some extent be polarized [16]. This prediction was confirmed by the DASI experiment [17], [18], [19]. Their results were confirmed by CBI [20] and CAPMAP [21] as well as more recently by Boomerang [22], [23], [24]. The experiments DASI, CBI, and Boomerang, and of course WMAP also observed correlations between temperature perturbations and polarization. The data taken by these experiments as well as supernova observations made by the Supernova Search Team and the Supernova Cosmology Project [25], [26], [27], [28] as well as large scale structure surveys such as 2dF [29], [30], [31], [32] and the Sloan Digital Sky Survey [33], [34],[35], [36] have led us to a very coherent picture for the history of the universe from about one second after the Big Bang until the present, and have even provided us with hints about the physics governing the evolution of the universe as early as a mere 10^{-34} seconds after the beginning of the universe as we know it.

The model that has emerged from all these experiments is a six-parameter model known as the Λ CDM model. We now know that the geometry of our universe on large scales is well approximated by a flat Friedman-Robertson-Walker metric. With appropriately chosen coordinates (t, \mathbf{x}) , this metric is of the form

$$\bar{g}_{00} = -1, \tag{2.0.1}$$

$$\bar{g}_{0i} = 0, \tag{2.0.2}$$

$$\bar{g}_{ij} = a^2(t)\delta_{ij}, \tag{2.0.3}$$

with corresponding line element

$$ds^2 = -dt^2 + a^2(t)d\mathbf{x}^2. \quad (2.0.4)$$

The time evolution of the scale factor $a(t)$ is governed by the Friedman equation

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2 \left[\Omega_\Lambda + (\Omega_b + \Omega_c) \frac{1}{a^3(t)} + \Omega_r \frac{1}{a^4(t)} \right]. \quad (2.0.5)$$

In the flat case, the absolute scale of $a(t)$ has no physical meaning as it can be changed by a coordinate transformation. In writing the Friedman equation in this form, we have chosen coordinates such that its current value is set to unity, *i.e.* $a(t_0) = 1$, where t_0 denotes the present time. The dot denotes a derivative with respect to the time coordinate t . The Hubble constant H_0 measures the expansion rate of the universe at the present time and is commonly parameterized as $H_0 = h \times 100$ km/s/Mpc; Ω_b parameterizes the current ratio between the energy density stored in baryons and the total energy density of the universe; Ω_c parameterizes the ratio between the energy density stored in dark matter and the total energy density. The current ratio between the energy density stored in radiation, *i.e.* in photons and neutrinos, and the total energy density is denoted by Ω_r and is not counted as a parameter because $\Omega_r h^2$ is a function only of the temperature of the cosmic microwave background radiation at the present time, which is well known experimentally. $\Omega_\Lambda \equiv (1 - \Omega_c - \Omega_b - \Omega_r)$ parameterizes the ratio between the energy density stored in vacuum energy and the total energy density of the universe. The last parameter describing the background geometry is the optical depth of the medium to the surface at which the majority of the photons experience their last interaction with the electrons in the baryonic plasma. It is usually denoted τ . The fluctuations in the Λ CDM model are assumed to

be adiabatic, nearly scale invariant, Gaussian fluctuations.¹ Their spectrum is parameterized by the strength of these fluctuations $\Delta_{\mathcal{R}}^2$ and a scalar spectral index that characterizes the deviation from scale invariance denoted n_s . We will define these quantities more carefully in subsections 2.4.2. Altogether, the six parameters can then for example be taken as $\Omega_b h^2$, $\Omega_c h^2$, H_0 , τ , $\Delta_{\mathcal{R}}^2$, and n_s . This model is consistent with experimental data over a large range of scales, and all parameters in this model are by now known at the per cent level. The values derived from the five year WMAP data alone are given by²

$$\Omega_b h^2 = 0.02273 \pm 0.00062, \quad (2.0.6)$$

$$\Omega_c h^2 = 0.1099 \pm 0.0062, \quad (2.0.7)$$

$$H_0 = 71.9_{-2.7}^{+2.6} \text{ km/s/Mpc}, \quad (2.0.8)$$

$$\tau = 0.087 \pm 0.017, \quad (2.0.9)$$

$$\Delta_{\mathcal{R}}^2 = (2.41 \pm 0.11) \times 10^{-9}, \quad (2.0.10)$$

$$n_s = 0.963_{-0.015}^{+0.014}. \quad (2.0.11)$$

The quoted errors indicate 68% confidence level. The flatness of the universe, which enters the model as an assumption, has also been confirmed experimentally at the per cent level. Furthermore, it should be noted that there is currently no evidence for other deviations from the Λ CDM model such as a momentum dependence of the spectral index, non-Gaussianities, or departures from adiabaticity.

The cosmic microwave background will play a central role throughout this thesis and is what links the different topics together. To keep this work

¹These properties will be introduced as we go along.

²As will be explained later, the quantity $\Delta_{\mathcal{R}}^2$ is momentum dependent. The value given corresponds to a momentum of 0.002 Mpc^{-1} .

more or less self-contained, this chapter thus contains a brief review of the main concepts that are important for an understanding of the cosmic microwave background radiation. The discussion will be somewhat condensed. For the reader looking for a more detailed treatment of the subject, reviews such as [37] and references therein may be useful. The reference I will follow the closest is naturally [38]. The interested reader will find many details and explanations not only regarding the cosmic microwave background but also other topics in cosmology in there.

The interested reader is also encouraged to take a look at the original references. Linear theory of cosmological perturbations was first applied to the CMB anisotropies in [39], [40] for a universe consisting of baryons and photons. This was extended to include dark matter in [41], [42], [43]. The theory for the study of the polarization of the cosmic microwave background was developed in [44], [45], [41], [46], [43], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57]

2.1 Observables – Intensity Matrix and Stokes Parameters

It will be convenient to think of the detector in an idealized fashion as a device that simply counts the number of photons hitting it per unit of time per receiving area, *i.e.* a device that measures the intensity of the radiation it is exposed to. While this is certainly not the most accurate way to think about radiometers, it is a rather good way to think about the bolometer detectors used in ground based experiments such as QUAD and BICEP, balloon borne experiments such as Boomerang, and the Planck-HFI instrument. To extract as much information as possible from the cosmic microwave background radi-

ation, we should not only measure its intensity but also its polarization. To achieve this, imagine equipping the detector with a polarization filter whose angle can be adjusted. One can then measure the intensity as a function of this angle. We denote the measured intensity by

$$\mathcal{I}(\mathbf{x}, \hat{n}, t, \nu, \gamma). \quad (2.1.1)$$

It depends on the position in the universe \mathbf{x} where the experiment takes place, the direction in the sky \hat{n} the detector is pointing at, the time t at which the data is taken, the frequency (band) ν the detector is sensitive to, and finally the orientation of the polarization filter, which we will denote by γ .³

As was expected theoretically and beautifully confirmed experimentally by the COBE satellite, the frequency dependence of the cosmic microwave background radiation is the same as that of a black body with a temperature of $T_0 = 2.725 K$. In an ideal world, one would thus gain nothing by making measurements in different frequency bands. There are, however, foregrounds, *e.g.* from synchrotron emission, thermal dust emission, or free-free emission. These have a different frequency dependence and taking data in multiple frequency bands (typically between tenths and hundreds of GHz) allows these foregrounds to be subtracted rather efficiently. Understanding these foregrounds and how to subtract them is one of the big challenges for future experiments such as CMBPol and deserves detailed treatment, but we will for now simply assume that this can be done and drop the frequency dependence of the measured intensity from our discussion.

The measured intensity contains information both about the ensemble of photons and about the detector. The two components can be disentangled

³To be specific, this angle γ is conventionally measured relative to the Galactic meridian.

by noticing that it takes the form

$$\mathcal{I}(\mathbf{x}, \hat{n}, t, \gamma) = \sum_{i,j} \mathcal{I}_{ij}(\mathbf{x}, \hat{n}, t) e_i(\gamma) e_j^*(\gamma), \quad (2.1.2)$$

where $\mathcal{I}_{ij}(\mathbf{x}, \hat{n}, t)$ is called the intensity matrix. It characterizes the ensemble of photons, while the polarization vectors $\mathbf{e}(\gamma)$ characterize the orientation of the polarization filter on the detector.

As we will use this quantity later, let us introduce a dimensionless intensity matrix J_{ij} as

$$a^4(t) \bar{\rho}_\gamma(t) J_{ij}(\mathbf{x}, -\hat{n}, t) \equiv \mathcal{I}_{ij}(\mathbf{x}, \hat{n}, t) - \bar{\mathcal{I}}_{ij}(t). \quad (2.1.3)$$

Here $a(t)$ denotes the Robertson-Walker scale factor, $\bar{\rho}_\gamma(t)$ is the energy density stored in photons, and $\bar{\mathcal{I}}_{ij}(t)$ is the unpolarized background contribution that is independent of the position of the experiment and the direction in the sky. To understand the dimensionless intensity matrix better, consider the special case such that $\hat{n} = \hat{z}$, a coordinate system is chosen such that the position at which the experiment takes place is $\mathbf{x} = 0$, and the time at which it takes place is t_0 . Since the intensity matrix is of second order in the electromagnetic field, it contains two polarization vectors. Transversality of these polarization vectors⁴ to the direction the photon propagates in together with the Hermitian nature of the matrix as implied by the reality of the measured intensity then guarantees that it must be possible to parameterize the dimensionless intensity matrix as

$$J_{ij}(0, -\hat{z}, t_0) = \frac{2}{T_0} \begin{pmatrix} \Delta T(\hat{z}) + Q(\hat{z}) & U(\hat{z}) - iV(\hat{z}) & 0 \\ U(\hat{z}) + iV(\hat{z}) & \Delta T(\hat{z}) - Q(\hat{z}) & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (2.1.4)$$

⁴For definiteness, we work in Coulomb gauge.

Here $\Delta T(\hat{z})$ is the deviation of the temperature in the \hat{z} -direction from the background temperature $T_0 = 2.725 K$ and except for the background temperature itself is the most traditional observable. A convenient way to extract it from the dimensionless density matrix is to take the trace, which experimentally amounts to removing the polarization filter.⁵ The quantities $Q(\hat{z})$, $U(\hat{z})$, and $V(\hat{z})$ are known as Stokes parameters and are the key quantities in the study of the polarization of the cosmic microwave background. A natural choice for the polarization vector characterizing the orientation of the polarization filter is

$$\mathbf{e}(\gamma) = \hat{x} \cos \gamma + \hat{y} \sin \gamma. \quad (2.1.5)$$

Using equation (2.1.2), the measured intensity and hence the current read out on the instrument as a function of the orientation of the polarization filter is then given by⁶

$$\mathcal{I}(0, \hat{z}, t_0, \gamma) - \bar{\mathcal{I}}(t_0) \propto \Delta T + \sqrt{Q^2 + U^2} \cos \left(2\gamma - \arctan \frac{U}{Q} \right). \quad (2.1.6)$$

Especially for ground based experiments the polarization filter will often be spinning at substantial rate to beat down the noise. The measured current will then be sinusoidal in time with a phase and amplitude set by the Stokes parameters Q and U , which allows them to be extracted rather directly. The amplitude $P \equiv \sqrt{Q^2 + U^2}$ is sometimes called the polarized intensity. The angle that maximizes the polarization signal, *i.e.* $2\gamma = \arctan \frac{U}{Q}$, is often called the polarization direction. The Stokes parameter V disappeared from

⁵The factor four relating the trace of the intensity matrix to the temperature fluctuation of course has to do with the fact that the intensity goes like the fourth power of the temperature.

⁶In practice $\bar{\mathcal{I}}(t_0)$ is much larger than both temperature fluctuations and Stokes parameters. Instead of an absolute measurement in a single direction it is thus common to perform differential measurements for two different points in the sky such that $\bar{\mathcal{I}}(t_0)$ drops out.

the formula, but it is not important. The polarization of the cosmic microwave background is generated by the scattering of light by non-relativistic electrons which does not lead to circularly polarized light. One thus expects \mathcal{I}_{ij} to be real and $V = 0$.

Having introduced the temperature fluctuations and Stokes parameters for one point in the sky \hat{z} , let us generalize the definition for arbitrary directions \hat{n} . We can conveniently define the temperature anisotropy via the trace of the dimensionless intensity matrix as

$$\frac{\Delta T(\hat{n})}{T_0} = \frac{1}{4} J_{ii}(0, -\hat{n}, t_0). \quad (2.1.7)$$

It has become traditional to represent the experimental results for the temperature anisotropy in the form of color-coded sky maps. As example the five-year WMAP temperature sky maps [58] are shown in Figure 2.1.

To define the Stokes parameters for an arbitrary direction \hat{n} , it is convenient to notice that one has

$$Q(\hat{z}) \pm iU(\hat{z}) = \frac{T_0}{2} e_{\pm i}(\hat{z}) e_{\pm j}(\hat{z}) J_{ij}(0, -\hat{z}, t_0), \quad (2.1.8)$$

where

$$\mathbf{e}_{\pm}(\hat{z}) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm i \\ 0 \end{pmatrix}, \quad (2.1.9)$$

are the polarization vectors for a photon coming to us from the \hat{z} -direction. In analogy, we can then define the Stokes parameters for an arbitrary direction \hat{n} as

$$Q(\hat{n}) \pm iU(\hat{n}) = \frac{T_0}{2} e_{\pm i}(\hat{n}) e_{\pm j}(\hat{n}) J_{ij}(0, -\hat{n}, t_0), \quad (2.1.10)$$

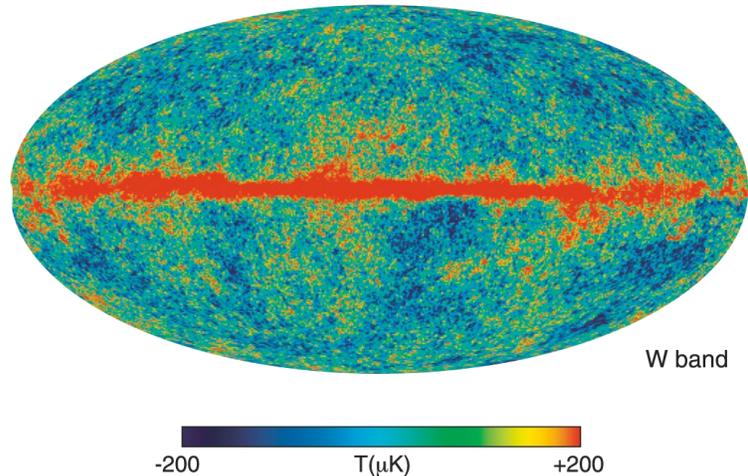


Figure 2.1: This figure shows the WMAP five-year temperature sky map (Mollweide projection in Galactic coordinates) in the W band (94 GHz) smoothed with a 0.2° Gaussian beam.

where we define the polarization vectors for a photon coming to us from the \hat{n} direction as

$$\mathbf{e}_\pm(\hat{n}) = \frac{1}{\sqrt{2}}(\hat{\theta} \pm i\hat{\phi}). \quad (2.1.11)$$

To be explicit, for coordinates on the sphere such that

$$\hat{n} = \begin{pmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{pmatrix}, \quad (2.1.12)$$

the unit vectors $\hat{\theta}$ and $\hat{\phi}$ are given by

$$\hat{\theta} = \begin{pmatrix} \cos \theta \cos \phi \\ \cos \theta \sin \phi \\ -\sin \theta \end{pmatrix} \quad \text{and} \quad \hat{\phi} = \begin{pmatrix} -\sin \phi \\ \cos \phi \\ 0 \end{pmatrix}. \quad (2.1.13)$$

Like the temperature anisotropy data, the polarization data is also commonly represented in the form of color-coded sky maps. The different colors represent different values for the polarized intensity and the direction of maximal polarization is indicated by a white bar. As an example we show the five-year WMAP results on polarization [58] in Figure 2.2.

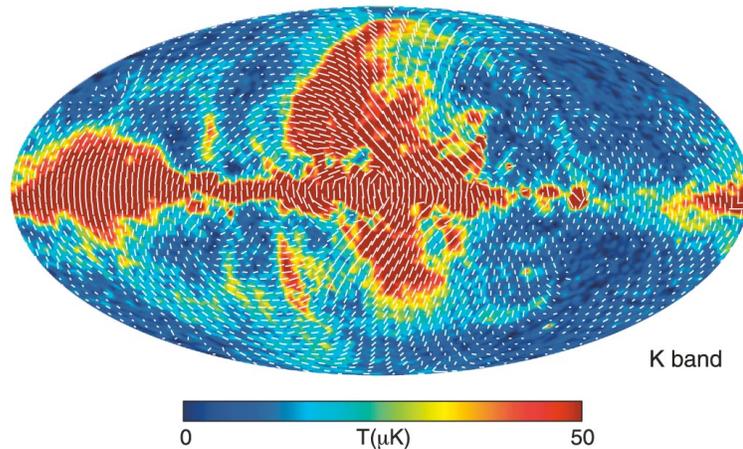


Figure 2.2: This figure shows the WMAP five-year polarization sky map (Mollweide projection in Galactic coordinates) in the K band (23 GHz) smoothed to an effective Gaussian beam of 2.0° . The color scale indicates polarized intensity, $P = \sqrt{Q^2 + U^2}$, and the line segments indicate polarization direction in pixels whose signal-to-noise ratio exceeds 1.

One should have in mind that this was a rather simplified discussion and ignored many of the challenges that arise in experiments such as foreground subtraction, understanding beam profiles, pointing abilities, *etc.*, but it will certainly be good enough for our present purposes.

2.2 Comparison with Theory – The Multipole Coefficients

As explained in the last section, experimentally we have rather direct access to the Stokes parameters, or equivalently the intensity matrix. These quantities do depend, however, on the position in the universe the experiment takes place. Equivalently, they depend on initial conditions which can not be predicted theoretically. All we can predict theoretically are the probability distributions these initial conditions obey. To an extent that we will make more precise in what follows, the information about the early universe is then encoded in the correlations between the temperature fluctuations and Stokes parameters at various points in the sky. The obvious first step is to look for correlations in these quantities at two different points in the sky, *i.e.* to study two-point functions. It has become common practice to represent the information contained in these two-point functions in the form of so-called multipole coefficients which we shall introduce in detail below. Even though there is already a large literature on the 3-point functions of the temperature anisotropies, the study of these as well as higher n -point functions is still in its infancy, to a large extent because the data only allows us to bound departures from Gaussianity but not to detect them. As the quality of the data improves, the way the information stored in 3- and higher n -point functions will certainly have to be refined.

The most commonly studied two-point functions are

$$\langle \Delta T(\hat{n}) \Delta T(\hat{n}') \rangle \quad , \quad (2.2.1)$$

$$\langle \Delta T(\hat{n}) [Q(\hat{n}') + iU(\hat{n}')] \rangle \quad , \quad (2.2.2)$$

$$\langle [Q(\hat{n}) + iU(\hat{n})] [Q(\hat{n}') + iU(\hat{n}')] \rangle \quad , \quad (2.2.3)$$

$$\langle [Q(\hat{n}) + iU(\hat{n})] [Q(\hat{n}') - iU(\hat{n}')] \rangle \quad . \quad (2.2.4)$$

In these formulas, we think of the temperature fluctuations and Stokes parameters not as the experimentally measured quantities, but rather as the theoretical predictions for these for a given history, and the averages denote an average over all possible histories in a given model. It has become customary to represent the information contained in these two-point functions in the form of multipole coefficients which we will briefly review now. The most widely used ones are those corresponding to the temperature anisotropies. So we will begin with these and define the others in analogy.

The temperature fluctuation $\Delta T(\hat{n})$ is a section of the normal bundle of the two-sphere in \mathbb{R}^3 , or, since this bundle is trivial, simply a function. We know that functions on a sphere can be expanded in terms of the eigenfunctions of the Laplacian on the sphere, the spherical harmonics. We will write this expansion as

$$\Delta T(\hat{n}) = \sum_{\ell, m} a_{T, \ell m} Y_{\ell}^m(\hat{n}). \quad (2.2.5)$$

Here $Y_{\ell}^m(\hat{n})$ are the spherical harmonics and $a_{T, \ell m}$ are the expansion coefficients and the subscript T indicates that we are dealing with the temperature fluctuations. Using the orthonormality of the spherical harmonics, the expansion coefficients can be extracted by integration

$$a_{T, \ell m} = \int d^2\hat{n} Y_{\ell}^{m*}(\hat{n}) \Delta T(\hat{n}). \quad (2.2.6)$$

The spherical harmonics most widely used in the study of the cosmic microwave background behave under complex conjugation as⁷

$$Y_{\ell}^m(\hat{n})^* = Y_{\ell}^{-m}(\hat{n}). \quad (2.2.7)$$

⁷These conventions are different from the ones used in most quantum mechanics books. So we briefly summarize them in Appendix B.

The reality of the temperature fluctuation, $\Delta T(\hat{n})$, then guarantees that the expansion coefficients satisfy

$$a_{T,\ell m}^* = a_{T,\ell -m}. \quad (2.2.8)$$

These expansion coefficients should be thought of as random variables that obey a (possibly only very nearly) Gaussian probability distribution. It is usually also assumed to respect rotational invariance. In this case, the two-point functions of these coefficients have to be of the form

$$\langle a_{T,\ell m} a_{T,\ell' m'}^* \rangle = C_{TT,\ell} \delta_{\ell\ell'} \delta_{mm'}. \quad (2.2.9)$$

Using $\sum_m Y_\ell^m(\hat{n}) Y_\ell^{m*}(\hat{n}') = (2\ell + 1)/4\pi P_\ell(\hat{n} \cdot \hat{n}')$ with $P_\ell(x)$ the Legendre polynomials, the correlation function (2.2.1) becomes

$$\langle \Delta T(\hat{n}) \Delta T(\hat{n}') \rangle = \sum_\ell C_{TT,\ell} P_\ell(\hat{n} \cdot \hat{n}'), \quad (2.2.10)$$

and we see that all the information about the two-point correlations of the temperature fluctuations is contained in the coefficients $C_{TT,\ell}$.

Different from the temperature fluctuation $\Delta T(\hat{n})$, the Stokes parameters are coefficients of a section in the symmetric product of two copies of the tangent bundle of the two-sphere, or components of a symmetric tensor. As a consequence, we should expand them in terms of eigensections of the corresponding Laplacian. The expansion analogous to equation (2.2.17) then takes the form

$$Q(\hat{n}) + iU(\hat{n}) = \sum_{\ell,m} a_{P,\ell m} {}_2Y_\ell^m(\hat{n}), \quad (2.2.11)$$

where ${}_2Y_\ell^m(\hat{n})$ are the spin-2 spherical harmonics. Their basic properties are reviewed in Appendix B. Using their orthonormality property, the expansion

coefficients can again be extracted by integration

$$a_{P,\ell m} = \int d^2\hat{n} {}_2Y_\ell^{m*}(\hat{n}) (Q(\hat{n}) + iU(\hat{n})) . \quad (2.2.12)$$

Since we are expanding a complex quantity, the new expansion coefficients do not satisfy a simple reality condition like the coefficients $a_{T,\ell m}$. It is convenient to introduce two sets of coefficients that do

$$a_{E,\ell m} \equiv -(a_{P,\ell m} + a_{P,\ell -m}^*)/2 \quad \text{and} \quad a_{B,\ell m} \equiv i(a_{P,\ell m} - a_{P,\ell -m}^*)/2 . \quad (2.2.13)$$

The signs are chosen to match the conventions in [57]. The crucial difference between the new expansion coefficients $a_{E,\ell m}$ and $a_{B,\ell m}$ is their behavior under space inversion. While $a_{T,\ell m}$ and $a_{E,\ell m}$ under a space inversion go to $(-1)^\ell a_{T,\ell m}$ and $(-1)^\ell a_{E,\ell m}$, respectively, $a_{B,\ell m}$ goes to $-(-1)^\ell a_{B,\ell m}$. The subscripts E and B were chosen in analogy to static electric and magnetic fields, as an E -mode, *i.e.* only $a_{E,\ell m}$ non-zero, corresponds to a curl-free polarization field while the B -mode, *i.e.* only $a_{B,\ell m}$ non-zero, corresponds to a divergence-free polarization field.

$$\langle a_{T,\ell m} a_{E,\ell' m'}^* \rangle = C_{TE,\ell} \delta_{\ell\ell'} \delta_{mm'} , \quad (2.2.14)$$

$$\langle a_{E,\ell m} a_{E,\ell' m'}^* \rangle = C_{EE,\ell} \delta_{\ell\ell'} \delta_{mm'} , \quad (2.2.15)$$

$$\langle a_{B,\ell m} a_{B,\ell' m'}^* \rangle = C_{BB,\ell} \delta_{\ell\ell'} \delta_{mm'} . \quad (2.2.16)$$

If the distribution governing these coefficients respects parity, these are the only non-vanishing coefficients and they contain all the information encoded in the correlations of temperature fluctuations and Stokes parameters. In a theory that violates parity, there will typically also be the analogously defined $C_{TB,\ell}$ and $C_{EB,\ell}$ multipole coefficients, but there is currently no indication that parity is violated, and we will ignore this possibility in this work. It

follows from the reality conditions of the coefficients $a_{T,\ell m}$, $a_{E,\ell m}$, and $a_{B,\ell m}$ that the multipole coefficients $C_{TE,\ell}$ are real while the multipole coefficients $C_{TT,\ell}$, $C_{EE,\ell}$, and $C_{BB,\ell}$ are real and positive. Our conventions match those of [57] and the multipole coefficients have units of square temperature.

One can write down expressions analogous to equation (2.2.10) to express the two-point functions (2.2.2), (2.2.3), (2.2.4) in terms of the multipole coefficients $C_{TE,\ell}$, $C_{TE,\ell}$, and $C_{BB,\ell}$, but these expressions are hardly used. So we will not review them here.

If the distribution governing the expansion coefficients $a_{T,\ell m}$, $a_{E,\ell m}$, and $a_{B,\ell m}$ is Gaussian, the multipole coefficients $C_{TT,\ell}$, $C_{TE,\ell}$, $C_{EE,\ell}$, and $C_{BB,\ell}$ contain everything there is to know about the cosmic microwave background as the odd n -point functions vanish and the higher even n -point functions are given in terms of sums of products of the two-point function. It is natural to expect, however, that this distribution is in fact not exactly Gaussian and that there are small non-Gaussianities. In this case, additional information would be contained in the higher n -point functions. The simplest observable to test for non-Gaussianities is, of course, the three-point function, but the sensitivity of current experiments is not good enough for a detection [59]. Planck data will possibly improve the situation.

So far the discussion has been rather theoretical involving stochastic parameters and averages over all possible histories of bilinears of these parameters. In practice, we can neither average over all possible histories nor, what would be equivalent by the ergodic theorem, over all positions in the universe, and we have to content ourselves with a measurement of the cosmic microwave background in the one universe we live in from within our solar system. A natural question then is how these multipole coefficients can be

extracted from such a measurement. We shall briefly describe how this is possible for an ideal scenario in which one has a noiseless full-sky map and refer the reader to the literature for the more realistic case where noise has to be taken into account and foregrounds limit us to parts of the sky. We will discuss the temperature fluctuations first. The Stokes parameters can then again be treated analogously.

Assuming the rather idealized situation that an experiment provides us with a noiseless full-sky map, we can expand the temperature fluctuation as

$$\Delta T(\hat{n}) = \sum_{\ell m} a_{T,\ell m}^{\text{obs}} Y_{\ell}^m(\hat{n}). \quad (2.2.17)$$

The expansion coefficients $a_{\ell m}^{\text{obs}}$, obtained by integration from the temperature map, are now simply complex numbers satisfying the same reality properties as their theoretical counterparts corresponding to one particular history. We can thus not take an average as we did before, but we can define what is called an estimator for the temperature multipole coefficients as

$$C_{TT,\ell}^{\text{obs}} \equiv \frac{1}{2\ell + 1} \sum_m |a_{T,\ell m}^{\text{obs}}|^2 \quad (2.2.18)$$

This provides an unbiased estimate of the multipole coefficients $C_{TT,\ell}$ in the sense that the average of the analogously defined quantity for the stochastic parameters $a_{T,\ell m}$ satisfies

$$\langle \tilde{C}_{TT,\ell}^{\text{obs}} \rangle \equiv \frac{1}{2\ell + 1} \left\langle \sum_m |a_{T,\ell m}|^2 \right\rangle = C_{TT,\ell}. \quad (2.2.19)$$

However, since there are only $2\ell + 1$ modes per ℓ -value, even for an ideal noiseless full-sky map this estimate of the multipole coefficients comes with an uncertainty

$$\left\langle \left(\frac{\tilde{C}_{TT,\ell}^{\text{obs}} - C_{TT,\ell}}{C_{TT,\ell}} \right)^2 \right\rangle = \frac{2}{2\ell + 1}. \quad (2.2.20)$$

This is known as cosmic variance.

In a more realistic setting, noise and other sources of errors cause this estimator to be biased. Furthermore, the data has to be masked near the Galactic plane as foregrounds there cannot be subtracted reliably. The expansion (2.2.17) then has to be modified to account for this, and more sophisticated estimators have to be used. The current state of the art is to use a pixel-based maximum likelihood estimator for low ℓ and a pseudo- C_ℓ estimator for higher ℓ . For details we refer the reader to [60] and references therein.

For the idealized scenario, similar estimators to the one in (2.2.18) for $C_{TT,\ell}$ can be defined for the multipole coefficients $C_{TE,\ell}$, $C_{EE,\ell}$, and $C_{BB,\ell}$ from the coefficients $a_{E,\ell m}^{\text{obs}}$ and $a_{B,\ell m}^{\text{obs}}$ extracted from the observed Stokes parameters. Just like for the temperature analysis, the state of the art is a hybrid estimator using a pixel-based maximum likelihood estimator for low multipole and a pseudo- C_ℓ estimator for higher multipoles. In addition to the subtleties due to noise and sky cut already present in the analysis of temperature fluctuations, in the case of polarization the sky cut introduces the additional complication that the separation into E - and B -mode is no longer unique. For the details on the polarization analysis, we refer the reader to [61] and references therein.

2.3 Derivation of the Multipole Coefficients

As reviewed in Section 2.1, the observables that are most directly accessible to cosmic microwave background experiments are the temperature anisotropies and the Stokes parameters Q and U . As explained in the previous section, these cannot directly be used to compare with theory because they depend on initial conditions that cannot be predicted. What can be pre-

dicted theoretically are the probability distributions that govern these initial conditions. If these probability distributions are Gaussian, they can be completely specified through the multipole coefficients $C_{TT,\ell}$, $C_{TE,\ell}$, $C_{EE,\ell}$, and $C_{BB,\ell}$. These can be extracted fairly efficiently from the sky maps of temperature fluctuations and Stokes parameters and thus provide the main point of contact between theory and experiment. To complete the picture, we will now review how these multipole coefficients can be calculated for a given theory.

The key quantity in this will be the dimensionless intensity matrix introduced in Section 2.1. We will calculate it for arbitrary initial conditions and then evaluate the average over initial conditions of bilinears of the components of the intensity matrix to obtain the multipole coefficients. The dimensionless intensity matrix is closely related to the number density of photons in phase space. Its time evolution in a given background is then governed by the Boltzmann equation for the ensemble of photons. This equation involves metric perturbations whose dynamics is of course governed by the linearized Einstein equations. We will thus briefly review linear perturbation theory around a flat FRW background in subsection 2.3.1 before deriving the Boltzmann equations for photons in subsection 2.3.2 and those for neutrinos in 2.3.3. We will put everything together and review the expressions for the multipole coefficients and the equations needed to calculate them in the Λ CDM model in subsection 2.3.4. For definiteness, we will work in Coulomb gauge as far as the photons are concerned and in a particular synchronous gauge that we will define below as far as the metric perturbations are concerned.

2.3.1 Linear perturbation theory in a flat FRW universe

In this subsection, we will summarize the linearized Einstein equations as well as the equations that follow from the covariant conservation of the stress-energy tensor at the linear level in a flat FRW universe. The background metric in such a universe is given by equations (2.0.1)-(2.0.3), and the contributions of the various constituents to the background stress-energy tensor are all of the perfect fluid form $\bar{T}_{\mu\nu} = \bar{p}\bar{g}_{\mu\nu} + (\bar{p} + \bar{\rho})\bar{u}_\mu\bar{u}_\nu$ with the velocity vector field \bar{u}_μ satisfying $\bar{g}^{\mu\nu}\bar{u}_\mu\bar{u}_\nu = -1$.⁸ In the rest frame of the fluid, $\bar{u}_i = 0$ and $\bar{u}_0 = -1$, leading to a stress-energy tensor of the form

$$\bar{T}_{00} = \bar{\rho}, \quad (2.3.1)$$

$$\bar{T}_{0i} = 0, \quad (2.3.2)$$

$$\bar{T}_{ij} = a^2(t)\bar{p}\delta_{ij}. \quad (2.3.3)$$

The Einstein equations for this background take the form⁹

$$\frac{3\dot{a}^2}{a^2} = 8\pi G\bar{\rho}, \quad (2.3.4)$$

and

$$\frac{\dot{a}^2}{a^2} + \frac{2\ddot{a}}{a} = -8\pi G\bar{p}, \quad (2.3.5)$$

where $\bar{\rho}$ and \bar{p} stand for the total energy density and total pressure from all constituents, respectively. As already briefly summarized at the beginning of this chapter, for the Λ CDM model these are a cosmological constant, cold dark matter, baryonic matter, as well as radiation. For the cosmological constant, the pressure is related to the energy density as $\bar{p}_\Lambda = -\bar{\rho}_\Lambda$. For the time periods

⁸It is conventional to use bars to indicate that we are dealing with background quantities.

⁹The first of these equations is often called the Friedmann equation.

we are interested in, chiefly temperatures of $10^9 K$ and lower, the baryonic plasma consists of non-relativistic electrons and protons whose pressure is negligible compared to their energy density, so that $\bar{p}_b = 0$. The cold dark matter particles are assumed to be non-relativistic as well, so that $\bar{p}_c = 0$. For the radiation component, *i.e.* the particles that remain relativistic at these low temperatures, photons and neutrinos, the pressure is related to the energy density as $\bar{p}_r = \bar{\rho}_r/3$. The Einstein equations together with the equations that follow from the covariant conservation of the stress energy tensor for the various components can then be brought into the form of equation (2.0.5).

To incorporate fluctuations around this background, one considers a metric of the form

$$g_{\mu\nu}(t, \mathbf{x}) = \bar{g}_{\mu\nu}(t) + h_{\mu\nu}(t, \mathbf{x}). \quad (2.3.6)$$

The fluctuations in the inverse metric $h^{\mu\nu}(t, \mathbf{x}) \equiv g^{\mu\nu}(t, \mathbf{x}) - \bar{g}^{\mu\nu}(t)$ are given in terms of these as $h^{\mu\nu} = -\bar{g}^{\mu\rho}\bar{g}^{\nu\sigma}h_{\rho\sigma}$.

To obtain the perturbed stress-energy tensor, it is convenient to start from that of a perfect fluid, and then include perturbations that parameterize the deviation from the perfect fluid case. The stress-energy tensor for a perfect fluid is of the form $T_{\mu\nu} = p g_{\mu\nu} + (p + \rho)u_\mu u_\nu$, where the velocity vector field u_μ satisfies $g^{\mu\nu}u_\mu u_\nu = -1$. If we decompose the velocity vector field as $u_\mu = \bar{u}_\mu + \delta u_\mu$, this implies $\delta u^0 = \delta u_0 = h_{00}/2$. The first order perturbations to the stress energy tensor of a perfect fluid, $\delta T_{\mu\nu} \equiv T_{\mu\nu} - \bar{T}_{\mu\nu}$, can then be parameterized as

$$\delta T_{00} = -\bar{\rho}h_{00} + \delta\rho, \quad (2.3.7)$$

$$\delta T_{0i} = \bar{p}h_{0i} - (\bar{\rho} + \bar{p})\delta u_i, \quad (2.3.8)$$

$$\delta T_{ij} = \bar{p}h_{ij} + a^2\delta_{ij}\delta p. \quad (2.3.9)$$

In the study of linear perturbations around a flat FRW background, it is convenient to make use of the symmetries of the background. The spatial slices are simply \mathbb{R}^3 with the flat metric and are thus invariant under translations as well as rotations. One can classify the perturbations around this background according to their transformation properties with respect to rotations into scalar modes, vector modes, and tensor modes. This decomposition is useful because the symmetry ensures that these modes cannot mix at the linear level.

As far as the metric fluctuations are concerned, this leads to a parameterization of the fluctuations of the form¹⁰

$$h_{00}(t, \mathbf{x}) = -E(t, \mathbf{x}), \quad (2.3.10)$$

$$h_{0i}(t, \mathbf{x}) = a [\partial_i F(t, \mathbf{x}) + G_i(t, \mathbf{x})], \quad (2.3.11)$$

$$h_{ij}(t, \mathbf{x}) = a^2 [A(t, \mathbf{x})\delta_{ij} + \partial_i \partial_j B(t, \mathbf{x}) + \partial_i C_j(t, \mathbf{x}) + \partial_j C_i(t, \mathbf{x}) + D_{ij}(t, \mathbf{x})], \quad (2.3.12)$$

where E , F , A , and B are scalar modes, C_i and G_i are vector modes, and D_{ij} are tensor modes or the graviton fluctuations. The vector fluctuations are divergence-free, *i.e.* $\partial_i C_i = \partial_i G_i = 0$, and the tensor fluctuations are divergence-free and traceless, *i.e.* $\partial_i D_{ij} = \partial_i D_{ji} = 0$ and $D_{ii} = 0$, with all repeated indices being summed over.

For the fluctuations of the stress energy tensor, the decomposition leads

¹⁰We will later work in synchronous gauge. These perturbations are related to the ones commonly used in calculations in synchronous gauge by $h = 3A + \nabla^2 B$ and $\eta = -A/2$. The quantity $\alpha = -\dot{B}/2$ is also sometimes used.

to the following parameterization

$$\delta T_{00} = -\bar{\rho} h_{00} + \delta\rho, \quad (2.3.13)$$

$$\delta T_{0i} = \bar{p} h_{0i} - (\bar{\rho} + \bar{p}) (\partial_i \delta u + \delta u_i^V), \quad (2.3.14)$$

$$\delta T_{ij} = \bar{p} h_{ij} + a^2 (\delta_{ij} \delta p + \partial_i \partial_j \pi^S + \partial_i \pi_j^V + \partial_j \pi_i^V + \pi_{ij}^T), \quad (2.3.15)$$

where the metric fluctuations are those of equations (2.3.10)-(2.3.12), and we have decomposed δu_i into a scalar velocity potential, δu , and a divergenceless vector δu_i^V . Furthermore, we have introduced the anisotropic inertia π^S , the divergenceless π_i^V vector anisotropic inertia, and the divergenceless, traceless tensor anisotropic inertia, π_{ij}^T . It is sometimes more convenient to work with the perturbations of the stress energy tensor with one upper and one lower index. These are given by

$$\delta T^0_0 = -\delta\rho, \quad (2.3.16)$$

$$\delta T^0_i = (\bar{\rho} + \bar{p}) (\partial_i \delta u + \delta u_i^V), \quad (2.3.17)$$

$$\delta T^i_j = \delta_{ij} \delta p + \partial_i \partial_j \pi^S + \partial_i \pi_j^V + \partial_j \pi_i^V + \pi_{ij}^T. \quad (2.3.18)$$

For this ansatz the Einstein tensor in a general gauge is of the form

$$\delta G_{00} = -\frac{\nabla^2}{a^2} A + 3\frac{\dot{a}}{a} \dot{A} + \frac{\dot{a}}{a} \nabla^2 \dot{B} - \frac{2\dot{a}}{a^2} \nabla^2 F, \quad (2.3.19)$$

$$\begin{aligned} \delta G_{0i} &= \partial_i \left(\frac{\dot{a}}{a} E - \dot{A} - \left(\frac{\dot{a}^2}{a^2} + 2\frac{\ddot{a}}{a} \right) aF \right) \\ &\quad - \left(\frac{\dot{a}^2}{a^2} + \frac{2\ddot{a}}{a} \right) aG_i - \frac{1}{2} \frac{\nabla^2}{a^2} aG_i + \frac{1}{2} \nabla^2 \dot{C}_i, \end{aligned} \quad (2.3.20)$$

$$\begin{aligned} \delta G_{ij} &= \delta_{ij} \left(\dot{a}^2 E + 2a\ddot{a}E + \frac{1}{2} \nabla^2 E + a\dot{a}\dot{E} \right. \\ &\quad \left. - (\dot{a}^2 + 2a\ddot{a})A + \frac{1}{2} \nabla^2 A - a^2 \ddot{A} - 3a\dot{a}\dot{A} \right. \\ &\quad \left. - \frac{1}{2} a^2 \nabla^2 \dot{B} - \frac{3}{2} a\dot{a} \nabla^2 \dot{B} + 2\dot{a} \nabla^2 F + a \nabla^2 \dot{F} \right) \end{aligned} \quad (2.3.21)$$

$$\begin{aligned}
& +\partial_i\partial_j \left(-\frac{1}{2}E - \frac{1}{2}A - (\dot{a}^2 + 2a\ddot{a})B \right. \\
& \quad \left. + \frac{1}{2}a^2\ddot{B} + \frac{3}{2}a\dot{a}\dot{B} - 2\dot{a}F - a\dot{F} \right) \\
& -\dot{a}(\partial_i G_j + \partial_j G_i) - \frac{1}{2}a(\partial_i \dot{G}_j + \partial_j \dot{G}_i) \\
& -(\dot{a}^2 + 2a\ddot{a})(\partial_i C_j + \partial_j C_i) + \frac{1}{2}a^2(\partial_i \ddot{C}_j + \partial_j \ddot{C}_i) \\
& + \frac{3}{2}a\dot{a}(\partial_i \dot{C}_j + \partial_j \dot{C}_i) \\
& + \frac{1}{2}a^2\ddot{D}_{ij} + \frac{3}{2}a\dot{a}\dot{D}_{ij} - \frac{1}{2}\nabla^2 D_{ij} - (\dot{a}^2 + 2a\ddot{a})D_{ij}.
\end{aligned}$$

With these expressions at hand, we can write down the linearized Einstein equations in a flat FRW background. They give rise to four equations governing the evolution of the scalar modes. The first arises from the 00-Einstein equations and in synchronous gauge, *i.e.* a gauge in which $h_{00} = h_{0i} = 0$, takes the form

$$-\frac{\nabla^2}{a^2}A + 3\frac{\dot{a}}{a}\dot{A} + \frac{\dot{a}}{a}\nabla^2\dot{B} = 8\pi G\delta\rho. \quad (2.3.22)$$

The second equation arises from the part of the $0i$ -Einstein equation proportional to ∂_i and in synchronous becomes

$$\dot{A} = 8\pi G(\bar{\rho} + \bar{p})\delta u. \quad (2.3.23)$$

The third and fourth equations arise from the parts of the ij -Einstein equation proportional to δ_{ij} and $\partial_i\partial_j$, respectively. They take the form

$$\frac{1}{2}\frac{\nabla^2}{a^2}A - \ddot{A} - 3\frac{\dot{a}}{a}\dot{A} - \frac{1}{2}\nabla^2\ddot{B} - \frac{3}{2}\frac{\dot{a}}{a}\nabla^2\dot{B} = 8\pi G\delta p, \quad (2.3.24)$$

and

$$-A + a^2\ddot{B} + 3a\dot{a}\dot{B} = 16\pi G a^2 \pi^S. \quad (2.3.25)$$

In writing these equations we have used the ij -Einstein equation for the flat FRW background (2.3.5)

The Einstein equations give rise to two equations governing the time evolution of the vector modes. The first arises from the divergenceless part of the $0i$ -Einstein equation. In synchronous gauge it is of the form

$$\nabla^2 \dot{C}_i = -16\pi G(\bar{\rho} + \bar{p})\delta u_i^V. \quad (2.3.26)$$

The second one arises from the ij -Einstein equation and, using equations of motion for the background, in synchronous gauge can be written as

$$\ddot{C}_i + 3\frac{\dot{a}}{a}\dot{C}_i = 16\pi G\pi_i^V. \quad (2.3.27)$$

Finally, the equation governing the time evolution of the tensor modes arises from the divergenceless, traceless part of the ij -Einstein equations. It takes the form

$$\ddot{D}_{ij} + 3\frac{\dot{a}}{a}\dot{D}_{ij} - \frac{\nabla^2}{a^2}D_{ij} = 16\pi G\pi_{ij}^T. \quad (2.3.28)$$

In addition to the Einstein equation, the equations obtained from linearizing the covariant conservation of the stress-energy tensor, $\nabla_\mu T^\mu_\nu$ will also be useful. These give rise to two equations for the scalar modes. Energy conservation gives

$$\begin{aligned} \dot{\delta\rho} + \frac{3\dot{a}}{a}(\delta\rho + \delta p) + \frac{\nabla^2}{a^2}[(\bar{\rho} + \bar{p})\delta u + a\dot{a}\pi^S] \\ + \frac{1}{2}(\bar{\rho} + \bar{p})(3\dot{A} + \nabla^2\dot{B}) = 0. \end{aligned} \quad (2.3.29)$$

The part of the equation corresponding to momentum conservation proportional to ∂_i gives the equation

$$\delta p + \nabla^2\pi^S + \frac{1}{a^3}\frac{\partial}{\partial t}[a^3(\bar{\rho} + \bar{p})\delta u] = 0. \quad (2.3.30)$$

Finally, covariant conservation of the stress-energy tensor leads to an equation for the vector modes of the form

$$\nabla^2 \pi_i^V + \frac{1}{a^3} \frac{\partial}{\partial t} [a^3 (\bar{\rho} + \bar{p}) \delta u_i^V] = 0. \quad (2.3.31)$$

If there are several non-interacting fluids, these equations have to be obeyed individually for each of the fluids. This is, of course, stronger than the condition that the total stress energy tensor be covariantly conserved, which is what follows from the Einstein equations.

From equations (2.3.27) and (2.3.31), we see that in the absence of anisotropic stress, the vector modes decay rapidly with time. So as long as they are not continuously being generated, these will be negligible and we will drop them from our analysis in the remainder of this work. One should, however, have in mind that these are important in the study of models with textures or strings.

We thus see that the time evolution of the fluctuations is governed by systems of coupled partial differential equations. To simplify the solution of this system of equations, it is convenient to make use of the translational invariance of the system and look for a solution in Fourier space, *i.e.* we look for a solution in the form of a superposition of plane waves $e^{i\mathbf{q}\cdot\mathbf{x}}$. The time evolution of the coefficients of $e^{i\mathbf{q}\cdot\mathbf{x}}$ is then governed by a system of coupled linear ordinary differential equations. The general solution of such a system of differential equations can be written as a superposition of several linearly independent solutions, which can be labeled by a discrete label n . Rotational invariance of the background ensures that the equations can only depend on the magnitude $q \equiv |\mathbf{q}|$ but not its direction, and the basis set of linearly independent solutions can then also be chosen to depend only on the magnitude q and not

the direction. The only directional dependence then arises through the initial conditions, *i.e.* the coefficients of these solutions.

For concreteness, let us first consider scalar perturbations and then turn to the tensor perturbations. If we collectively denote the scalar modes by a vector $\mathbf{X}(t, \mathbf{x})$ whose components are the scalar perturbations $\delta\rho$, δp , *etc.*, then the general solution can be written in the form

$$\mathbf{X}(t, \mathbf{x}) = \sum_n \int d^3q \alpha_n(\mathbf{q}) \mathbf{X}_{nq}(t) e^{i\mathbf{q}\cdot\mathbf{x}}. \quad (2.3.32)$$

The coefficients $\alpha_n(\mathbf{q})$ should be thought of as stochastic parameters governed by a Gaussian probability distribution. They of course underly the expansion coefficients $a_{T,\ell m}$, $a_{E,\ell m}$, and $a_{B,\ell m}$ that appeared in Section 2.3.4. Again, the information about the probability distribution they obey is entirely contained in the two-point function of these. Assuming that the probability distribution respects translational and rotational invariance, this two-point function must be of the form

$$\langle \alpha_n(\mathbf{q}) \alpha_m^*(\mathbf{q}') \rangle = P_{nm}(q) \delta(\mathbf{q} - \mathbf{q}'). \quad (2.3.33)$$

It follows from this expression that the matrix $P_{nm}(q)$ must be hermitian. Only a combination of this matrix and the normalization of the basis set of linearly independent solutions is, of course, meaningful, and one could choose a basis such that

$$\langle \tilde{\alpha}_n(\mathbf{q}) \tilde{\alpha}_m^*(\mathbf{q}') \rangle = \delta_{nm} \delta(\mathbf{q} - \mathbf{q}'), \quad (2.3.34)$$

One of the assumptions of the Λ CDM model is that only the adiabatic mode is excited. This mode is characterized by the property that the quantity $\mathcal{R}_q \equiv A_q/2 + H\delta u_q$ approaches a constant, \mathcal{R}_q^o , in the limit $q/a \ll H$. It

furthermore has the property that the perturbation for any scalar perturbation $s(t, \mathbf{x})$ like the energy density of the various fluids, $\rho(t, \mathbf{x})$, their pressures, $p(t, \mathbf{x})$, *etc.* are given by

$$\delta s_q(t) = -\frac{\mathcal{R}_q^o \dot{\bar{s}}(t)}{a(t)} \int^t a(t') dt', \quad (2.3.35)$$

and the perturbation to the velocity potential is given by

$$\delta u_q(t) = \frac{\mathcal{R}_q^o}{a(t)} \int^t a(t') dt'. \quad (2.3.36)$$

There is also a second adiabatic mode with $\mathcal{R}_q^o = 0$, but it decays with time and can be ignored in models in which the perturbations are generated early.

This constancy in \mathcal{R}_q has been used in special cases for a long time [62], [63] but proofs that such a mode must exist under rather general circumstances have only been given more recently for the linear case [64], [65], [66] as well as the non-linear case [67], [68], [69], [70].

As we will see in 2.4, the perturbations generated in single field inflation are necessarily adiabatic. Deviations from adiabaticity are of course looked for as they would be an indication for physics beyond the Λ CDM model, but at this point there is no real indication that other modes, often called entropic modes or more commonly isocurvature perturbations, are excited. In the adiabatic case, the solution can then be chosen such that

$$\langle \alpha(\mathbf{q}) \alpha^*(\mathbf{q}') \rangle = \delta(\mathbf{q} - \mathbf{q}'), \quad (2.3.37)$$

and the solution for $\mathbf{X}(t, \mathbf{x})$ reduces to

$$\mathbf{X}(t, \mathbf{x}) = \int d^3q \alpha(\mathbf{q}) \mathbf{X}_q(t) e^{i\mathbf{q}\cdot\mathbf{x}}, \quad (2.3.38)$$

with $\mathbf{X}_q(t)$ denoting the solutions for the scalar perturbations corresponding to the adiabatic mode, and the amplitude of all of its components is of course

set by \mathcal{R}_q^o . We will from now on only consider the adiabatic case and assume isocurvature perturbations are absent.

The discussion for the tensor modes is very similar. On the one hand it is slightly more complicated because of the index structure on the tensor perturbations, on the other hand it is simpler because there are only two independent modes. The solution to equation (2.3.28) can be written in the form

$$D_{ij}(t, \mathbf{x}) = \sum_{N, \lambda} \int d^3q \beta_N(\mathbf{q}, \lambda) e_{ij}(\hat{q}, \lambda) \mathcal{D}_{Nq}(t) e^{i\mathbf{q}\cdot\mathbf{x}}. \quad (2.3.39)$$

Here λ denotes the helicity of the mode, the index N labels the two linearly independent modes, $\beta(\mathbf{q}, \pm 2)$ and the polarization tensor $e_{ij}(\hat{q}, \pm 2)$ are the stochastic parameter and polarization tensor for helicity $\lambda = \pm 2$, normalized so that¹¹

$$\langle \beta_N(\mathbf{q}, \lambda) \beta_{N'}^*(\mathbf{q}', \lambda') \rangle = \delta_{NN'} \delta_{\lambda\lambda'} \delta^3(\mathbf{q} - \mathbf{q}'), \quad (2.3.40)$$

and for \hat{q} in the 3-direction

$$e_{11}(\hat{q}, \pm 2) = -e_{22}(\hat{q}, \pm 2) = 1/\sqrt{2}, e_{12}(\hat{q}, \pm 2) = e_{21}(\hat{q}, \pm 2) = \pm i/\sqrt{2}. \quad (2.3.41)$$

In writing equation (2.3.40), we have assumed not only that the distribution obeys translational invariance and rotational invariance but also that it respects parity. If this were not the case, the two different helicity states could have different powers. With an ansatz for the anisotropic stress of the same form

$$\pi_{ij}^T(t, \mathbf{x}) = \sum_{N\lambda} \int d^3q \beta_N(\mathbf{q}, \lambda) e_{ij}(\hat{q}, \lambda) \pi_{Nq}^T(t) e^{i\mathbf{q}\cdot\mathbf{x}}, \quad (2.3.42)$$

¹¹We should caution the reader that other normalizations frequently occur in the literature.

the equation for the tensor modes becomes

$$\ddot{\mathcal{D}}_{Nq} + 3\frac{\dot{a}}{a}\dot{\mathcal{D}}_{Nq} + \frac{q^2}{a^2}\mathcal{D}_{Nq} = 16\pi G\pi_{Nq}^T \quad (2.3.43)$$

In the absence of anisotropic inertia and for $q/a \ll H$, *i.e.* outside the horizon, it is easy to see that one of the solutions is a constant, which we will denote \mathcal{D}_q^o with the superscript *o* indicating that it is the value outside the horizon. This mode plays the same role for the tensor perturbations the adiabatic mode plays for the scalars. The other mode decays like $\int dt a(t)^{-3}$, and is negligible in any scenario where the fluctuations get generated early, *e.g.* in inflation. The tensor perturbation of the metric can then be written as

$$D_{ij}(t, \mathbf{x}) = \sum_{\lambda} \int d^3q \beta(\mathbf{q}, \lambda) e_{ij}(\hat{q}, \lambda) \mathcal{D}_q(t) e^{i\mathbf{q}\cdot\mathbf{x}}. \quad (2.3.44)$$

To make further progress in solving these equations, one has to specify a model. We will return to this in the next subsection after introducing the Boltzmann equations for neutrinos and photons.

2.3.2 The Boltzmann equation for photons

We can now turn to the Boltzmann equation that governs the time evolution of an ensemble of photons. We treat the photons as a gas of point particles that interact with the non-relativistic electrons through Thompson scattering. We will start by reviewing the collisionless Boltzmann equation for a gas of scalar particles and then generalize this to the case of photons.

Consider a map $X : \mathbb{R} \rightarrow M$ embedding the worldline of a scalar particle into the spacetime (M, g) with coordinates (t, x^i) . The world line action can be then written in the form

$$S = \frac{1}{2} \int d\tau e(\tau) g_{\mu\nu}(X(\tau)) \dot{X}^\mu \dot{X}^\nu, \quad (2.3.45)$$

where τ is an affine parameter parameterizing the worldline of the particle, $e(\tau)$ is an einbein that can be thought of as a Lagrange multiplier whose equation of motion enforces that the particle be on-shell, $g_{\mu\nu}$ is the metric on spacetime, and X^μ are the compositions of the embedding X with the coordinate functions, so that $g_{\mu\nu}(X(\tau))\dot{X}^\mu\dot{X}^\nu$ is the pullback of the metric on spacetime to the worldline of the particle. Varying with respect to the metric, fixing a gauge for the affine parameter such that $e(\tau) = 1$, and performing the integral over τ , we find that the stress energy tensor for a massless scalar particle is given by

$$T_{\mu\nu} = \frac{1}{\sqrt{-g}}\delta(\mathbf{x} - \mathbf{x}(t))\frac{p_\mu(t)p_\nu(t)}{p^0(\mathbf{x}, \mathbf{p}, t)}, \quad (2.3.46)$$

where we have introduced $p_\mu(t) \equiv g_{\mu\nu}\dot{X}^\nu$, used the notation $p^0(\mathbf{x}, \mathbf{p}, t) = \sqrt{g^{ij}(t, \mathbf{x})p_i p_j}$, and denoted the vector with components $X^i(t)$ by $\mathbf{x}(t)$. For a gas of non-interacting massless scalar particles, the stress energy tensor is simply the sum of the stress energy tensors of the individual particles in the gas and one has

$$\begin{aligned} T_{\mu\nu} &= \frac{1}{\sqrt{-g}} \sum_r \delta(\mathbf{x} - \mathbf{x}_r(t)) \frac{p_\mu^r(t)p_\nu^r(t)}{p_r^0(t)} \\ &= \frac{1}{\sqrt{-g}} \int d^3p \sum_r \delta(\mathbf{x} - \mathbf{x}_r(t)) \delta(\mathbf{p} - \mathbf{p}_r(t)) \frac{p_\mu p_\nu}{p^0} \\ &= \frac{1}{\sqrt{-g}} \int d^3p n(\mathbf{x}, \mathbf{p}, t) \frac{p_\mu p_\nu}{p^0}, \end{aligned} \quad (2.3.47)$$

where r labels the different particles, and we have defined the phase space density

$$n(\mathbf{x}, \mathbf{p}, t) \equiv \sum_r \delta(\mathbf{x} - \mathbf{x}_r(t)) \delta(\mathbf{p} - \mathbf{p}_r(t)). \quad (2.3.48)$$

To be specific, the components of \mathbf{x} are x^i and those of \mathbf{p} are p_i . This will be useful in its own right as we can describe the free-streaming neutrinos in this way, but we will for now move on to photons and return to neutrinos later.

In the case of an ensemble of photons, in addition to keeping track of the particles positions and momenta, we would also like to keep track of their polarizations. Instead of the phase space density we can then define a phase space density matrix

$$n^{ij}(\mathbf{x}, \mathbf{p}, t) \equiv \sum_{r,\lambda} \delta(\mathbf{x} - \mathbf{x}_r(t)) \delta(\mathbf{p} - \mathbf{p}_r(t)) \times P_r(\lambda) e^i(\mathbf{p}_r(t), \lambda) e^{j*}(\mathbf{p}_r(t), \lambda). \quad (2.3.49)$$

In this equation $P_r(\lambda)$ is the probability that the photon r has polarization λ , and the vectors $\mathbf{e}(\mathbf{p}, \lambda)$ are the polarization vectors for a photon travelling with three-momentum \mathbf{p} and polarization λ , and of course satisfy $p_i e^i = 0$ and $g_{ij} e^i(\mathbf{p}, \lambda) e^{j*}(\mathbf{p}, \lambda') = \delta_{\lambda\lambda'}$.

Since we will need this for the Einstein equations, let us also give the stress-energy tensor for the ensemble of photons in terms of the phase space density matrix. It is given by

$$T_{\mu\nu} = \frac{1}{\sqrt{-g}} \int d^3p g_{ij} n^{ij}(\mathbf{x}, \mathbf{p}, t) \frac{p_\mu p_\nu}{p^0}. \quad (2.3.50)$$

In the collisionless case, the time evolution of this density matrix can be worked out simply by taking its derivative with respect to time and collecting all the terms. This requires knowledge of the time derivatives of $\mathbf{x}(t)$, $\mathbf{p}(t)$, and $\mathbf{e}(\mathbf{p}_r(t), \lambda)$. The first of these can be taken from the definition of p_i written in the form

$$\frac{dX^i}{dt} = \frac{p^i}{p^0}. \quad (2.3.51)$$

where we have raised the index using $p^i \equiv g^{ij}p_j$. The second one can be taken from the geodesic equation in the form

$$\frac{dp_i}{dt} = \frac{p^k p^l}{2p^0} \frac{\partial g_{kl}}{\partial x^i}. \quad (2.3.52)$$

The time dependence of the polarization vector is somewhat more involved but can be obtained by parallel transporting the polarization vector along the worldline of the photon. Using $p_i e^i = 0$ and $g_{ij} e^i(\lambda) e^{j*}(\lambda') = \delta_{\lambda\lambda'}$ and decomposing \dot{e}^i into \dot{p}^i and the polarization vectors $e^i(\mathbf{p}, \lambda)$, one can choose a basis of polarization vectors such that

$$\frac{\partial e^i}{\partial t} = -e^m \frac{p^k p^l p^i}{2(p^0)^3} \frac{\partial g_{kl}}{\partial x^m} - \frac{1}{2} \left(\dot{g}_{kl} + \frac{p^m}{p^0} \frac{\partial g_{kl}}{\partial x^m} \right) e^k \left(g^{li} - \frac{p^l p^i}{(p^0)^2} \right), \quad (2.3.53)$$

The Boltzmann equation for the density matrix then becomes

$$\begin{aligned} \frac{\partial n^{ij}}{\partial t} + \frac{p^k}{p^0} \frac{\partial n^{ij}}{\partial x^k} + \frac{1}{2} \frac{p^k p^l}{p^0} \frac{\partial g^{kl}}{\partial x^m} \left(\frac{\partial n^{ij}}{\partial p_m} + \frac{p^j}{(p^0)^2} n^{mi} + \frac{p^i}{(p^0)^2} n^{mj} \right) \\ + \frac{1}{2} \left(\dot{g}_{kl} + \frac{p^m}{p^0} \frac{\partial g_{kl}}{\partial x^m} \right) \left[\left(g^{li} - \frac{p^l p^i}{(p^0)^2} \right) n^{kj} + \left(g^{lj} - \frac{p^l p^j}{(p^0)^2} \right) n^{ki} \right] = \mathcal{C}^{ij}, \end{aligned} \quad (2.3.54)$$

where \mathcal{C}^{ij} is the collision term. In our case it incorporates the interactions of the photons with the non-relativistic electrons in the plasma. Its derivation would lead us too far away from the main line of thought and we will content ourselves by quoting the result. The interested reader is referred to Appendix H of [38] for a derivation.

Let us at this point also specialize to the study of the evolution of the fluctuations of the density matrix around its equilibrium value rather than the full density matrix. To this end, let us write the density matrix as

$$n^{ij}(\mathbf{x}, \mathbf{p}, t) = \frac{1}{2} \bar{n}_\gamma(a(t)p^0) \left(g^{ij} - \frac{p^i p^j}{(p^0)^2} \right) + \delta n^{ij}(\mathbf{x}, \mathbf{p}, t). \quad (2.3.55)$$

The equation governing the time evolution of δn^{ij} is then given by

$$\begin{aligned} \frac{\partial \delta n^{ij}}{\partial t} + \frac{\hat{p}_m}{a} \frac{\partial \delta n^{ij}}{\partial x^m} + 2 \frac{\dot{a}}{a} \delta n^{ij} \\ - \frac{1}{4a^2} p \bar{n}'_\gamma \hat{p}_k \hat{p}_l \frac{\partial}{\partial t} \left(\frac{h_{kl}}{a^2} \right) (\delta^{ij} - \hat{p}_i \hat{p}_j) = \delta \mathcal{C}^{ij}. \end{aligned} \quad (2.3.56)$$

Here we have introduced the notation $p = \sqrt{\bar{p}_i \bar{p}_i}$, $\hat{p} = \mathbf{p}/p$ and the linearized collision term on the right hand side is given by

$$\begin{aligned} \delta \mathcal{C}^{ij}(\mathbf{x}, \mathbf{p}, t) = & -\omega_c(t) \delta n^{ij} + \frac{3\omega_c(t)}{8\pi} \int d^2 \hat{p}_1 [\delta n^{ij}(\mathbf{x}, |\mathbf{p}| \hat{p}_1, t) \\ & - \hat{p}_i \hat{p}_k \delta n^{kj}(\mathbf{x}, |\mathbf{p}| \hat{p}_1, t) - \hat{p}_j \hat{p}_k \delta n^{ik}(\mathbf{x}, |\mathbf{p}| \hat{p}_1, t) \\ & + \hat{p}_i \hat{p}_j \hat{p}_k \hat{p}_l \delta n^{kl}(\mathbf{x}, |\mathbf{p}| \hat{p}_1, t)] \\ & - \frac{\omega_c(t)}{2a^3(t)} (p_k \delta u_k) \bar{n}'_\gamma(p) (\delta^{ij} - \hat{p}_i \hat{p}_j). \end{aligned} \quad (2.3.57)$$

where $\omega_c(t)$ is the collision rate of a photon with the electrons of the plasma. It is given by $\omega_c(t) = n_e(t) \sigma_T$, where $n_e(t)$ is the number density of electrons in the baryonic plasma¹² and σ_T is the Thompson cross section given by¹³

$$\sigma_T = \frac{8\pi}{3} \frac{\alpha^2}{m^2}. \quad (2.3.58)$$

For the linearized Einstein equations, it will be useful to know how the perturbations in the stress energy tensor are related to the fluctuations of the density matrix. A straightforward but somewhat tedious calculation shows

¹²This number density requires knowledge of the ionized fraction of hydrogen and helium during recombination and later during reionization. The derivation is full of interesting physics, but it would take us too far away from the main line of thought in our review. The interested reader is invited to read about it in [38], but for practical purposes the reader need not know the derivation and can use the publicly available recast code [71], [72].

¹³See *e.g.* [73] for a derivation.

that the first order perturbations of (2.3.50) are given by

$$\delta T^0_0 = -\frac{1}{a^4} \int d^3p a^2 \delta n^{kk}(\mathbf{x}, \mathbf{p}, t) p, \quad (2.3.59)$$

$$\delta T^0_i = \frac{1}{a^3} \int d^3p a^2 \delta n^{kk}(\mathbf{x}, \mathbf{p}, t) p_i, \quad (2.3.60)$$

$$\delta T^i_j = \frac{1}{a^4} \int d^3p a^2 \delta n^{kk}(\mathbf{x}, \mathbf{p}, t) \frac{p_i p_j}{p}. \quad (2.3.61)$$

The dimensionless intensity matrix is related to the fluctuations in the density matrix according to

$$a^4(t) \bar{\rho}_\gamma(t) J_{ij}(\mathbf{x}, \hat{p}, t) = a^2(t) \int_0^\infty 4\pi p^3 dp \delta n^{ij}(\mathbf{x}, |\mathbf{p}| \hat{p}, t), \quad (2.3.62)$$

At this point it will be convenient to continue separately for the scalar and tensor modes, but since this will be needed later, let us first briefly give the perturbations in the stress-energy tensor in terms of J_{ij} . They are given by

$$\delta T^0_0 = -\bar{\rho}_\gamma(t) \int \frac{d^2\hat{p}}{4\pi} J_{kk}(\mathbf{x}, \hat{p}, t), \quad (2.3.63)$$

$$\delta T^0_i = a(t) \bar{\rho}_\gamma(t) \int \frac{d^2\hat{p}}{4\pi} J_{kk}(\mathbf{x}, \hat{p}, t) \hat{p}_i, \quad (2.3.64)$$

$$\delta T^i_j = \bar{\rho}_\gamma(t) \int \frac{d^2\hat{p}}{4\pi} J_{kk}(\mathbf{x}, \hat{p}, t) \hat{p}_i \hat{p}_j. \quad (2.3.65)$$

2.3.2.1 Scalar modes

Introducing the Fourier transforms of the intensity matrix

$$J_{ij}(\mathbf{x}, \hat{p}, t) = \int d^3q J_{ij}(\mathbf{q}, \hat{p}, t) e^{i\mathbf{q}\cdot\mathbf{x}}, \quad (2.3.66)$$

and of the velocity potential

$$\delta u_i(t, \mathbf{x}) = \int d^3q \delta u_i(t, \mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{x}}, \quad (2.3.67)$$

we can then use equations (2.3.56) and (2.3.57) to derive an equation for the dimensionless intensity matrix. For the scalar modes this equation becomes

$$\begin{aligned}
\frac{\partial J_{ij}}{\partial t} + i \frac{\hat{\mathbf{p}} \cdot \mathbf{q}}{a(t)} J_{ij} + \alpha(\mathbf{q}) \left[\dot{A}_q - (\mathbf{q} \cdot \hat{\mathbf{p}})^2 \dot{B}_q \right] (\delta_{ij} - \hat{p}_i \hat{p}_j) = \\
-\omega_c(t) J_{ij} + \frac{3\omega_c(t)}{8\pi} \int d^2 \hat{\mathbf{p}}_1 [J_{ij}(\mathbf{q}, \hat{\mathbf{p}}_1, t) - \hat{p}_i \hat{p}_k J_{kj}(\mathbf{q}, \hat{\mathbf{p}}_1, t) \\
-\hat{p}_j \hat{p}_k J_{ik}(\mathbf{q}, \hat{\mathbf{p}}_1, t) + \hat{p}_i \hat{p}_j \hat{p}_k \hat{p}_l J_{kl}(\mathbf{q}, \hat{\mathbf{p}}_1, t)] \\
+ \frac{2\omega_c(t)}{a(t)} (\hat{p}_k \delta u_k(t, \mathbf{q})) (\delta^{ij} - \hat{p}_i \hat{p}_j) . \tag{2.3.68}
\end{aligned}$$

We have dropped the arguments, but J_{ij} stands for the Fourier transform of the dimensionless intensity matrix. Remember also that we work under the assumption that only the adiabatic mode is excited, implying that there is only a single stochastic parameter $\alpha(\mathbf{q})$. Except for a trivial overall dependence on $\alpha(\mathbf{q})$, there is no preferred direction in the problem. This allows for the commonly used decomposition of the dimensionless intensity matrix of the form

$$\begin{aligned}
J_{ij}(\mathbf{q}, \hat{\mathbf{p}}, t) = \alpha(\mathbf{q}) \left\{ \frac{1}{2} \left(\Delta_T^{(S)}(q, \hat{\mathbf{q}} \cdot \hat{\mathbf{p}}, t) - \Delta_P^{(S)}(q, \hat{\mathbf{q}} \cdot \hat{\mathbf{p}}, t) \right) (\delta_{ij} - \hat{p}_i \hat{p}_j) \right. \\
\left. + \Delta_P^{(S)}(q, \hat{\mathbf{q}} \cdot \hat{\mathbf{p}}, t) \left[\frac{(\hat{q}_i - (\hat{\mathbf{q}} \cdot \hat{\mathbf{p}}) \hat{p}_i) (\hat{q}_j - (\hat{\mathbf{q}} \cdot \hat{\mathbf{p}}) \hat{p}_j)}{1 - (\hat{\mathbf{p}} \cdot \hat{\mathbf{q}})^2} \right] \right\} . \tag{2.3.69}
\end{aligned}$$

Here the superscripts (S) indicate that we are dealing with the scalar modes, the subscript T stands for temperature and the subscript P for polarization. The coefficients are chosen such that the temperature fluctuations are proportional to $\Delta_T^{(S)}$ and the polarization is proportional to $\Delta_P^{(S)}$. If we further introduce so called source functions $\Phi(q, t)$ and $\Pi(q, t)$ via

$$\int \frac{d^2 \hat{\mathbf{p}}}{4\pi} J_{ij}(\mathbf{q}, \hat{\mathbf{p}}, t) = \alpha(\mathbf{q}) \left[\delta_{ij} \Phi(q, t) + \frac{1}{2} \hat{q}_i \hat{q}_j \Pi(q, t) \right] , \tag{2.3.70}$$

and use the notation $\mu = \hat{q} \cdot \hat{p}$, we are lead to two coupled Boltzmann equations for the quantities $\Delta_T^{(S)}$ and $\Delta_P^{(S)}$ of the form

$$\begin{aligned} \dot{\Delta}_P^{(S)}(q, \mu, t) + i \frac{q\mu}{a(t)} \Delta_P^{(S)}(q, \mu, t) = \\ -\omega_c(t) \Delta_P^{(S)}(q, \mu, t) + \frac{3}{4} \omega_c(t) (1 - \mu^2) \Pi(q, t) \end{aligned} \quad (2.3.71)$$

$$\begin{aligned} \dot{\Delta}_T^{(S)}(q, \mu, t) + i \frac{q\mu}{a(t)} \Delta_T^{(S)}(q, \mu, t) = -\omega_c(t) \Delta_T^{(S)}(q, \mu, t) \\ + \frac{3}{4} \omega_c(t) (1 - \mu^2) \Pi(q, t) + 3\omega_c(t) \Phi(q, t) \\ + \frac{4iq\mu}{a(t)} \omega_c(t) \delta u_{Bq}(t) - 2\dot{A}_q(t) + 2q^2 \mu^2 \dot{B}_q(t). \end{aligned} \quad (2.3.72)$$

One way to solve this system of equations is by decomposing $\Delta_T^{(S)}$ and $\Delta_P^{(S)}$ into multipole moments as

$$\Delta_T^{(S)}(q, \mu, t) = \sum_{\ell=0}^{\infty} (-i)^\ell (2\ell + 1) P_\ell(\mu) \Delta_{T,\ell}^{(S)}(q, t), \quad (2.3.73)$$

$$\Delta_P^{(S)}(q, \mu, t) = \sum_{\ell=0}^{\infty} (-i)^\ell (2\ell + 1) P_\ell(\mu) \Delta_{P,\ell}^{(S)}(q, t). \quad (2.3.74)$$

In terms of these, the Boltzmann equations for $\Delta_T^{(S)}$ and $\Delta_P^{(S)}$ give rise to the Boltzmann hierarchy for $\Delta_{T,\ell}^{(S)}$ and $\Delta_{P,\ell}^{(S)}$

$$\begin{aligned} \dot{\Delta}_{P,\ell}^{(S)}(q, t) + \frac{q}{a(2\ell + 1)} \left[(\ell + 1) \Delta_{P,\ell+1}^{(S)}(q, t) - \ell \Delta_{P,\ell-1}^{(S)}(q, t) \right] \\ = -\omega_c(t) \Delta_{P,\ell}^{(S)}(q, t) + \frac{1}{2} \omega_c(t) \Pi(q, t) (\delta_{\ell,0} + \delta_{\ell,2}), \end{aligned} \quad (2.3.75)$$

$$\begin{aligned} \dot{\Delta}_{T,\ell}^{(S)}(q, t) + \frac{q}{a(2\ell + 1)} \left[(\ell + 1) \Delta_{T,\ell+1}^{(S)}(q, t) - \ell \Delta_{T,\ell-1}^{(S)}(q, t) \right] \\ = -\omega_c(t) \Delta_{T,\ell}^{(S)}(q, t) - 2\dot{A}_q \delta_{\ell,0} + 2q^2 \dot{B}_q \left(\frac{1}{3} \delta_{\ell,0} - \frac{2}{15} \delta_{\ell,2} \right) \\ + \omega_c \left(3\Phi + \frac{1}{2} \Pi \right) \delta_{\ell,0} + \frac{1}{2} \omega_c \Pi \delta_{\ell,2} - \frac{4q}{3a} \omega_c \delta u_{Bq} \delta_{\ell,1}. \end{aligned} \quad (2.3.76)$$

The source functions Φ and Π can also be expressed in terms of these multipole coefficients

$$\Phi = \frac{1}{6} \left[2\Delta_{T,0}^{(S)} - \Delta_{T,2}^{(S)} - \Delta_{P,0}^{(S)} - \Delta_{P,2}^{(S)} \right], \quad (2.3.77)$$

$$\Pi = \Delta_{P,0}^{(S)} + \Delta_{T,2}^{(S)} + \Delta_{P,2}^{(S)}. \quad (2.3.78)$$

To turn this into a closed system, one thus only needs to know the metric components appearing in these equations as well as the velocity potential. Their time evolution is of course governed by the Einstein equations, and we give the complete system of equations in the next subsection. To do this, we will also need to know the components of the stress-energy tensor in terms of the multipole moments. Using equations (2.3.16), (2.3.32), and (2.3.69), one finds

$$\delta\rho_{\gamma q} = \bar{\rho}_\gamma \Delta_{T,0}^{(S)}, \quad (2.3.79)$$

$$\delta p_{\gamma q} = \frac{\bar{\rho}_\gamma}{3} \left(\Delta_{T,0}^{(S)} + \Delta_{T,2}^{(S)} \right), \quad (2.3.80)$$

$$\delta u_{\gamma q} = -\frac{3a}{4q} \Delta_{T,1}^{(S)}, \quad (2.3.81)$$

$$q^2 \pi_{\gamma q}^S = \bar{\rho}_\gamma \Delta_{T,2}^{(S)}. \quad (2.3.82)$$

Another way to find a somewhat formal solution to the coupled Boltzmann equations for $\Delta_T^{(S)}$ and $\Delta_P^{(S)}$ is to treat the system as an inhomogeneous system of ordinary differential equations and solve it as if the sources were known. This solution is commonly referred to as integrating the equations

along the line of sight [74] and takes the form

$$\begin{aligned} \Delta_T^{(S)}(q, \mu, t_0) &= \int_{t_1}^{t_0} dt \exp[-iqr(t)\mu] P(t) \\ &\times \left[3\Phi(q, t) + \frac{4i\mu q}{a(t)} \delta u_q(t) + \frac{3}{4}(1 - \mu^2)\Pi(q, t) \right] \\ &+ \int_{t_1}^{t_0} dt \exp \left[-iqr(t)\mu - \int_t^{t_0} \omega_c(t') dt' \right] \\ &\times \left(-2\dot{A}_q(t) + 2\mu^2 q^2 \dot{B}_q(t) \right), \quad (2.3.83) \end{aligned}$$

$$\Delta_P^{(S)}(q, \mu, t_0) = \frac{3}{4}(1 - \mu^2) \int_{t_1}^{t_0} dt \exp[-iqr(t)\mu] P(t)\Pi(q, t). \quad (2.3.84)$$

In these equations, t_1 is any time taken early enough before recombination so that any photon present at time t_1 would have collided many times before the present, $r(t) = \int_t^{t_0} dt'/a(t')$ is the co-moving radial coordinate of a source from which light emitted at time t would reach us at the origin at the present time t_0 , and we have introduced the probability distribution of last scattering $P(t) = \omega_c(t) \exp[-\int_t^{t_0} dt' \omega_c(t')]$. It is convenient to rewrite $\Delta_T^{(S)}$ slightly. Using the identity

$$\begin{aligned} \exp[-iqr(t)\mu] \mu^2 q^2 \dot{B}_q(t) &= \\ &- \exp[-iqr(t)\mu] \frac{d}{dt} \left[a^2 \ddot{B}_q(t) + a\dot{a}\dot{B}_q(t) \right] \\ &+ \frac{d}{dt} \left\{ \exp[-iqr(t)\mu] \left[a^2 \ddot{B}_q(t) + a\dot{a}\dot{B}_q(t) + iaq\mu\dot{B}_q(t) \right] \right\}, \quad (2.3.85) \end{aligned}$$

and integrating by parts. This yields

$$\Delta_T^{(S)}(q, \mu, t_0) = \int_{t_1}^{t_0} dt \exp[-iqr(t)\mu] P(t)$$

$$\begin{aligned}
& \times \left[3\Phi(q, t) - 2a^2(t)\ddot{B}_q(t) - 2a(t)\dot{a}(t)\dot{B}_q(t) \right. \\
& \quad \left. + 4i\mu q \left(\delta u_q(t)/a(t) + a(t)\dot{B}_q(t)/2 \right) + \frac{3}{4}(1 - \mu^2)\Pi(q, t) \right] \\
& - \int_{t_1}^{t_0} dt \exp \left[-iqr(t)\mu - \int_t^{t_0} \omega_c(t')dt' \right] \\
& \quad \times \frac{d}{dt} \left(2A_q(t) + 2a^2(t)\ddot{B}_q(t) + 2a(t)\dot{a}(t)\dot{B}_q(t) \right), \quad (2.3.86)
\end{aligned}$$

$$\Delta_P^{(S)}(q, \mu, t_0) = \frac{3}{4}(1 - \mu^2) \int_{t_1}^{t_0} dt \exp[-iqr(t)\mu] P(t)\Pi(q, t). \quad (2.3.87)$$

Notice that the integrand of the first contribution to $\Delta_T^{(S)}$ as well as the integrand in $\Delta_P^{(S)}$ are proportional to $P(t)$ implying that they only receive contributions when free electrons are present, *i.e.* around the time of last scattering and once the universe has become reionized. In contrast, the last term in the temperature amplitude is not proportional to $P(t)$ and receives contributions from all times. This contribution is referred to as the integrated Sachs-Wolfe contribution. It is dominated by the period when the vacuum energy becomes important, and contributes significantly only for $\ell \lesssim 20$.

The beauty of these equations is that once the time evolution of the sources, the metric fluctuations, and the velocity potential are known, they can be used to calculate $\Delta_{T,\ell}^{(S)}$ and $\Delta_{P,\ell}^{(S)}$ for arbitrary ℓ . Also, notice that the source functions as well as the Einstein equations only require knowledge of $\Delta_{T,\ell}^{(S)}$ and $\Delta_{P,\ell}^{(S)}$ up to $\ell = 2$. An efficient way to calculate $\Delta_{T,\ell}^{(S)}$ and $\Delta_{P,\ell}^{(S)}$ then is to use a truncation of the Boltzmann hierarchy at a value of ℓ that is high enough so that $\Delta_{T,\ell}^{(S)}$ and $\Delta_{P,\ell}^{(S)}$ are known to the desired accuracy up to $\ell = 2$, and then use the line of sight integrals to evaluate the higher ℓ -values [74]. To be specific, the default setting in CMBfast for the value at which the Boltzmann

hierarchy for the photons is truncated is $\ell = 12$ for the scalar modes.

2.3.2.2 Tensor modes

Using an ansatz for the dimensionless intensity matrix of the form

$$J_{ij}(\mathbf{x}, \hat{p}, t) = \sum_{\lambda=\pm 2} \int d^3q \beta(\mathbf{q}, \lambda) J_{ij}(\mathbf{q}, \hat{p}, t, \lambda) e^{i\mathbf{q}\cdot\mathbf{x}}, \quad (2.3.88)$$

we can again use equation (2.3.56) to derive an equation for $J_{ij}(\mathbf{q}, \hat{p}, t, \lambda)$. For the tensor mode this equation becomes

$$\begin{aligned} \frac{\partial J_{ij}}{\partial t} + i \frac{\hat{p} \cdot \mathbf{q}}{a(t)} J_{ij} + \hat{p}_k \hat{p}_l e_{kl}(\hat{q}, \lambda) \dot{D}_q(t) (\delta_{ij} - \hat{p}_i \hat{p}_j) = \\ -\omega_c(t) J_{ij} + \frac{3\omega_c(t)}{8\pi} \int d^2\hat{p}_1 [J_{ij}(\mathbf{q}, \hat{p}_1, t, \lambda) - \hat{p}_i \hat{p}_k J_{kj}(\mathbf{q}, \hat{p}_1, t, \lambda) \\ - \hat{p}_j \hat{p}_k J_{ik}(\mathbf{q}, \hat{p}_1, t, \lambda) + \hat{p}_i \hat{p}_j \hat{p}_k \hat{p}_l J_{kl}(\mathbf{q}, \hat{p}_1, t, \lambda)]. \end{aligned} \quad (2.3.89)$$

We can again introduce a source function. Rotational invariance ensures that the integrals

$$\int \frac{d^2\hat{p}_1}{4\pi} J_{ij}(\mathbf{q}, \hat{p}_1, t, \lambda), \quad (2.3.90)$$

must be proportional to $e_{ij}(\hat{q}, \lambda)$ times some source function of q and time.

This source function is commonly defined as

$$\int \frac{d^2\hat{p}_1}{4\pi} J_{ij}(\mathbf{q}, \hat{p}_1, t, \lambda) = -\frac{2}{3} e_{ij}(\hat{q}, \lambda) \Psi(q, t). \quad (2.3.91)$$

Furthermore, in the tensor case, we can decompose the dimensionless intensity matrix as

$$\begin{aligned} J_{ij}(\mathbf{q}, \hat{p}, t, \lambda) = \\ \frac{1}{2} (\delta_{ij} - \hat{p}_i \hat{p}_j) \hat{p}_k \hat{p}_l e_{kl}(\hat{q}, \lambda) \left(\tilde{\Delta}_T^{(T)}(q, \hat{p} \cdot \hat{q}, t) + \tilde{\Delta}_P^{(T)}(q, \hat{p} \cdot \hat{q}, t) \right) \\ + \left(e_{ij}(\hat{q}, \lambda) - \hat{p}_i \hat{p}_k e_{kj}(\hat{q}, \lambda) - \hat{p}_j \hat{p}_k e_{ik}(\hat{q}, \lambda) + \hat{p}_i \hat{p}_j \hat{p}_k \hat{p}_l e_{kl}(\hat{q}, \lambda) \right) \\ \times \tilde{\Delta}_P^{(T)}(q, \hat{p} \cdot \hat{q}, t). \end{aligned} \quad (2.3.92)$$

The superscript (T) indicates that we are dealing with the tensor modes, the subscripts T and P stand for temperature and polarization, respectively. Again, the coefficients are chosen such that the temperature fluctuations are proportional to $\tilde{\Delta}_T^{(T)}$ and the polarization is proportional to $\tilde{\Delta}_P^{(T)}$. This decomposition leads to two coupled Boltzmann equations for $\tilde{\Delta}_T^{(T)}$ and $\tilde{\Delta}_P^{(T)}$. In the tensor case, these are of the form

$$\begin{aligned} \dot{\tilde{\Delta}}_T^{(T)}(q, \mu, t) + i \frac{q\mu}{a(t)} \tilde{\Delta}_T^{(T)}(q, \mu, t) = \\ - \omega_c(t) \tilde{\Delta}_T^{(T)}(q, \mu, t) - 2\dot{\mathcal{D}}_q(t) + \omega_c(t) \Psi(q, t), \end{aligned} \quad (2.3.93)$$

and

$$\begin{aligned} \dot{\tilde{\Delta}}_P^{(T)}(q, \mu, t) + i \frac{q\mu}{a(t)} \tilde{\Delta}_P^{(T)}(q, \mu, t) = \\ - \omega_c(t) \tilde{\Delta}_P^{(T)}(q, \mu, t) - \omega_c(t) \Psi(q, t). \end{aligned} \quad (2.3.94)$$

Traditionally, this system of equations is again solved by decomposing $\tilde{\Delta}_T^{(T)}$ and $\tilde{\Delta}_P^{(T)}$ into multipole moments as

$$\tilde{\Delta}_T^{(T)}(q, \mu, t) = \sum_{\ell=0}^{\infty} (-i)^\ell (2\ell + 1) P_\ell(\mu) \tilde{\Delta}_{T,\ell}^{(T)}(q, t), \quad (2.3.95)$$

$$\tilde{\Delta}_P^{(T)}(q, \mu, t) = \sum_{\ell=0}^{\infty} (-i)^\ell (2\ell + 1) P_\ell(\mu) \tilde{\Delta}_{P,\ell}^{(T)}(q, t). \quad (2.3.96)$$

This leads to the Boltzmann hierarchy

$$\begin{aligned} \dot{\tilde{\Delta}}_{T,\ell}^{(T)}(q, t) + \frac{q}{a(2\ell + 1)} \left[(\ell + 1) \tilde{\Delta}_{T,\ell+1}^{(T)}(q, t) - \ell \tilde{\Delta}_{T,\ell-1}^{(T)}(q, t) \right] \\ = \left(-2\dot{\mathcal{D}}_q(t) + \omega_c(t) \Psi(q, t) \right) \delta_{\ell,0} - \omega_c(t) \tilde{\Delta}_{T,\ell}^{(T)}(q, t), \end{aligned} \quad (2.3.97)$$

$$\begin{aligned} \dot{\tilde{\Delta}}_{P,\ell}^{(T)}(q, t) + \frac{q}{a(2\ell + 1)} \left[(\ell + 1) \tilde{\Delta}_{P,\ell+1}^{(T)}(q, t) - \ell \tilde{\Delta}_{P,\ell-1}^{(T)}(q, t) \right] \\ = -\omega_c(t) \Psi(q, t) \delta_{\ell,0} - \omega_c(t) \tilde{\Delta}_{P,\ell}^{(T)}(q, t), \end{aligned} \quad (2.3.98)$$

with the source function $\Psi(q, t)$ given in terms of these multipole moments as

$$\begin{aligned} \Psi(q, t) = & \frac{1}{10} \tilde{\Delta}_{T,0}^{(T)}(q, t) + \frac{1}{7} \tilde{\Delta}_{T,2}^{(T)}(q, t) + \frac{3}{70} \tilde{\Delta}_{T,4}^{(T)}(q, t) \\ & - \frac{3}{5} \tilde{\Delta}_{P,0}^{(T)}(q, t) + \frac{6}{7} \tilde{\Delta}_{P,2}^{(T)}(q, t) - \frac{3}{70} \tilde{\Delta}_{P,4}^{(T)}(q, t) . \end{aligned} \quad (2.3.99)$$

To write down the linearized Einstein equations for the tensor mode, we also need to know the anisotropic stress in terms of the multipole moments. It is given by

$$\pi_{\gamma q}^T(t) = 2\bar{\rho}_\gamma(t) \left[\frac{1}{15} \tilde{\Delta}_{T,0}^{(T)}(q, t) + \frac{2}{21} \tilde{\Delta}_{T,2}^{(T)}(q, t) + \frac{1}{35} \tilde{\Delta}_{T,4}^{(T)}(q, t) \right] . \quad (2.3.100)$$

Again a formal solution to the coupled Boltzmann equations for $\tilde{\Delta}_T^{(T)}$ and $\tilde{\Delta}_P^{(T)}$ can be found by treating the system as an inhomogeneous system of ordinary differential equations and solve it as if the sources were known. The line of sight solution for the tensor modes takes the form

$$\begin{aligned} \tilde{\Delta}_T^{(T)}(q, \mu, t_0) = & -\tilde{\Delta}_P^{(T)}(q, \mu, t_0) \\ & -2 \int_{t_1}^{t_0} dt \exp \left[-iq\mu \int_t^{t_0} \frac{dt'}{a(t')} \right] \exp \left[-\int_t^{t_0} \omega_c(t') dt' \right] \dot{\mathcal{D}}_q(t) , \end{aligned} \quad (2.3.101)$$

$$\tilde{\Delta}_P^{(T)}(q, \mu, t_0) = - \int_{t_1}^{t_0} dt \exp \left[-iq\mu \int_t^{t_0} \frac{dt'}{a(t')} \right] P(t) \Psi(q, t) , \quad (2.3.102)$$

where t_1 is again any time taken early enough before recombination so that any photon present at time t_1 would have collided many times before the present, and $P(t) = \omega_c(t) \exp[-\int_t^{t_0} dt' \omega_c(t')]$ is the probability distribution of last scattering. Notice that the integrand in $\tilde{\Delta}_P^{(T)}$ is again proportional to $P(t)$ so that the polarization amplitude only receives contributions from a period around the time of last scattering and once the universe has become reionized while the temperature amplitude again receives contributions from all times.

In the tensor case the source function as well as the Einstein equations require knowledge of $\tilde{\Delta}_{T,\ell}^{(T)}$ and $\tilde{\Delta}_{P,\ell}^{(T)}$ up to $\ell = 4$. These can again be calculated using a truncation of the Boltzmann hierarchy. Once these are known, the line of sight solution can again be used to evaluate the higher ℓ -values. The default setting in CMBfast for the value at which the Boltzmann hierarchy is truncated is $\ell = 10$ in the case of the tensor modes.

One can also use the line of sight integral as an alternative of the truncated Boltzmann hierarchy altogether. One can derive an integral equation for the source function by integrating equation (2.3.92) over \hat{p} and equating coefficients of e_{ij} . The resulting integral equation is [75], [76]

$$\begin{aligned} \Psi(q, t) = & \frac{3}{2} \int_{t_1}^t dt' \exp \left[- \int_{t'}^t \omega_c(t'') dt'' \right] \\ & \times \left[-2\dot{D}_q(t') K \left(q \int_{t'}^t \frac{dt''}{a(t'')} \right) + \omega_c(t') F \left(q \int_{t'}^t \frac{dt''}{a(t'')} \right) \Psi(q, t') \right], \end{aligned} \quad (2.3.103)$$

where $K(v)$ and $F(v)$ are the functions

$$K(v) \equiv j_2(v)/v^2 \quad \text{and} \quad F(v) \equiv j_0(v) - 2j_1(v)/v + 2j_2(v)/v^2, \quad (2.3.104)$$

with $j_\ell(v)$ the spherical Bessel functions.

A similar approach for the calculation of the source functions could be taken for the scalar modes, but the equations there become somewhat long, and the combination of truncated Boltzmann hierarchy and line of sight integration seems easier to implement efficiently.

2.3.3 The Boltzmann equation for neutrinos

We can now return to the neutrino contribution and review their Boltzmann equations. Even though at least two species of neutrinos are known to

have a mass (see *e.g.* [77]), we will treat them as if they were massless¹⁴ and refer the interested reader to [79], [80], and [81]. We are far from observing a cosmic neutrino background and do not care about the intensity fluctuations in neutrinos directly, but it turns out to be convenient to use the same formalism as for the photons. The evolution of the neutrinos can conveniently be described with a phase space density as in equation (2.3.48). The Boltzmann equation governing its time evolution can be derived simply by taking the derivative with respect to time using

$$\frac{dX^i}{dt} = \frac{p^i}{p^0} \quad \text{and} \quad \frac{dp_i}{dt} = \frac{p^k p^l}{2p^0} \frac{\partial g_{kl}}{\partial x^i}. \quad (2.3.105)$$

One finds

$$\frac{\partial n}{\partial t} + \frac{p^k}{p^0} \frac{\partial n}{\partial x^k} + \frac{1}{2} \frac{p^k p^l}{p^0} \frac{\partial g^{kl}}{\partial x^m} \frac{\partial n}{\partial p_m} = 0. \quad (2.3.106)$$

The temperatures we are interested in are too low for neutrinos to interact efficiently and they can be treated as free-streaming. The collision term is thus absent. It is convenient to define the fluctuations by

$$n(\mathbf{x}, \mathbf{p}, t) = \bar{n}(a(t)p^0(\mathbf{x}, \mathbf{p}, t)) + \delta n(\mathbf{x}, \mathbf{p}, t), \quad (2.3.107)$$

where the equilibrium density \bar{n} is given by

$$\bar{n}(p) = \frac{1}{(2\pi)^3} \frac{1}{\exp[p/k_B a(t)\bar{T}(t)]}, \quad (2.3.108)$$

where $\bar{T}(t)$ is the neutrino temperature which, after electrons and positrons have annihilated, and is lower than the one of photons. It scales like $a^{-1}(t)$ so that $\bar{n}(p)$ is constant. The fluctuations satisfy the Boltzmann equation

$$\frac{\partial \delta n}{\partial t} + \frac{\hat{p}_m}{a} \frac{\partial \delta n}{\partial x^m} - \frac{1}{2} p \bar{n}_\gamma \hat{p}_k \hat{p}_l \frac{\partial}{\partial t} \left(\frac{h_{kl}}{a^2} \right) = 0. \quad (2.3.109)$$

¹⁴This is a good approximation when it comes to the cosmic microwave background, but cosmology is starting to put meaningful bounds on their masses [78].

The perturbations in the stress-energy tensor in terms of these are given by

$$\delta T^0_0 = -\frac{1}{a^4} \int d^3p \delta n(\mathbf{x}, \mathbf{p}, t) p, \quad (2.3.110)$$

$$\delta T^0_i = \frac{1}{a^3} \int d^3p \delta n(\mathbf{x}, \mathbf{p}, t) p_i, \quad (2.3.111)$$

$$\delta T^i_j = \frac{1}{a^4} \int d^3p \delta n(\mathbf{x}, \mathbf{p}, t) \frac{p_i p_j}{p}. \quad (2.3.112)$$

In analogy with the treatment of the photons, one can introduce a dimensionless density perturbation according to

$$a^4(t) \bar{\rho}_\nu(t) J(\mathbf{x}, \hat{p}, t) = N_\nu \int_0^\infty 4\pi p^3 dp \delta n(\mathbf{x}, |\mathbf{p}| \hat{p}, t). \quad (2.3.113)$$

In terms of it, the perturbations in the stress-energy tensor are of the form

$$\delta T^0_0 = -\bar{\rho}_\nu(t) \int \frac{d^2\hat{p}}{4\pi} J(\mathbf{x}, \hat{p}, t), \quad (2.3.114)$$

$$\delta T^0_i = a(t) \bar{\rho}_\nu(t) \int \frac{d^2\hat{p}}{4\pi} J(\mathbf{x}, \hat{p}, t) \hat{p}_i, \quad (2.3.115)$$

$$\delta T^i_j = \bar{\rho}_\nu(t) \int \frac{d^2\hat{p}}{4\pi} J(\mathbf{x}, \hat{p}, t) \hat{p}_i \hat{p}_j. \quad (2.3.116)$$

Let us now again continue the discussion separately for scalar and tensor modes.

2.3.3.1 Scalar modes

The Fourier components of $J(\mathbf{x}, \hat{p}, t)$, for the scalar case defined via

$$J(\mathbf{x}, \hat{p}, t) = \int d^3q \alpha(\mathbf{q}) \Delta_\nu^{(S)}(q, \hat{q} \cdot \hat{p}) e^{i\mathbf{q} \cdot \mathbf{x}}, \quad (2.3.117)$$

satisfy a Boltzmann equation of the form

$$\dot{\Delta}_\nu^{(S)}(q, \mu, t) + i \frac{q\mu}{a(t)} \Delta_\nu^{(S)}(q, \mu, t) = -2\dot{A}_q(t) + 2q^2 \mu^2 \dot{B}_q(t). \quad (2.3.118)$$

Just like the Boltzmann equations for the photons, it is useful to decompose $\Delta_\nu^{(S)}(q, \mu, t)$ into multipole moments according to

$$\Delta_\nu^{(S)}(q, \mu, t) = \sum_{\ell=0}^{\infty} (-i)^\ell (2\ell + 1) P_\ell(\mu) \Delta_{\nu, \ell}^{(S)}(q, t). \quad (2.3.119)$$

In terms of these, the Boltzmann equation for $\Delta_\nu^{(S)}$ gives rise to the Boltzmann hierarchy for $\Delta_{\nu, \ell}^{(S)}$

$$\begin{aligned} \dot{\Delta}_{\nu, \ell}^{(S)}(q, t) + \frac{q}{a(2\ell + 1)} \left[(\ell + 1) \Delta_{\nu, \ell+1}^{(S)}(q, t) - \ell \Delta_{\nu, \ell-1}^{(S)}(q, t) \right] = \\ - 2\dot{A}_q \delta_{\ell, 0} + 2q^2 \dot{B}_q \left(\frac{1}{3} \delta_{\ell, 0} - \frac{2}{15} \delta_{\ell, 2} \right). \end{aligned} \quad (2.3.120)$$

To write down the Einstein equations, we will again need the fluctuations of the stress-energy tensor in terms of the multipole moments. For the neutrinos, they are given by

$$\delta\rho_{\nu q} = \bar{\rho}_\nu \Delta_{\nu, 0}^{(S)}, \quad (2.3.121)$$

$$\delta p_{\nu q} = \frac{\bar{\rho}_\nu}{3} \left(\Delta_{\nu, 0}^{(S)} + \Delta_{\nu, 2}^{(S)} \right), \quad (2.3.122)$$

$$\delta u_{\nu q} = -\frac{3a}{4q} \Delta_{\nu, 1}^{(S)}, \quad (2.3.123)$$

$$q^2 \pi_{\nu q}^S = \bar{\rho}_\nu \Delta_{\nu, 2}^{(S)}. \quad (2.3.124)$$

For the neutrinos, we only need the low multipole moments. The truncation of the Boltzmann hierarchy is thus sufficient, and no line of sight integral is needed to calculate the higher multipole moments. This is what is done in the numerical codes like CMBfast [82]¹⁵ and CAMB [83].¹⁶

One can in this case, however use the line of sight solution *instead of* the Boltzmann hierarchy. This was used in some codes used to generate results

¹⁵<http://www.cfa.harvard.edu/~mzaldarr/CMBFAST/cmbfast.html>

¹⁶<http://camb.info/>.

for later chapters in the case of the tensor modes and we will have to say more about it there. For the scalar modes, we have not made use of it and will simply refer the interested reader to [38].

2.3.3.2 Tensor modes

For the tensor modes, we can define the Fourier components of $J(\mathbf{x}, \hat{p}, t)$ as

$$J(\mathbf{x}, \hat{p}, t) = \sum_{\lambda=\pm 2} \int d^3q e^{i\mathbf{q}\cdot\mathbf{x}} \beta(\mathbf{q}, \lambda) e_{ij}(\hat{q}, \lambda) \hat{p}_i \hat{p}_j \Delta_{\nu}^{(T)}(q, \hat{p} \cdot \hat{q}, t). \quad (2.3.125)$$

The quantity $\Delta_{\nu}^{(T)}(q, \hat{p} \cdot \hat{q}, t)$ satisfies a Boltzmann equation

$$\dot{\Delta}_{\nu}^{(T)}(q, \mu, t) + \frac{iq\mu}{a(t)} \Delta_{\nu}^{(T)}(q, \mu, t) = -2\dot{\mathcal{D}}_q(t), \quad (2.3.126)$$

that can again be solved by decomposing $\Delta_{\nu}^{(T)}(q, \mu, t)$ into multipole moments $\Delta_{\nu, \ell}^{(T)}(q, t)$ according to

$$\Delta_{\nu}^{(T)}(q, \mu, t) = \sum_{\ell} (-i)^{\ell} (2\ell + 1) P_{\ell}(\mu) \Delta_{\nu, \ell}^{(T)}(q, t). \quad (2.3.127)$$

The time evolution of the multipole moments $\Delta_{\nu, \ell}^{(T)}(q, t)$ is governed by the Boltzmann hierarchy

$$\dot{\Delta}_{\nu, \ell}^{(T)} + \frac{q}{a(2\ell + 1)} \left[(\ell + 1) \Delta_{\nu, \ell+1}^{(T)} - \ell \Delta_{\nu, \ell-1}^{(T)} \right] = -2\dot{\mathcal{D}}_q(t) \delta_{\ell 0}. \quad (2.3.128)$$

Together with the expression for the anisotropic stress in terms of the multipole coefficients

$$\pi_{\nu q}^T(t) = 2\bar{\rho}_{\nu}(t) \left[\frac{1}{15} \Delta_{\nu, 0}^{(T)}(q, t) + \frac{2}{21} \Delta_{\nu, 2}^{(T)}(q, t) + \frac{1}{35} \Delta_{\nu, 4}^{(T)}(q, t) \right], \quad (2.3.129)$$

the truncation of the Boltzmann hierarchy is again sufficient and no line of sight integration is needed to calculate the higher multipole moments.

In the case of the tensor mode, however, using the line of sight integral to avoid the Boltzmann hierarchy completely is also possible. The line of sight solution is simply given by

$$\Delta_{\nu}^{(T)}(q, \mu, t) = -2 \int_{t_1}^t dt' \exp\left(-iq\mu \int_{t'}^t \frac{dt''}{a(t'')}\right) \dot{\mathcal{D}}_q(t'). \quad (2.3.130)$$

This allows us to write the anisotropic stress due to neutrinos as [84]

$$\pi_{\nu q}^T(t) = -4 \int_{t_1}^t dt' K\left(-iq \int_{t'}^t \frac{dt''}{a(t'')}\right) \dot{\mathcal{D}}_q(t'), \quad (2.3.131)$$

where the function $K(v)$ is given by

$$K(v) \equiv j_2(v)/v^2 = -\frac{\sin v}{v^3} - \frac{3 \cos v}{v^4} + \frac{3 \sin v}{v^5}, \quad (2.3.132)$$

and the functions $j_{\ell}(v)$ are spherical Bessel functions.

2.3.4 The Multipole Coefficients

We now have all the ingredients that are necessary to give explicit expressions for the multipole coefficients in a given model, specifically the Λ CDM model. We will content ourselves with giving the expressions for the multipole coefficients and a summary of the system of equations that has to be solved to calculate them. A lot of progress can be made in solving these equations analytically and obtaining analytic expressions for the multipole coefficients, but we will not review this here. We refer the interested reader to [85], [38] and references therein.

It is convenient to use the definition of the multipole coefficients in the form

$$C_{XY,\ell} = \frac{1}{2\ell + 1} \sum_m \langle a_{X,\ell m} a_{Y,\ell m}^* \rangle, \quad (2.3.133)$$

where X and Y stand for T , E , and B . Using the definition of the expansion coefficients $a_{T,\ell m}$ and $a_{P,\ell m}$, equations (2.2.6), (2.2.12) as well as the expressions (2.1.7), (2.1.8) relating the temperature fluctuations and Stokes parameters to the dimensionless intensity matrix, one finds

$$a_{T,\ell m} = \frac{T_0}{4} \int d^2 \hat{n} Y_\ell^{m*}(\hat{n}) J_{ii}(0, -\hat{n}, t_0), \quad (2.3.134)$$

$$a_{P,\ell m} = \frac{T_0}{2} \int d^2 \hat{n} {}_2Y_\ell^{m*}(\hat{n}) e_{+i}(\hat{n}) e_{+j}(\hat{n}) J_{ij}(0, -\hat{n}, t_0). \quad (2.3.135)$$

Using the decomposition of the dimensionless intensity matrix (2.3.69), (2.3.92) together with the line of sight solutions (2.3.86), (2.3.87), (2.3.101), and (2.3.102), we will now work out expressions for the expansion coefficients and, by taking the average (2.3.133), the multipole coefficients.

2.3.4.1 Scalar modes

Using equation (2.3.69), one finds that the trace of the intensity matrix for the scalar modes is given by

$$J_{ii}(0, -\hat{n}, t_0) = \int d^3 q \alpha(\mathbf{q}) \Delta_T^{(S)}(q, -\hat{q} \cdot \hat{n}, t_0), \quad (2.3.136)$$

while

$$e_{+i}(\hat{n}) e_{+j}(\hat{n}) J_{ij}(0, -\hat{n}, t_0) = \int d^3 q \alpha(\mathbf{q}) \frac{(\mathbf{e}_+ \cdot \hat{q})^2}{1 - (\hat{q} \cdot \hat{n})^2} \Delta_P^{(S)}(q, -\hat{q} \cdot \hat{n}, t_0), \quad (2.3.137)$$

so that

$$a_{T,\ell m}^{(S)} = \frac{T_0}{4} \int d^3 q \alpha(\mathbf{q}) \int d^2 \hat{n} Y_\ell^{m*}(\hat{n}) \Delta_T^{(S)}(q, -\hat{q} \cdot \hat{n}, t_0), \quad (2.3.138)$$

$$a_{P,\ell m}^{(S)} = \frac{T_0}{2} \int d^3 q \alpha(\mathbf{q}) \int d^2 \hat{n} {}_2Y_\ell^{m*}(\hat{n}) \frac{(\mathbf{e}_+ \cdot \hat{q})^2}{1 - (\hat{q} \cdot \hat{n})^2} \Delta_P^{(S)}(q, -\hat{q} \cdot \hat{n}, t_0). \quad (2.3.139)$$

Though somewhat tedious, using the line of sight solutions (2.3.86), (2.3.87) for $\Delta_T^{(S)}(q, -\hat{q} \cdot \hat{n}, t_0)$ and $\Delta_P^{(S)}(q, -\hat{q} \cdot \hat{n}, t_0)$, the integrals over \hat{n} can be done analytically. One finds the following expressions for the coefficients $a_{T,\ell m}^{(S)}$ and $a_{P,\ell m}^{(S)}$:

$$a_{T,\ell m}^{(S)} = \pi T_0 i^\ell \int d^3q \alpha(\mathbf{q}) Y_\ell^{m*}(\hat{q}) \Delta_{T,\ell}^{(S)}(q, t_0), \quad (2.3.140)$$

$$a_{P,\ell m}^{(S)} = -\pi T_0 i^\ell \int d^3q \alpha(\mathbf{q}) Y_\ell^{m*}(\hat{q}) \Delta_{E,\ell}^{(S)}(q, t_0). \quad (2.3.141)$$

$$(2.3.142)$$

The quantities $\Delta_{T,\ell}^{(S)}(q, t_0)$ and $\Delta_{E,\ell}^{(S)}(q, t_0)$ are sometimes referred to as transfer functions and are given by

$$\begin{aligned} \Delta_{T,\ell}^{(S)}(q, t_0) &= \int_{t_1}^{t_0} dt P(t) \\ &\times \left\{ \left[3\Phi(q, t) - 2a(t) \frac{d}{dt} \left(a(t) \dot{B}_q(t) \right) + \frac{3}{4} \Pi(q, t) \right] j_\ell(qr(t)) \right. \\ &\quad \left. - 4q \left[\delta u_q(t)/a(t) + a(t) \dot{B}_q(t)/2 \right] j'_\ell(qr(t)) + \frac{3}{4} \Pi(q, t) j''_\ell(qr(t)) \right\} \\ &- \int_{t_1}^{t_0} dt \exp \left[- \int_t^{t_0} \omega_c(t') dt' \right] \\ &\quad \times \frac{d}{dt} \left[2A_q(t) + 2a(t) \frac{d}{dt} \left(a(t) \dot{B}_q(t) \right) \right] j_\ell(qr(t)), \end{aligned} \quad (2.3.143)$$

$$\Delta_{E,\ell}^{(S)}(q, t_0) = \frac{3}{4} \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \int_{t_1}^{t_0} dt \frac{P(t) \Pi(q, t)}{q^2 r^2(t)} j_\ell(qr(t)). \quad (2.3.144)$$

For vanishing Stokes parameter V , the dimensionless intensity matrix is real. Together with the definition of the sources functions (2.3.70), the product of stochastic parameters and $\Delta_{E,\ell}^{(S)}(q, t_0)$ then satisfies

$$\alpha(\mathbf{q})^* \Delta_{E,\ell}^{(S)}(q, t_0)^* = \alpha(-\mathbf{q}) \Delta_{E,\ell}^{(S)}(q, t_0). \quad (2.3.145)$$

Combined with the behavior of the spherical harmonics under space inversion $Y_\ell^m(\hat{q}) = (-1)^\ell Y_\ell^{-m*}(-\hat{q})$, this implies $a_{P,\ell m}^* = a_{P,\ell -m}$, or equivalently

$$a_{E,\ell m}^{(S)} = -a_{P,\ell m}^{(S)} \quad \text{and} \quad a_{B,\ell m} = 0. \quad (2.3.146)$$

Using equation (2.3.133), the non-vanishing contributions of the scalar modes to the multipole coefficients are then given by

$$C_{TT,\ell}^{(S)} = \pi^2 T_0^2 \int q^2 dq \left| \Delta_{T,\ell}^{(S)}(q, t_0) \right|^2, \quad (2.3.147)$$

$$C_{TE,\ell}^{(S)} = \pi^2 T_0^2 \int q^2 dq \Delta_{T,\ell}^{(S)}(q, t_0) \Delta_{E,\ell}^{(S)}(q, t_0), \quad (2.3.148)$$

$$C_{EE,\ell}^{(S)} = \pi^2 T_0^2 \int q^2 dq \left| \Delta_{E,\ell}^{(S)}(q, t_0) \right|^2. \quad (2.3.149)$$

At linear order, the scalar modes thus only contribute to the multipole coefficients, $C_{TT,\ell}$, $C_{TE,\ell}$, and $C_{EE,\ell}$, but not to $C_{BB,\ell}$, so that a B-mode signal would present evidence for vector or tensor modes. There is good evidence from the observed $C_{TE,\ell}$ that perturbations are generated early [86], so that the vector modes would have decayed by now, and a detection of a B-mode signal would be an indirect detection of gravitational waves. This is somewhat oversimplified as in the presence of an E-mode signal gravitational lensing will lead to a B-mode signal. This is an important effect, and at least for $\ell \gtrsim 100$ is expected to be the dominant contribution. Lensing is reasonably well understood, and it seems to be possible to extract a primordial B-mode signal as long as the tensor to scalar ratio, which will be introduced in subsections 2.4.2 and 2.4.3, is large enough, roughly $r \gtrsim 0.001 - 0.01$. We will return to this in subsection 2.4.3 and explain more carefully what could be learned from such an observation.

Once the transfer functions are known, we thus know how to obtain the multipole coefficients. What remains is to give the equations governing

the time dependence of the quantities that enter in the calculation of the transfer functions. The source functions, metric perturbations, and velocity potential are obtained by integrating the truncated Boltzmann hierarchies for photons, equations (2.3.76) and (2.3.76) and neutrinos, equations (2.3.120) along with two conveniently chosen linear combinations of equations (2.3.22), (2.3.23), (2.3.24) and (2.3.25) for the metric components. Where they appear, the fluctuations in the energy density, pressure, velocity potential, and the anisotropic stress for the photons are expressed in terms of the multipole coefficients using equations (2.3.79), (2.3.80), (2.3.81), and (2.3.82); those for neutrinos are expressed in terms of the corresponding multipole coefficients using equations (2.3.121), (2.3.122), (2.3.123), and (2.3.124). The fluctuations in energy density, pressure, velocity potential, and anisotropic inertia for the remaining constituents are determined from their equations of motion. To be specific, for the Λ CDM model, the missing equations are those governing the evolution of the baryons and the dark matter particles. In the particular synchronous gauge in which the velocity potential for the cold dark matter vanishes, energy conservation for the cold dark matter perturbations takes the form

$$\delta\dot{\rho}_{cq} + \frac{3\dot{a}}{a}\delta\rho_{cq} + \frac{1}{2}\bar{\rho}_{cq}\left(3\dot{A}_q - q^2\dot{B}_q\right) = 0, \quad (2.3.150)$$

while energy conservation for the baryons gives

$$\delta\dot{\rho}_{bq} + \frac{3\dot{a}}{a}\delta\rho_{bq} - \frac{q^2}{a^2}\bar{\rho}_b\delta u_{bq} + \frac{1}{2}\bar{\rho}_b\left(3\dot{A}_q - q^2\dot{B}_q\right) = 0. \quad (2.3.151)$$

While the exchange of energy between the baryons and the photons due to scattering is negligible and energy is conserved separately for baryons and photons, momentum is exchanged efficiently between electrons and photons.

The equation corresponding to momentum conservation in the plasma is¹⁷

$$\delta\dot{u}_{bq} + \frac{4\bar{\rho}_\gamma}{3\bar{\rho}_b}\omega_c(t) \left(\delta u_{bq} + \frac{3a}{4q}\Delta_{T,1}^{(S)}(q,t) \right) = 0, \quad (2.3.152)$$

where we have used equations (2.3.80), (2.3.82), and (2.3.81) to write the perturbations in the pressure and velocity potential for the photons as well as the anisotropic stress in terms of the multipole coefficients, equation (2.3.76) for $\ell = 1$ to eliminate $\dot{\Delta}_{T,1}^{(S)}(q,t)$. Finally, we can use equations (2.3.22) and (2.3.23) for the metric components. For the Λ CDM model, these are

$$\frac{q^2}{a^2}A_q + \frac{\dot{a}}{a} \left(3\dot{A}_q - q^2\dot{B}_q \right) = 8\pi G \left(\delta\rho_{qb} + \delta\rho_{qc} + \bar{\rho}_\gamma\Delta_{T,0}^{(S)} + \bar{\rho}_\nu\Delta_{\nu,0}^{(S)} \right), \quad (2.3.153)$$

and

$$\dot{A}_q = 8\pi G \left(\bar{\rho}_b\delta u_{bq} - \frac{a}{q}\bar{\rho}_\gamma\Delta_{T,1}^{(S)}(q,t) - \frac{a}{q}\bar{\rho}_\nu\Delta_{\nu,1}^{(S)}(q,t) \right). \quad (2.3.154)$$

These equations of motion are a closed set and can be solved numerically for any choice of initial conditions. What remains is to specify the initial conditions for adiabatic perturbations. They can be found by solving this system of equations far outside the horizon in the radiation dominated era. At early times, Thomson scattering is very efficient, and it can be seen from the Boltzmann hierarchy (2.3.76), (2.3.76) that it drives the temperature multipole moments $\Delta_{T,\ell}^{(S)}(q,t)$ for $\ell \geq 2$ as well as all polarization multipole moments $\Delta_{P,\ell}^{(S)}(q,t)$ to zero. In this limit, one can look for a solution of the remaining system of equations of the form

$$\Delta_{T,0}^{(S)} = \Delta_{\nu,0}^{(S)} = \frac{4}{3}\frac{\delta\rho_c}{\bar{\rho}_c} = \frac{4}{3}\frac{\delta\rho_b}{\bar{\rho}_b} \equiv \Delta_0^{(S)}, \quad (2.3.155)$$

¹⁷We have dropped the perturbation in the baryon pressure as in [38]. This term is kept in CMBfast and CAMB, but turns out to be negligible for wavelengths observable in the CMB. It is important at shorter scales and should be included for an accurate calculation of the matter power spectrum.

and

$$\Delta_{\nu,1}^{(S)} \propto \Delta_{T,1}^{(S)} = -\frac{4}{3} \frac{q}{a} \delta u_{bq} \equiv \Delta_1^{(S)}. \quad (2.3.156)$$

Choosing the normalization such that the quantity \mathcal{R}_q outside the horizon approaches the constant \mathcal{R}_q^o the initial conditions for the growing adiabatic mode to leading order in q/aH are given by

$$\Delta_0^{(S)}(q, t) = \frac{4}{3} \frac{q^2 t^2}{a^2(t)} \mathcal{R}_q^o, \quad (2.3.157)$$

$$\Delta_1^{(S)}(q, t) = \frac{8}{27} \frac{q^3 t^3}{a^3(t)} \mathcal{R}_q^o, \quad (2.3.158)$$

$$\Delta_{\nu,2}^{(S)}(q, t) = -\frac{16}{3(15 + 4f_\nu)} \frac{q^2 t^2}{a^2(t)} \mathcal{R}_q^o, \quad (2.3.159)$$

$$A_q(t) = \left(2 - \frac{2}{3} \frac{5 + 4f_\nu}{15 + 4f_\nu} \frac{q^2 t^2}{a^2(t)} \right) \mathcal{R}_q^o, \quad (2.3.160)$$

$$q^2 \dot{B}_q(t) = \frac{20}{15 + 4f_\nu} \frac{q^2 t}{a^2(t)} \mathcal{R}_q^o, \quad (2.3.161)$$

$$(2.3.162)$$

where f_ν is the fraction of the radiation energy density stored in neutrinos

$$f_\nu = \frac{\bar{\rho}_\nu}{(\bar{\rho}_\gamma + \bar{\rho}_\nu)} = \frac{N_\nu(7/8) (4/11)^{4/3}}{1 + N_\nu(7/8) (4/11)^{4/3}}, \quad (2.3.163)$$

with N_ν the number of light neutrino species, which we take to be $N_\nu = 3$ in agreement with particle physics, and $\Delta_{\nu,1}^{(S)}$ is related to $\Delta_1^{(S)}$ by

$$\Delta_{\nu,1}^{(S)}(q, t) = \frac{23 + 4f_\nu}{15 + 4f_\nu} \Delta_1^{(S)}(q, t). \quad (2.3.164)$$

The higher neutrino multipole moments are not being driven to zero, but they are higher order in q/aH and can be set to zero initially.

In practice, it may be advantageous not to start the integration too far in the radiation dominated era to keep the integration time as short as

possible. In that case the initial conditions should be corrected to account for the presence of dark matter and baryons. To do this, it is convenient to work with the variable $y = a(\Omega_b + \Omega_c)/\Omega_r$.

For a numerical treatment, rather than using equations (2.3.143) as they stand, it may be more convenient to eliminate the derivatives acting on the spherical Bessel functions by integration by parts so only the Bessel functions but not their derivatives have to be evaluated or stored. It may also be somewhat more convenient for numerical purposes to work with conformal time, $\tau = \int_{t_1}^t 1/a(t)$.

We have now reviewed how to calculate the contribution of the adiabatic mode to the multipole coefficients $C_{TT,\ell}^{(S)}$, $C_{TE,\ell}^{(S)}$, and $C_{EE,\ell}^{(S)}$ for an arbitrary choice of \mathcal{R}_q^o . To calculate the multipole coefficients and compare them to experiment, we have to make an assumption about the momentum dependence of \mathcal{R}_q^o . This is commonly done by specifying the quantity

$$\Delta_{\mathcal{R}}^2(q) = 4\pi |\mathcal{R}_q^o|^2 q^3. \quad (2.3.165)$$

In the Λ CDM model the perturbations are assumed to be nearly scale invariant, which translates into a momentum dependence of $\Delta_{\mathcal{R}}^2(q)$ of the form

$$\Delta_{\mathcal{R}}^2(q) = \Delta_{\mathcal{R}}^2(q_*) \left(\frac{q}{q_*}\right)^{n_s-1}, \quad (2.3.166)$$

where n_s is referred to as the scalar spectral index and $\Delta_{\mathcal{R}}^2(q_*)$ is the scalar amplitude at the pivot scale q_* . For $n_s = 1$ all decades in momentum contribute the same amount to the variance of $\mathcal{R}(t, \mathbf{x})$, so that this case is known as a scale invariant spectrum.

The theoretical predictions for the contributions of the scalar modes to the multipoles in the Λ CDM model with cosmological parameters as derived

from the five-year WMAP data alone (2.0.6)-(2.0.11) obtained by integrating the equations discussed in this section numerically are shown in Figure 2.3.

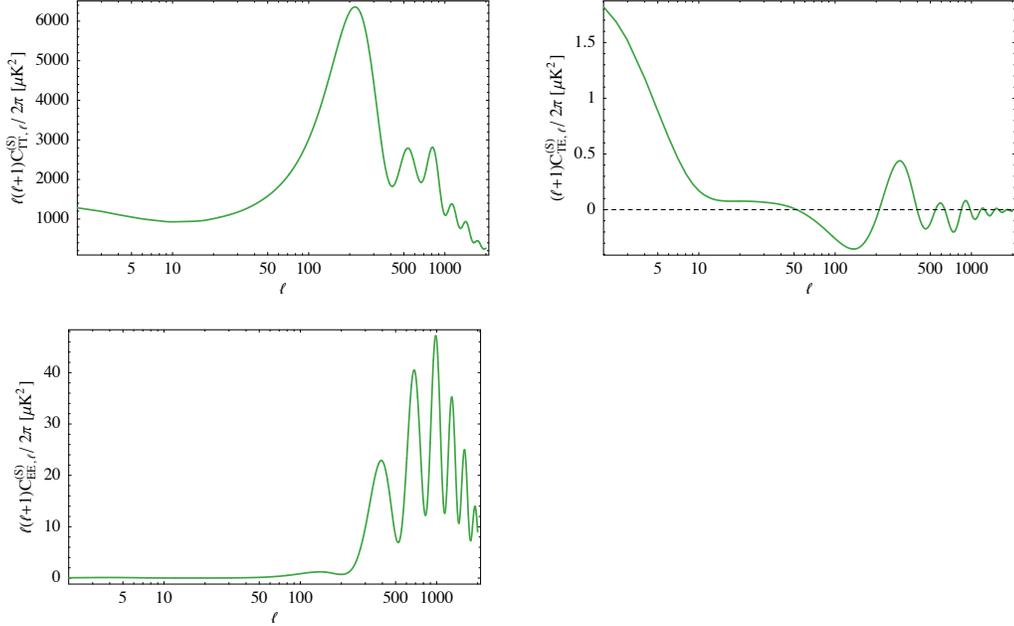


Figure 2.3: This plot shows the theoretical prediction for the scalar contributions multipole coefficients $C_{TT,\ell}^{(S)}$, $C_{TE,\ell}^{(S)}$ and $C_{EE,\ell}^{(S)}$ for the Λ CDM model with parameters as in equations (2.0.6)-(2.0.11).

The experimental results for the temperature multipole coefficients as well as the temperature polarization cross correlation from [86] are shown in Figure 2.4 along with the best-fit Λ CDM model.

What remains to be understood is what mechanism generated these adiabatic, Gaussian, perturbations with a nearly scale invariant spectrum in the first place. The leading paradigm is inflation and we will review the calculation of the power spectrum for this case in subsection 2.4.2.

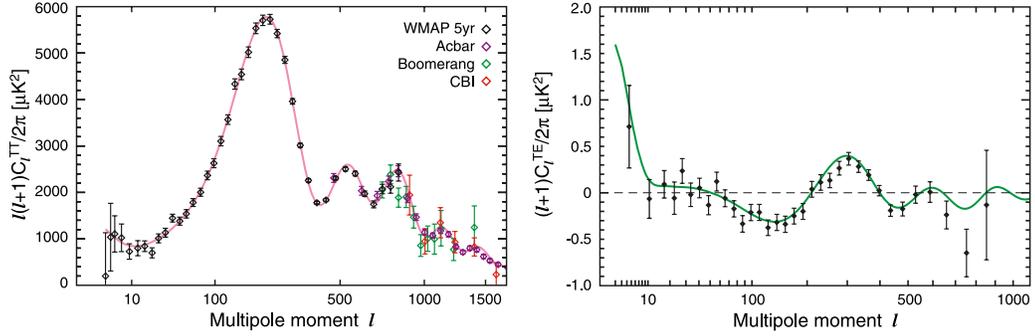


Figure 2.4: The left plot shows the five year WMAP results for the TT multipole coefficients along with results from the ACBAR [87], Boomerang [88], and CBI [89] experiments. The right plot shows the five year WMAP 5-year results for the TE multipole coefficients. The solid curves represent the best-fit Λ CDM model to the WMAP data.

2.3.4.2 Tensor modes

For the tensor modes, equation (2.3.92) implies that the trace of the dimensionless intensity matrix is of the form

$$J_{ii}(0, -\hat{n}, t_0) = \sum_{\lambda=\pm 2} \int d^3q \beta(\mathbf{q}, \lambda) \hat{n}_i \hat{n}_j e_{ij}(\hat{q}, \lambda) \tilde{\Delta}_T^{(T)}(q, -\hat{q} \cdot \hat{n}, t_0), \quad (2.3.167)$$

while

$$e_{+i}(\hat{n}) e_{+j}(\hat{n}) J_{ij}(0, -\hat{n}, t_0) = \sum_{\lambda=\pm 2} \int d^3q \beta(\mathbf{q}, \lambda) e_{+i}(\hat{n}) e_{+j}(\hat{n}) e_{ij}(\hat{q}, \lambda) \tilde{\Delta}_P^{(T)}(q, -\hat{q} \cdot \hat{n}, t_0), \quad (2.3.168)$$

leading to

$$a_{T,\ell m}^{(T)} = \frac{T_0}{4} \sum_{\lambda=\pm 2} \int d^3q \beta(\mathbf{q}, \lambda) \int d^2\hat{n} Y_\ell^{m*}(\hat{n}) \times \hat{n}_i \hat{n}_j e_{ij}(\hat{q}, \lambda) \tilde{\Delta}_T^{(T)}(q, -\hat{q} \cdot \hat{n}, t_0), \quad (2.3.169)$$

$$a_{P,\ell m}^{(T)} = \frac{T_0}{2} \sum_{\lambda=\pm 2} \int d^3q \beta(\mathbf{q}, \lambda) \int d^2\hat{n} {}_2Y_\ell^{m*}(\hat{n}) \\ \times e_{+i}(\hat{n})e_{+j}(\hat{n})e_{ij}(\hat{q}, \lambda)\tilde{\Delta}_P^{(T)}(q, -\hat{q} \cdot \hat{n}, t_0). \quad (2.3.170)$$

Using the line of sight solutions (2.3.101), (2.3.102) for $\Delta_T^{(T)}(q, -\hat{q} \cdot \hat{n}, t_0)$ and $\Delta_P^{(T)}(q, -\hat{q} \cdot \hat{n}, t_0)$, the integrals over \hat{n} can again be done analytically. The expressions for the coefficients $a_{T,\ell m}^{(T)}$ and $a_{P,\ell m}^{(T)}$ can be written in the form

$$a_{T,\ell m}^{(T)} = \pi T_0 i^\ell \frac{1}{\sqrt{2}} \sum_{\pm} \int d^3q \beta(\mathbf{q}, \pm 2) {}_{\mp 2}Y_\ell^{m*}(\hat{q}) \Delta_{T,\ell}^{(T)}(q, t_0), \quad (2.3.171)$$

$$a_{P,\ell m}^{(T)} = -\pi T_0 i^\ell \frac{1}{\sqrt{2}} \sum_{\pm} \int d^3q \beta(\mathbf{q}, \pm 2) {}_{\mp 2}Y_\ell^{m*}(\hat{q}) \\ \times \left(\Delta_{E,\ell}^{(T)}(q, t_0) \pm i \Delta_{B,\ell}^{(T)}(q, t_0) \right), \quad (2.3.172)$$

where $\Delta_{T,\ell}^{(T)}(q, t_0)$, $\Delta_{E,\ell}^{(T)}(q, t_0)$, and $\Delta_{B,\ell}^{(T)}(q, t_0)$ are given by

$$\Delta_{T,\ell}^{(T)}(q, t_0) = \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \int_{t_1}^{t_0} dt \exp \left[- \int_t^{t_0} \omega_c(t') dt' \right] \\ \times \left(2\dot{\mathcal{D}}_q(t) - \omega_c(t)\Psi(q, t) \right) \frac{j_\ell(qr(t))}{q^2 r^2(t)}, \quad (2.3.173)$$

$$\Delta_{E,\ell}^{(T)}(q, t_0) = - \int_{t_1}^{t_0} dt P(t)\Psi(q, t) \\ \times \left(12 + 8\rho \frac{\partial}{\partial \rho} - \rho^2 + \rho^2 \frac{\partial^2}{\partial \rho^2} \right) \frac{j_\ell(\rho)}{\rho^2} \Big|_{\rho=qr(t)}, \quad (2.3.174)$$

$$\Delta_{B,\ell}^{(T)}(q, t_0) = \int_{t_1}^{t_0} dt P(t)\Psi(q, t) \left(8\rho + 2\rho^2 \frac{\partial}{\partial \rho} \right) \frac{j_\ell(\rho)}{\rho^2} \Big|_{\rho=qr(t)}. \quad (2.3.175)$$

With the reality property for the stochastic parameter $\beta(\mathbf{q}, \lambda)^* = \beta(-\mathbf{q}, \lambda)$ and the property of the tensor spherical harmonics under space inversion ${}_\lambda Y_\ell^m(\hat{q}) =$

$(-1)^\ell Y_\ell^{-m*}(-\hat{q})$, one finds

$$a_{E,\ell m}^{(T)} = \pi T_0 i^\ell \frac{1}{\sqrt{2}} \sum_{\pm} \int d^3q \beta(\mathbf{q}, \pm 2)_{\mp 2} Y_\ell^{m*}(\hat{q}) \Delta_{E,\ell}^{(T)}(q, t_0), \quad (2.3.176)$$

$$a_{B,\ell m}^{(T)} = \pi T_0 i^\ell \frac{1}{\sqrt{2}} \sum_{\pm} \int d^3q \beta(\mathbf{q}, \pm 2)_{\mp 2} Y_\ell^{m*}(\hat{q}) \Delta_{B,\ell}^{(T)}(q, t_0) \lambda/2. \quad (2.3.177)$$

The contributions of the tensor modes to the multipole coefficients can then be calculated using (2.3.133)

$$C_{TT,\ell}^{(T)} = \pi^2 T_0^2 \int q^2 dq \left| \Delta_{T,\ell}^{(T)}(q, t_0) \right|^2, \quad (2.3.178)$$

$$C_{TE,\ell}^{(T)} = \pi^2 T_0^2 \int q^2 dq \Delta_{T,\ell}^{(T)}(q, t_0) \Delta_{E,\ell}^{(T)}(q, t_0), \quad (2.3.179)$$

$$C_{EE,\ell}^{(T)} = \pi^2 T_0^2 \int q^2 dq \left| \Delta_{E,\ell}^{(T)}(q, t_0) \right|^2, \quad (2.3.180)$$

$$C_{BB,\ell}^{(T)} = \pi^2 T_0^2 \int q^2 dq \left| \Delta_{B,\ell}^{(T)}(q, t_0) \right|^2. \quad (2.3.181)$$

Notice that the multipole coefficients $C_{EB,\ell}^{(T)}$ and $C_{TB,\ell}^{(T)}$ vanish because the contributions of positive and negative helicity gravitons are equal and opposite and cancel in the sum. In a parity violating theory, the power in positive and negative helicity modes may be different, so that a cancellation will no longer occur and these coefficients will generically be non-zero. There is no indication that the probability distribution obeyed by the stochastic parameters $\beta(\mathbf{q}, \lambda)$ violates parity at this point.

The source function and metric perturbation are obtained by integrating the truncated Boltzmann hierarchy for photons, equations (2.3.97) and (2.3.98), along with the equation of motion for the tensor mode (2.3.28). For the Λ CDM model, the anisotropic stress receives contributions from photons and neutrinos. For the anisotropic stress of the photons, equation (2.3.100) is used. To calculate the anisotropic stress due to neutrinos, one either uses

a truncation of the Boltzmann hierarchy for the neutrinos (2.3.128) together with the expression for the anisotropic stress (2.3.129), or formula (2.3.131). As briefly discussed earlier, one of the solutions of this system will be constant outside the horizon, and we will give the initial conditions for this mode for the case where the truncated Boltzmann hierarchy is used for both photons and neutrinos. They can be obtained by solving the system of equations deep in the radiation dominated era and far outside the horizon. At these early times, Thomson scattering is highly efficient and drives all temperature and polarization multipole moments to zero, implying that these can be set to zero during this period. Choosing the normalization such that \mathcal{D}_q approaches the constant \mathcal{D}_q^o in the limit of vanishing q/aH , the solution for the remaining system of equations then takes the form

$$\mathcal{D}_q(t) = \left(1 - \frac{10}{15 + 4f_\nu} \frac{q^2 t^2}{a^2(t)}\right) \mathcal{D}_q^o, \quad (2.3.182)$$

$$\Delta_{\nu,0}^{(T)}(q, t) = \frac{20}{15 + 4f_\nu} \frac{q^2 t^2}{a^2(t)} \mathcal{D}_q^o. \quad (2.3.183)$$

The remaining multipole moments for the neutrinos are higher order in q/aH , and can be set to zero initially.

It may again be more convenient for numerical calculations to eliminate the derivatives acting on the spherical Bessel functions in equations (2.3.173) by integration by parts and/or work with conformal time or $y = a\Omega_M/\Omega_R$ as independent variable.

We have now reviewed how to calculate the tensor contribution to the multipole coefficients for an arbitrary choice of \mathcal{D}_q^o . To calculate the multipole coefficients, we have to make an assumption about the momentum dependence of \mathcal{D}_q^o . This can be done by specifying the quantity

$$\Delta_{\mathcal{D}}^2(q) = 16\pi |\mathcal{D}_q^o|^2 q^3. \quad (2.3.184)$$

Since the scalar perturbations are observed to be nearly scale invariant, it is natural to assume that the tensor perturbations will be as well. The following parameterization of the primordial power spectrum for the tensor modes is commonly used

$$\Delta_{\mathcal{D}}^2(q) = \Delta_{\mathcal{D}}^2(q_*) \left(\frac{q}{q_*} \right)^{n_t}, \quad (2.3.185)$$

where n_t is referred to as the tensor spectral index and $\Delta_{\mathcal{D}}(q)$ is the tensor amplitude. In this case a scale invariant spectrum corresponds to $n_t = 0$.

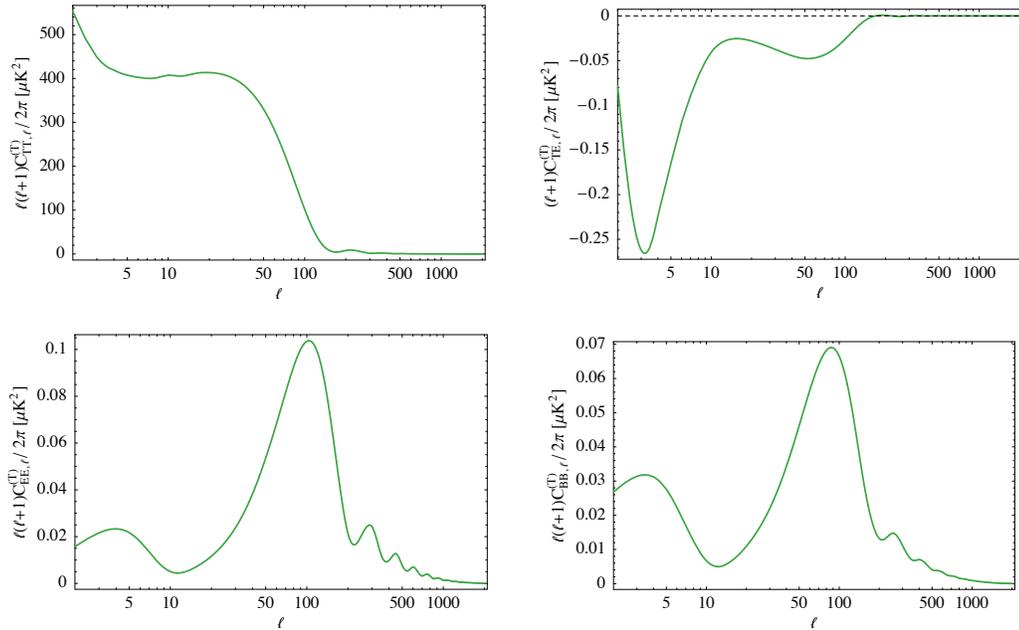


Figure 2.5: This plot shows the theoretical prediction for the tensor contribution to the multipole coefficients $C_{TT,\ell}^{(T)}$, $C_{TE,\ell}^{(T)}$, $C_{EE,\ell}^{(T)}$, and $C_{BB,\ell}^{(T)}$ for the Λ CDM model with parameters as in equations (2.0.6)-(2.0.11), as well as $n_t = 0$ and $r = 1$. Other values of r can be obtained by rescaling appropriately.

The theoretical predictions for the contributions of the tensor modes

to the multipole coefficients for a Λ CDM model with cosmological parameters given by equations (2.0.6)-(2.0.11) obtained by integrating the equations discussed in this section numerically are shown in Figure 2.5.

Inflation predicts a that primordial gravitational waves with a nearly scale invariant spectrum should have been produced and we will review the calculation of the power spectrum for this case in subsection 2.4.3. The detection of these primordial gravitational waves through the observation of a B-mode signal in the CMB is sometimes referred to as the smoking gun for inflation.

2.4 Inflation and the Primordial Power Spectrum

We have so far simply made the assumption that some small perturbations away from the FRW geometry are present, and worked out the predictions for the cosmic microwave background for adiabatic, Gaussian perturbations. It remains to be understood how these fluctuations were generated.

It was thought for some time that the perturbations we observe might have been seeded by defects such as cosmic strings or textures, but we now know experimentally that this cannot be the dominant mechanism. The most promising idea at present is that the fluctuations that lead to the structures we see around us arose as quantum fluctuations during a phase of exponential expansion of the universe called inflation.¹⁸ The spectrum of perturbations from inflation were first calculated by Mukhanov and Chibisov for Starobinsky's model [90], [91] and independently by Hawking [92], Guth and Pi [93],

¹⁸It is hard to put into words how violent this expansion must have been. The universe is thought to have expanded in size by a factor of 10^{26} in some 10^{-34} seconds.

Steinhardt, and Starobinsky [94], Bardeen and Turner [95], as well as Fischler, Ratra and Susskind [96]. As we will see, the predictions made by inflation are in remarkable agreement with experiment. This agreement certainly is the most convincing evidence that such a phase may indeed have occurred, but it should be noted that the existence of such a phase had not been postulated to fit the data but rather as a solution to three problems with the standard Big Bang theory. These problems are

- The monopole problem: There are some indications from the running of the gauge couplings as well as the observed particle content that the $SU(3)_c \times SU(2)_L \times U(1)_Y$ gauge group of the standard model of particle physics might be part of some larger gauge group. When this larger, grand unified gauge group gets broken to the standard model gauge group, one generically expects about one monopole per causal region at the time to be generated. No monopoles or other defects have been observed. This is known as the monopole problem.
- The flatness problem: The universe is now known to be spatially flat to within about one per cent. For the universe to be so close to being spatially flat today, it must have been spatially flat at early times to extremely high precision. This is known as the flatness problem.
- The horizon problem: The cosmic microwave radiation has been observed to have the same temperature in all directions to about one part in 10^5 . In a universe containing only matter and radiation, it is impossible for regions we see in different parts of the sky to have been in causal contact, and it is impossible for these different regions to have thermalized. This is known as the horizon problem.

It was first pointed out by Guth [97] that inflating the universe by about a factor $10^{30} = e^{69}$, or 69 e -folds, can solve the horizon as well as the flatness problem. As already pointed out by Guth in his original paper, the model suffered from a serious flaw called the graceful exit problem. The universe was supposed to inflate while a scalar field was stuck in a false vacuum. To end inflation, the scalar field had to tunnel to the true vacuum. For any period of inflation long enough to solve these problems, this would lead to the formation of bubbles that would not coalesce, leaving the universe in an extremely inhomogeneous state. This problem was solved independently by Linde [98] as well as Albrecht and Steinhardt [99] and the model had become viable. It should also be mentioned that an inflating universe without a graceful exit problem had already been studied earlier by Starobinsky [100]. His model was technically somewhat more involved and he had a very different motivation from Guth's that did not manage to get people's attention at the time.

After some historic remarks, some cautionary ones may be in order before discussing the physics of inflation. It may be that there simply is no grand unified gauge group that needs to be broken down to the standard model gauge group and neither need other gauge groups exist that get broken and leave relics behind. Even if such a grand unified group exists, the universe may never have been hot enough to undergo the phase transition that would leave monopoles behind. While we do have good evidence that the universe was once very dense and hot, this evidence comes from the existence of the cosmic microwave background as well as the success of nucleosynthesis. Both of them probe energy scales many orders of magnitudes lower than the energy scales associated with grand unified theories.

It is often stated that a flat FRW universe is unstable. As the absence

of instabilities in our review of perturbations around a flat FRW universe demonstrates, this statement is incorrect, and an FRW universe that is flat will remain so at all times. It might thus be that the universe is observed to be nearly flat today simply because it is flat rather than because some mechanism drove a spatially curved universe to a solution that is flat to a very good approximation, but one might then, of course, ask what was responsible for the universe to be exactly flat in the first place.

Finally, it should be mentioned that it is not completely understood whether it is correct to assert that inflation solves the horizon problem. Several studies seem to indicate that the universe is required to be homogeneous over a few times the horizon size to successfully enter an extended period of inflation [101], [102], [103], [104], [105], [106], [107], [108]. Even though this is several orders of magnitude better than the requirement that the universe be homogeneous over 10^5 causal regions as would be needed in the absence of inflation, acausal physics would still be required if these conclusions hold up, and a form of the horizon problem would persist. There is, however, also a numerical study that comes to the conclusion that it may be possible to have successful inflation despite significant inhomogeneities [109]. It seems more detailed numerical studies are in order to settle this. To end this paragraph on a positive note, one might mention, that inflation would explain why the universe we see is so large, and so close to being homogeneous and isotropic as well as several other puzzles [110].

Let us now give a very brief review of the physics of inflation starting from the action we will be working with for the rest of the section. It is that of a single scalar field minimally coupled to Einstein gravity

$$S = \frac{1}{16\pi G} \int d^4x \sqrt{-g} R - \int d^4x \sqrt{-g} \left(\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi + V(\phi) \right). \quad (2.4.1)$$

Since we will need it in the next to subsections, let us mention that the stress energy tensor for the scalar field that follows from this action is of the ideal fluid form, *i.e.* $T_{\mu\nu} = p g_{\mu\nu} + (p + \rho)u_\mu u_\nu$, with energy density ρ , pressure p , and u_μ given by

$$\rho = -\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + V(\phi), \quad (2.4.2)$$

$$p = -\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi), \quad (2.4.3)$$

$$u_\mu = -\frac{\partial_\mu\phi}{\sqrt{-g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi}}. \quad (2.4.4)$$

At the two-derivative level the action (2.4.1) is the most general action for a single scalar field minimally coupled to Einstein gravity, and is the simplest system that gives rise to an inflationary cosmology. Many modifications and extensions are possible, such as allowing for a non-minimal coupling, considering modifications of gravity, non-canonical kinetic terms, or allowing multiple scalar fields to participate in the inflationary dynamics. Since data is currently consistent with the most naive single field model, we will focus on this simplest case and refer the interested reader to the reviews [111], [110], as well as [38] for more sophisticated possibilities as well as a more detailed treatment of the single field case.

Conditions under which we know that the universe will enter an inflationary phase are that the scalar field can be treated as some time dependent background $\bar{\phi}(t)$ plus small spacetime dependent perturbations $\delta\phi(t, \mathbf{x})$, *i.e.*

$$\phi(t, \mathbf{x}) = \bar{\phi}(t) + \delta\phi(t, \mathbf{x}), \quad (2.4.5)$$

that the geometry be close to an FRW geometry, and that the background scalar field take a value so that its energy density is dominated by the potential energy. In this case, as long as the potential energy in the scalar field

is positive, it will have an effect similar to a cosmological constant and lead to a nearly exponential expansion. We will briefly summarize the background dynamics in subsection 2.4.1 and then turn to the calculation of the spectrum of scalar and tensor fluctuations generated in an inflationary universe in subsections 2.4.2 and 2.4.3.

2.4.1 The Background Evolution

During inflation, the flat FRW background is an attractor, and the spatial curvature rapidly becomes negligible. We will thus work with a flat FRW metric with corresponding line element

$$ds^2 = -dt^2 + a^2(t)d\mathbf{x}^2 \quad (2.4.6)$$

The Friedmann equation and the equation of motion for the scalar field for the background fields then take the form

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3} \left(\frac{1}{2}\dot{\bar{\phi}}^2 + V(\bar{\phi}) \right), \quad (2.4.7)$$

and

$$\ddot{\bar{\phi}} + 3\frac{\dot{a}}{a}\dot{\bar{\phi}} + V'(\bar{\phi}) = 0. \quad (2.4.8)$$

For the universe to undergo a period of nearly exponential expansion, the Hubble expansion rate

$$H = \frac{\dot{a}}{a}, \quad (2.4.9)$$

should be nearly constant. This will be the case provided that the fractional rate of change of the expansion rate is small compared to the rate of expansion of the universe, *i.e.* as long as

$$\epsilon \equiv -\frac{\dot{H}}{H^2} \ll 1. \quad (2.4.10)$$

Using the relation

$$\dot{H} = -4\pi G \dot{\phi}^2, \quad (2.4.11)$$

which can be derived by taking the derivative of the Hubble rate (2.4.7) and using the equation of motion for the scalar field, equation (2.4.8), it is easy to see that this will be the case provided

$$\dot{\phi}^2 \ll V(\bar{\phi}). \quad (2.4.12)$$

The kinetic energy of the field should thus be negligible compared to the potential energy, or put differently, the scalar field should be rolling slowly. To what extent this is the case is measured by the quantity ϵ , explaining why it is usually referred to as slow-roll parameter. Since an extended period of inflation is required to address the horizon problem, we have to require that the kinetic energy remain small for an extended period of time. To ensure this, the fractional rate of change of the time derivative of the field should be small compared to the rate of expansion of the universe, *i.e.* the absolute value of the quantity¹⁹

$$\delta \equiv \frac{\ddot{\phi}}{H\dot{\phi}}, \quad (2.4.13)$$

should satisfy $|\delta| \ll 1$. This implies that the first term in the equation of motion of the field, equation (2.4.8), is negligible compared to the second one. If it is also negligible compared to the last term, it can be dropped from the equation, and the dynamics of the scalar field is approximately given by

$$\dot{\phi} \approx -\frac{V'(\bar{\phi})}{3H}. \quad (2.4.14)$$

¹⁹We will come across this quantity again written as $\delta = \frac{\ddot{H}}{2\dot{H}H}$. Using equation (2.4.11), it is easy to see that they are identical. Another commonly used quantity is $\eta = \frac{\dot{\epsilon}}{\epsilon H}$.

This approximation is commonly referred to as the slow-roll approximation. Combining equations (2.4.11), (2.4.9), (2.4.7), and (2.4.14), the slow-roll parameter ϵ in this approximation can be written as

$$\epsilon \approx \frac{1}{16\pi G} \left(\frac{V'(\bar{\phi})}{V(\bar{\phi})} \right)^2, \quad (2.4.15)$$

while δ becomes

$$\delta \approx \frac{1}{16\pi G} \left(\frac{V'(\bar{\phi})}{V(\bar{\phi})} \right)^2 - \frac{1}{8\pi G} \frac{V''(\bar{\phi})}{V(\bar{\phi})}. \quad (2.4.16)$$

The conditions necessary for slow roll can thus be thought of as flatness conditions on the potential requiring that the quantities

$$\epsilon_V \equiv \frac{1}{16\pi G} \left(\frac{V'(\bar{\phi})}{V(\bar{\phi})} \right)^2, \quad (2.4.17)$$

and

$$\eta_V = \frac{1}{8\pi G} \frac{V''(\bar{\phi})}{V(\bar{\phi})}, \quad (2.4.18)$$

should satisfy $\epsilon_V \ll 1$ and $|\eta_V| \ll 1$. As long as these conditions hold and the potential is positive, the universe will undergo a rapidly accelerated expansion. The amount of expansion can easily be related to the motion of the field using equations (2.4.7) and (2.4.14). One finds

$$\frac{a(t_e)}{a(t_i)} = \exp \left[- \int_{\phi_i}^{\phi_e} d\phi \frac{8\pi G V(\bar{\phi})}{V'(\bar{\phi})} \right]. \quad (2.4.19)$$

The potential decreases as the field moves from its initial value ϕ_i to its final value ϕ_e , implying that the integrand is positive and the universe grows by some number N_e of e-foldings given by

$$N_e \equiv \ln \frac{a(t_e)}{a(t_i)} = - \int_{\phi_i}^{\phi_e} d\phi \frac{8\pi G V(\bar{\phi})}{V'(\bar{\phi})} \gg \sqrt{4\pi G} |\phi_e - \phi_i|, \quad (2.4.20)$$

where the inequality follows from the condition $\epsilon_V \ll 1$. The universe will thus expand by a large number of e-foldings as long as the field travels over a distance of $1/\sqrt{4\pi G} = \sqrt{2}M_P$ or larger. One should stress that this does not mean that the classical description of gravity will break down as the energy density in the field may well be far below the Planck scale throughout the field's motion. It does, however, imply that the potential needs to be flat over large distances in field space and that the inflationary dynamics may provide a probe of physics at very high energy scales.

At some point this period of accelerated expansion of the universe has to come to an end and the energy stored in the scalar field must be transferred to the standard model fields. This is referred to as reheating. Even though this is of crucial importance for a successful model of inflation, we will skip a discussion of reheating and now turn to the calculation of the spectrum of fluctuations during this period. The reader is encouraged to consult the review [112], as well as [38] and references therein for a discussion of reheating and further details and examples for the background evolution of single field inflation.

2.4.2 Primordial spectrum for scalar modes

After this brief review of the evolution of the background geometry in single field inflation, we will now turn to the quantization of the perturbations in the scalar field around such a background with the goal to extract the quantity \mathcal{R}_q^o that enters the calculation of the multipole coefficients as explained in the last section. We will be working in the Heisenberg picture throughout, meaning that the states will be time independent and the operators carry all the time dependence. For the quantization of the scalar field, it will be

convenient to work in Newtonian gauge, *i.e.* perform an infinitesimal diffeomorphism to set the metric components B and F to zero. The perturbations of the stress energy tensor in this gauge are given by

$$\delta\rho = \dot{\bar{\phi}}\delta\dot{\phi} + V'(\bar{\phi})\delta\phi - \frac{1}{2}E\dot{\bar{\phi}}^2, \quad (2.4.21)$$

$$\delta p = \dot{\bar{\phi}}\delta\dot{\phi} - V'(\bar{\phi})\delta\phi - \frac{1}{2}E\dot{\bar{\phi}}^2, \quad (2.4.22)$$

$$\delta u = -\frac{\delta\phi}{\sqrt{\frac{\dot{\bar{\phi}}^2}{\phi}}}, \quad (2.4.23)$$

while π^S , π_i^V , and π_{ij}^T all vanish. As can be seen from equation (2.3.21), the equation following from the part of the Einstein equations proportional to $\partial_i\partial_j$ then implies $E = -A \equiv 2\Psi$. It will be convenient to use the following linear combinations of the constraints

$$\dot{\Psi} + H\Psi = 4\pi G\dot{\bar{\phi}}\delta\phi, \quad (2.4.24)$$

and

$$\dot{H}\Psi - \frac{\nabla^2}{a^2}\Psi = 4\pi G\left(\ddot{\bar{\phi}}\delta\phi - \dot{\bar{\phi}}\delta\dot{\phi}\right), \quad (2.4.25)$$

and use the equation governing the time evolution of the scalar field perturbation

$$\delta\ddot{\phi} + 3H\delta\dot{\phi} - \frac{\nabla^2}{a^2}\delta\phi + V''(\bar{\phi})\delta\phi = 4\dot{\Psi}\dot{\bar{\phi}} - 2\Psi V'(\bar{\phi}). \quad (2.4.26)$$

Translational invariance implies that it will be useful to look for a solution of the form

$$\delta\phi(t, \mathbf{x}) = \int d^3q [\delta\phi_q(t)e^{i\mathbf{q}\cdot\mathbf{x}}\alpha(\mathbf{q}) + h.c.], \quad (2.4.27)$$

$$\Psi(t, \mathbf{x}) = \int d^3q [\delta\Psi_q(t)e^{i\mathbf{q}\cdot\mathbf{x}}\alpha(\mathbf{q}) + h.c.], \quad (2.4.28)$$

where \mathbf{q} is again the comoving momentum of the mode. The functions $\delta\phi_q(t)$ and $\Psi_q(t)$ then satisfy the above equations with ∇^2 replaced by $-q^2$.

During inflation, the scale factor a grows very rapidly while the Hubble rate H is nearly constant. This implies that for any value of comoving momentum the terms proportional to q^2/a^2 will dominate as we go back far enough in time because the scale factor becomes small. At these early times the modes oscillated so rapidly that they ceased to feel the effects of the expanding universe and behaved essentially like in Minkowski space. The metric perturbation Ψ becomes negligible in this limit and so does the friction term. One can then look for a WKB solution to this system of equation. Normalizing the solution such that $\alpha(\mathbf{q})$ and $\alpha^*(\mathbf{q})$ satisfy canonical commutation relations, the positive frequency solution at early times takes the form

$$\delta\phi_q(t) = \frac{1}{(2\pi)^{3/2}a(t)\sqrt{2q}} \exp \left[-iq \int_{t_*}^t \frac{dt'}{a(t')} \right], \quad (2.4.29)$$

$$\Psi_q(t) = \frac{4\pi i G \dot{\phi}}{(2\pi)^{3/2}q\sqrt{2q}} \exp \left[-iq \int_{t_*}^t \frac{dt'}{a(t')} \right]. \quad (2.4.30)$$

The subsequent evolution of the field and the metric perturbation can then be obtained by integrating the above system of equations with these initial conditions. What remains is to choose the state the system should be in. The standard assumption is that the universe at these early times is in what one might call the in-vacuum, in this context usually called the Bunch-Davies vacuum, defined by the requirement that it be annihilated by $\alpha(\mathbf{q})$, *i.e.*

$$\alpha(\mathbf{q}) |0\rangle = 0. \quad (2.4.31)$$

No completely convincing argument for why this should be the correct state has been given so far, and other possibilities have been considered such as so

called truncated α -vacua or thermal states. The study of modifications of this state is sometimes referred to collectively as trans-Planckian physics. We will see a well motivated modification in Chapter 5; a spectrum derived using the Bunch-Davies vacuum in a background with a modulated potential turns out to be indistinguishable from the spectrum for an excited state in a universe without these modulations.

The natural observables in this system are correlation functions of the fluctuations at different points in spacetime evaluated in this state of the form

$$\langle 0 | \delta\phi(t, \mathbf{x}) \delta\phi(t, \mathbf{y}) | 0 \rangle , \quad (2.4.32)$$

$$\langle 0 | \Psi(t, \mathbf{x}) \delta\phi(t, \mathbf{y}) | 0 \rangle , \quad (2.4.33)$$

and so on. We are currently working at the linear level so that the theory is Gaussian and all higher n -point functions either vanish for n odd, or are given in terms of sums of products of these two-point functions if n is even. To make contact with our previous discussions where averages corresponded to ensemble averages one important ingredient is missing. Decoherence must set in and turn the quantum state into one particular classical state of an ensemble whose ensemble averages are the same as these quantum averages. How this occurs is not understood in detail.

Let us now proceed on our way towards a calculation of the quantity $\mathcal{R}(t, \mathbf{x}) = -\Psi(t, \mathbf{x}) - H\delta\phi(t, \mathbf{x})/\dot{\phi}$. In analogy with the expansions (2.4.27), (2.4.28) we will also expand $\mathcal{R}(t, \mathbf{x})$ as

$$\mathcal{R}(t, \mathbf{x}) = \int d^3q [\mathcal{R}_q(t) e^{i\mathbf{q}\cdot\mathbf{x}} \alpha(\mathbf{q}) + h.c.] . \quad (2.4.34)$$

We are interested in the value $\mathcal{R}_q(t)$ approaches outside the horizon, *i.e.* when the terms proportional to q^2/a^2 become negligible. One could evolve the equations for Ψ and $\delta\phi$ and evaluate \mathcal{R} in this limit, but it turns out to be more

convenient to derive an equation for \mathcal{R} directly. This equation is known as the Mukhanov-Sasaki equation [113], [114]. To derive it, it is convenient to choose a gauge such that $\delta\phi = 0$ and $B = 0$. Using

$$H^2 = \frac{8\pi G}{3} \left(\frac{1}{2} \dot{\bar{\phi}}^2 + V(\bar{\phi}) \right), \quad (2.4.35)$$

$$\dot{H} = -4\pi G \dot{\bar{\phi}}^2, \quad (2.4.36)$$

the perturbations in the stress energy tensor in this gauge can be brought into the form

$$\delta T_{00} = \frac{1}{8\pi G} E \left(\dot{H} + 3H^2 \right), \quad (2.4.37)$$

$$\delta T_{0i} = \frac{1}{8\pi G} a \partial_i F \left(-2\dot{H} - 3H^2 \right), \quad (2.4.38)$$

$$\delta T_{ij} = \frac{1}{8\pi G} a^2 \delta_{ij} \left(-4\mathcal{R}\dot{H} - 6\mathcal{R}H^2 + E\dot{H} \right). \quad (2.4.39)$$

It is convenient to use the Hamiltonian and momentum constraint as well as the equation following from the part of the ij equations proportional to $\partial_i \partial_j$. These take the form

$$\dot{H}E + 3H^2E - 6H\dot{\mathcal{R}} + 2\frac{\nabla^2}{a^2}\mathcal{R} + 2H\frac{\nabla^2}{a^2}aF = 0, \quad (2.4.40)$$

$$HE - 2\dot{\mathcal{R}} = 0, \quad (2.4.41)$$

$$\frac{1}{2}E + \mathcal{R} + 2aHF + a\dot{F} = 0. \quad (2.4.42)$$

The last two equations can be used to eliminate E and F from the first one, which then leads to the equation governing the time evolution of the Fourier components of \mathcal{R} . It can be brought into the form

$$\frac{d^2\mathcal{R}_q}{d\tau^2} + \frac{2}{z} \frac{dz}{d\tau} \frac{d\mathcal{R}_q}{d\tau} + q^2\mathcal{R}_q = 0, \quad (2.4.43)$$

where $\tau = \int_{t_*}^t dt' 1/a(t')$ is the conformal time and

$$z = \frac{1}{H} \frac{d\bar{\phi}}{d\tau}. \quad (2.4.44)$$

Alternatively, it can be written as

$$\frac{d^2\mathcal{R}_q}{d\tau^2} + 2aH(1 + \delta + \epsilon) \frac{d\mathcal{R}_q}{d\tau} + q^2\mathcal{R}_q = 0, \quad (2.4.45)$$

where $\epsilon = -\dot{H}/H^2$ is the slow roll parameter defined in the last subsection and we will now use δ in the form

$$\delta = \frac{\ddot{H}}{2\dot{H}H}. \quad (2.4.46)$$

It is immediately clear from the Mukhanov-Sasaki equation (2.4.45) that there will be a solution that approaches a constant far outside the horizon, *i.e.* when term proportional to q^2 can be neglected.

What remains is to give the initial conditions for $\mathcal{R}_q(\tau)$. These are determined by equations (2.4.29) and (2.4.30) together with

$$\mathcal{R}_q = -\Psi_q - H\delta\phi_q/\dot{\phi}, \quad (2.4.47)$$

and are given by

$$\mathcal{R}_q(\tau) = -\frac{1}{(2\pi)^{3/2}\sqrt{2q}z(\tau)} e^{-iq\tau}. \quad (2.4.48)$$

So far this has been exact, and one could simply integrate this numerically in a straightforward way.²⁰ This would however limit us to a case by case study. To understand the generic prediction for the momentum dependence of \mathcal{R}_q made by inflation, it is helpful to make further progress analytically. One possibility is to consider specific potentials such as the exponential potential for which the equation can be solved exactly. We will not pursue this and refer the interested reader to [38]. Another option is to solve the equation in the

²⁰To this end it is convenient to split the equation up into two, one for the real part of \mathcal{R}_q and one for the imaginary part.

slow-roll approximation, *i.e.* for $\epsilon \ll 1$, $\delta \ll 1$ as well as $\dot{\delta} \ll 1$.²¹ We will turn to this now, and we will find that in this approximation single field inflation with a canonical kinetic term minimally coupled to gravity predicts just the nearly scale-invariant spectrum needed to explain experiments.

To obtain the Mukhanov-Sasaki equation in the slow-roll approximation note that

$$\frac{daH}{d\tau} = a^2 H^2 (1 - \epsilon), \quad (2.4.49)$$

To leading order in the slow-roll parameter ϵ one thus has $aH = -(1 + \epsilon)/\tau$.

Together with

$$\frac{1}{z} \frac{dz}{d\tau} = aH (1 + \delta + \epsilon) \approx -\frac{(1 + \delta + 2\epsilon)}{\tau}, \quad (2.4.50)$$

the equation (2.4.45) in the slow-roll approximation can then be brought into the form

$$\frac{d^2 \mathcal{R}_q}{d\tau^2} + \frac{2(1 + \delta + 2\epsilon)}{\tau} \frac{d\mathcal{R}_q}{d\tau} + q^2 \mathcal{R}_q = 0. \quad (2.4.51)$$

It can easily be checked that the solution to this equation with initial conditions satisfying (2.4.45) to leading order in slow roll parameters takes the form

$$\mathcal{R}_q(\tau) = -\frac{(-q\tau)^{-\nu+1/2}}{4\pi\sqrt{2qz(\tau)}} e^{i\pi\nu/2+i\pi/4} (-q\tau)^\nu H_\nu^{(1)}(-q\tau), \quad (2.4.52)$$

with $\nu = \frac{3}{2} + 2\epsilon + \delta$, and $H_\nu^{(1)}$ denotes the Hankel function of the first kind. The solution is written in this somewhat peculiar way because $z(\tau) \propto \tau^{-\nu+1/2}$, as can be seen from equation (2.4.50), so that the fraction is in fact a constant. Using the limiting behavior of the Hankel functions for $x \ll \sqrt{\nu+1}$

$$H_\nu^{(1)}(x) \rightarrow -i\Gamma(\nu)2^\nu x^{-\nu}/\pi, \quad (2.4.53)$$

²¹The time derivative of ϵ is automatically higher order in slow roll parameters. A case in which the time derivative of δ plays an interesting role will appear in Chapter 5.

we find that \mathcal{R}_q^o in the slow-roll approximation takes the form

$$\mathcal{R}_q^o = i \frac{(-\tau)^{-\nu+1/2} \Gamma(\nu)}{2\sqrt{\pi}(2\pi)^{3/2} z(\tau)} e^{i\pi\nu/2+i\pi/4} 2^\nu \frac{1}{q^\nu}. \quad (2.4.54)$$

implying in particular $\mathcal{R}_q^o \propto q^{-\nu}$. In terms of the somewhat more conventional quantity $\Delta_{\mathcal{R}}^2(q) = 4\pi |\mathcal{R}_q^o|^2 q^3$ this implies²²

$$\Delta_{\mathcal{R}}^2(q) = \Delta_{\mathcal{R}}^2(q_*) \left(\frac{q}{q_*} \right)^{-4\epsilon-2\delta}. \quad (2.4.55)$$

Comparison with equation (2.3.166) then shows that single field slow-roll inflation predicts a nearly scale invariant spectrum with scalar spectral index

$$n_s = 1 - 4\epsilon - 2\delta. \quad (2.4.56)$$

At times it will also be useful to have an expression for the scalar spectral index in terms of the slow roll parameters ϵ_V and η_V rather than ϵ and δ . In the slow-roll approximation, $\delta = \epsilon_V - \eta_V$, and one finds

$$n_s = 1 - 6\epsilon_V + 2\eta_V. \quad (2.4.57)$$

What remains is to give a nice form for the amplitude of the fluctuations. To do this one recalls that \mathcal{R}_q^o is a constant, and can thus be evaluated at any time that is convenient. It will be convenient to evaluate it at the time at which a given mode exits the horizon defined by

$$\frac{q}{a(t_q)} = H(t_q). \quad (2.4.58)$$

²²Notice that the factors are chosen to give the variance of \mathcal{R} a simple form. One has $\langle \mathcal{R}(t_*, 0)^2 \rangle = \int d \ln q \Delta_{\mathcal{R}}^2(q)$, where t_* is chosen so that the modes of interest are outside the horizon and $\mathcal{R}_q(t_*) = \mathcal{R}_q^o$.

Using $z(\tau) = \pm a(\tau)\sqrt{\epsilon(\tau)}/\sqrt{4\pi G}$, and setting $\nu = 3/2$ in the argument of the Γ -function, the exponent of 2, and the phase, one finds

$$\mathcal{R}_q^o = \mp i \frac{\sqrt{8\pi G} H(t_q)}{2(2\pi)^{3/2} \sqrt{\epsilon(t_q)}} \frac{1}{q^{3/2}} \quad (2.4.59)$$

so that

$$\Delta_{\mathcal{R}}^2(q) = \frac{1}{8\pi^2} \frac{8\pi G H^2(t_q)}{\epsilon(t_q)}, \quad (2.4.60)$$

or equivalently

$$\Delta_{\mathcal{R}}^2(q) = \Delta_{\mathcal{R}}^2(q_*) \left(\frac{q}{q_*}\right)^{n_s-1}, \quad (2.4.61)$$

with

$$\Delta_{\mathcal{R}}^2(q_*) = \frac{1}{8\pi^2} \frac{8\pi G H^2(t_{q_*})}{\epsilon(t_{q_*})} \quad \text{and} \quad n_s = 1 - 4\epsilon(t_{q_*}) - 2\delta(t_{q_*}). \quad (2.4.62)$$

We have now calculated the spectrum of primordial scalar perturbations in slow roll inflation. As we have seen, the modes become constant once outside the horizon so that the power spectrum stays the same during the remaining period of inflation. However, inflation ends at some point, and physics that we do not understand, such as reheating, the electroweak phase transition, the QCD phase transition, *etc.*, occurs between the period when the modes we observe in the CMB exit the horizon, and the period we begin our integration of the Boltzmann hierarchy. One might thus worry that the spectrum we calculated will get modified during this intervening period. However, it can be shown that whatever the contents of the universe are, this adiabatic mode is conserved outside the horizon [64]. This remarkable property thus gives us a chance to learn about the universe at a very early time through the study of the cosmic microwave background. The adiabatic mode still exists for multifield

inflation, but in that case it will generically not be the only one that is excited so that the quantity \mathcal{R}_q will change outside the horizon and we cannot hope to learn much about the details of inflation in that case. How much can still be learned has not been conclusively answered and requires more careful study. The results for single field inflation to leading order in slow roll parameters were first given by [115]. For an extension of these results to higher orders in the slow-roll parameters we refer the reader to [116].

2.4.3 Primordial spectrum for tensor modes

We can now turn to the calculation of the spectrum of gravitational waves generated during inflation. To this end, we will have to quantize the transverse, traceless perturbation of the metric in the same way we quantized the scalar field in the last subsection. After using the equations of motion for the background, the action for these modes takes the form

$$S = \frac{1}{16\pi G} \int dt d^3x a^3(t) \left[\frac{1}{4} \dot{D}_{ij} \dot{D}_{ij} + \frac{1}{4} D_{ij} \nabla^2 D_{ij} \right], \quad (2.4.63)$$

where repeated indices are summed over. The equation governing the time evolution of the tensor perturbations is then simply

$$\ddot{D}_{ij} + 3\frac{\dot{a}}{a}\dot{D}_{ij} - \frac{\nabla^2}{a^2}D_{ij} = 0. \quad (2.4.64)$$

It will again be helpful to look for a solution of the form

$$D_{ij}(t, \mathbf{x}) = \sum_{\lambda} \int d^3q [\beta(\mathbf{q}, \lambda) e_{ij}(\hat{q}, \lambda) \mathcal{D}_q(t) e^{i\mathbf{q}\cdot\mathbf{x}} + h.c.]. \quad (2.4.65)$$

Equation (2.4.64) then becomes an ordinary differential equation governing the time evolution of the mode function $\mathcal{D}_q(t)$. Using conformal time as independent variable, it takes the form

$$\frac{d^2\mathcal{D}_q}{d\tau^2} + 2aH\frac{d\mathcal{D}_q}{d\tau} + q^2\mathcal{D}_q = 0. \quad (2.4.66)$$

This is the analogue of the Mukhanov-Sasaki equation, and it is again clear that it will have a solution that approaches a constant outside the horizon. Since the scale factor grows rapidly during inflation, if we go back in time far enough any mode will eventually be far inside the horizon, and the friction term will become negligible. In this limit, we can look for a WKB solution. To determine the normalization of the mode such that $\beta(\mathbf{q}, \lambda)$ and $\beta^*(\mathbf{q}, \lambda)$ satisfy canonical commutation relations

$$[\beta(\mathbf{q}, \lambda), \beta^*(\mathbf{q}', \lambda')] = \delta_{\lambda\lambda'} \delta(\mathbf{q} - \mathbf{q}'), \quad (2.4.67)$$

it is useful to notice that the action that follows for the two helicity modes of the graviton from (2.4.63), using $e_{ij}(\hat{q}, \lambda)e_{ij}^*(\hat{q}, \lambda') = 2\delta_{\lambda\lambda'}$, is the same as that for the Fourier components of two free, massless, minimally coupled scalar fields. One finds that for the modes to be canonically normalized, we should rescale D_{ij} by a factor $\sqrt{16\pi G}$. The initial conditions for \mathcal{D}_q should thus be the same as those of $\delta\phi_q$ given in equation (2.4.29) multiplied by a factor $\sqrt{16\pi G}$, *i.e.*

$$\mathcal{D}_q(\tau) = \frac{\sqrt{16\pi G}}{(2\pi)^{3/2} \sqrt{2qa}} e^{-iq\tau}. \quad (2.4.68)$$

Just like for the scalar modes, we will assume that the universe is in a state satisfying

$$\beta(\mathbf{q}, \lambda) |0\rangle = 0. \quad (2.4.69)$$

The equation (2.4.66) with initial conditions (2.4.68) could then simply be solved numerically for any given model, but it will again be interesting to look for an analytic solution in the slow-roll approximation. Using $aH \approx -(1+\epsilon)/\tau$, the equation (2.4.66) becomes

$$\frac{d^2 \mathcal{D}_q}{d\tau^2} + 2 \frac{1+\epsilon}{\tau} \frac{d\mathcal{D}_q}{d\tau} + q^2 \mathcal{D}_q = 0. \quad (2.4.70)$$

To leading order in the slow-roll parameter ϵ , the solution to this equation with initial conditions as in equation (2.4.68) is given by

$$\mathcal{D}_q(\tau) = \frac{\sqrt{16\pi G}(-q\tau)^{-\mu+1/2}}{4\pi\sqrt{2qa}} e^{i\pi\mu/2+i\pi/4} (-q\tau)^\mu H_\mu^{(1)}(-q\tau), \quad (2.4.71)$$

where $\mu \equiv \frac{3}{2} + \epsilon$, and the expression is again written such that the fraction is constant in time.²³ Using the asymptotic behavior for $H_\mu^{(1)}(x)$ in the limit $x \ll \sqrt{\mu+1}$ we obtain the value of \mathcal{D}_q^o

$$\mathcal{D}_q^o = -i \frac{\sqrt{16\pi G}(-\tau)^{-\mu+1/2}\Gamma(\mu)}{2\sqrt{\pi}(2\pi)^{3/2}a(\tau)} e^{i\pi\mu/2+i\pi/4} 2^\mu \frac{1}{q^\mu}, \quad (2.4.72)$$

implying that $\mathcal{D}_q^o \propto q^{-\mu}$. In terms of the quantity $\Delta_{\mathcal{D}}^2(q) = 16\pi |\mathcal{D}_q^o|^2 q^3$ this implies²⁴

$$\Delta_{\mathcal{D}}^2(q) = \Delta_{\mathcal{D}}^2(q_*) \left(\frac{q}{q_*}\right)^{-2\epsilon}, \quad (2.4.73)$$

so that single field slow-roll inflation predicts a nearly scale invariant spectrum of gravitational waves with tensor spectral index

$$n_t = -2\epsilon. \quad (2.4.74)$$

Notice that the slow roll parameter ϵ is positive definite as long as the null energy condition holds. An observation of what is sometimes referred to as a blue spectrum, $n_t > 0$, would thus require a violation of the null energy condition and in principle rule out single field slow-roll inflation. In practice,

²³It follows from $\frac{1}{a} \frac{da}{d\tau} = aH \approx -\frac{\mu-1/2}{\tau}$ that $a \propto \tau^{-\mu+1/2}$.

²⁴Notice the factor of 4 relative to the scalar case. In the tensor case, the numerical factors are chosen so that $\langle \mathcal{D}_{ij}(t_*, 0) \mathcal{D}_{ij}(t_*, 0) \rangle = \int d \ln q \Delta_{\mathcal{D}}^2(q)$, where t_* is a chosen so that the modes of interest are outside the horizon and $\mathcal{D}_q(t_*) = \mathcal{D}_q^o$. The factor 4 thus arises as a factor 2 from $e_{ij}(\hat{q}, \lambda) e_{ij}^*(\hat{q}, \lambda') = 2\delta_{\lambda\lambda'}$, and another factor 2 for the 2 helicity modes of the graviton.

it seems unlikely that this parameter will be measured sufficiently well to do so with confidence.

The amplitude can again be written nicely if we evaluate the expression at the time of horizon crossing.

$$\mathcal{D}_q^o = i \frac{\sqrt{8\pi G} H(t_q)}{(2\pi)^{3/2}} \frac{1}{q^{3/2}}, \quad (2.4.75)$$

so that

$$\Delta_{\mathcal{D}}^2(q) = \frac{2}{\pi^2} 8\pi G H^2(t_q), \quad (2.4.76)$$

or equivalently

$$\Delta_{\mathcal{D}}^2(q) = \Delta_{\mathcal{D}}^2(q_*) \left(\frac{q}{q_*} \right)^{n_t}, \quad (2.4.77)$$

with

$$\Delta_{\mathcal{D}}^2(q_*) = \frac{2}{\pi^2} 8\pi G H^2(t_{q_*}) \quad \text{and} \quad n_t = -2\epsilon(t_{q_*}). \quad (2.4.78)$$

Notice that the detection of a tensor signal and the measurement of its amplitude would directly tell us about the Hubble rate during inflation, or equivalently, about the energy scale of inflation. An important quantity in this context is the tensor to scalar ratio r . It is simply defined as

$$r(q) \equiv \frac{\Delta_{\mathcal{D}}^2(q)}{\Delta_{\mathcal{R}}^2(q)} = 16\epsilon(t_q). \quad (2.4.79)$$

Strictly speaking this ratio has a momentum dependence, but this dependence is expected to be rather mild, and it is quite common to ignore this and simply evaluate it at the pivot scale.

One can also use this expression to eliminate ϵ from the expression for the tensor spectral index. This leads to the slow-roll consistency condition

$$n_t = -\frac{r}{8}, \quad (2.4.80)$$

that directly relates two observable quantities and could in principle be used to test single field slow-roll inflation. Once again, in practice it seems unlikely that we will know these quantities well enough to use this as a conclusive test. For a mission like CMBPol, $\Delta n_t = 0.072$ for a tensor-to-scalar ratio of $r = 0.01$ in the absence of foregrounds and substantially worse in the presence of foregrounds. [117]. In theories with non-standard kinetic terms this expression gets modified by the speed of sound.

As was first noticed in [118] and studied more carefully in [119], [120], the tensor-to-scalar contains information about the distance traveled by the inflaton field during inflation. It so happens, that if a tensor signal is observable, the inflaton must have traversed a distance of order the Planck mass or larger. A naive way to see this is to treat $\epsilon_V = (V'/V)^2/16\pi G$ as roughly constant in equation (2.4.20). The number of e-folds of inflation N is then related to the distance traveled by the field $\Delta\phi$ as

$$N = - \int_{\phi_i}^{\phi_e} d\phi \frac{8\pi G V(\bar{\phi})}{V'(\bar{\phi})} \approx \frac{\sqrt{4\pi G} \Delta\phi}{\sqrt{\epsilon_V}}, \quad (2.4.81)$$

implying

$$\Delta\phi = N \left(\frac{r}{8}\right)^{1/2} M_P. \quad (2.4.82)$$

The number of e-folds required in a given model depend on the details of reheating. Using a conservative bound of $N \gtrsim 30$, this becomes²⁵

$$\Delta\phi \gtrsim 1.06 \left(\frac{r}{0.01}\right)^{1/2} M_P. \quad (2.4.83)$$

²⁵Strictly speaking we only know that the field was slowly rolling during the period when the modes we observe in the cosmic microwave background exited the horizon. At the very least this corresponds to $\Delta N \approx 5$, which gives a result that is smaller by a factor of $\sqrt{6}$, but does not change the qualitative conclusion.

A somewhat more careful analysis in [119] comes to the conclusion that models with a scalar spectral index in the range $0.92 < n_s < 1.06$ in fact satisfy the stronger bound²⁶

$$\Delta\phi \approx 10 \left(\frac{r}{0.01} \right)^{1/4} M_P. \quad (2.4.84)$$

This implies that the scalar field should have traversed a distance larger than M_P for a tensor-to-scalar ratio $r > 0.001$. This happens to be just what might be observable by a mission like CMBPol [117]. Furthermore, equation (2.4.62) implies

$$V_{\text{inf}} = \frac{3\pi^2}{2} \Delta_{\mathcal{R}}^2(q_*) r M_P^4. \quad (2.4.85)$$

Together with the value of the scalar amplitude derived from the five-year WMAP data, (2.0.10), this yields

$$V_{\text{inf}}^{1/4} = 1.06 \times 10^{16} \text{ GeV} \left(\frac{r}{0.01} \right)^{1/4}, \quad (2.4.86)$$

so that a measurement of the tensor amplitude, or equivalently the tensor-to-scalar ratio would immediately tell us about the energy scale of inflation.

The best bounds on the tensor-to-scalar ratio are currently derived from a combination of the WMAP data together with baryon acoustic oscillations observed in galaxy surveys and supernova data. The strongest constraint in this case arises from the tensor contribution to the temperature multipole coefficients at low multipole coefficients. The bound on the tensor-to-scalar ratio derived from the five-year WMAP data alone is $r < 0.43$ at 95% confidence level. The bound from WMAP as well as baryon acoustic oscillations and supernova data is $r < 0.22$ at 95% confidence level. The results of [78] are shown in Figure (2.6).

²⁶Reference [119] give their result in terms of the Planck mass. We use the reduced Planck mass throughout and converted their bound accordingly.

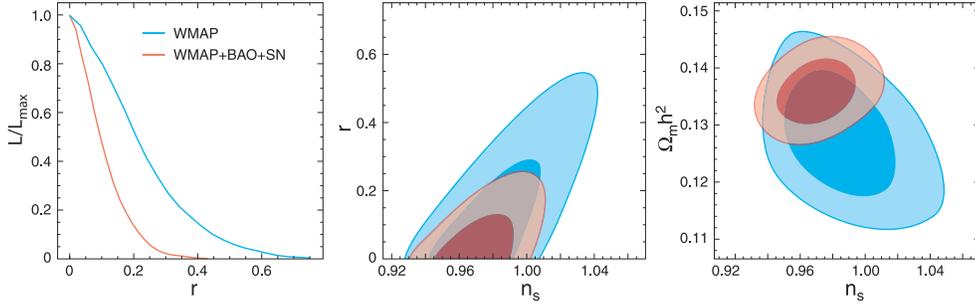


Figure 2.6: The left plot shows the one-dimensional marginalized distribution of r , showing the limit derived from WMAP alone, $r < 0.43$ (95% CL), and WMAP combined with baryon acoustic oscillations and supernova data, $r < 0.22$ (95% CL). The plot in the middle shows the two-dimensional marginalized distribution (68% and 95% CL), showing a strong degeneracy between n_s and r . The right plot shows a degeneracy between n_s and $\Omega_m h^2$. In all panels the results derived from WMAP alone are shown in blue, those for WMAP combined with baryon acoustic oscillations and supernovae in red.

The degeneracy between r and n_s can be understood from the fact that the main constraint on the tensor-to-scalar ratio comes from the low- ℓ TT spectrum, where the tensor signal has a plateau (see *e.g.* Figure 2.5). Raising n_s lowers the power in the scalar spectrum at low ℓ thus allowing for larger r . On the other hand raising n_s becomes possible only because of a degeneracy between $\Omega_m h^2$ and n_s so that adding the data from baryon acoustic oscillations and supernovae, which break the degeneracy between $\Omega_m h^2$ and n_s , significantly strengthens the bound. The degeneracy will disappear as soon as the main constraint comes from the polarization data. A first competitive result from the B -mode alone is $r < 0.73$ at 95% confidence level from the BICEP experiment [121].

Already at the present stage these results put meaningful constraints on inflationary models. Assuming that the potential is a power law $V \propto \phi^p$,

the tensor-to-scalar ratio and the scalar spectral index are given by

$$r = \frac{4p}{N} \quad \text{and} \quad n_s = 1 - \frac{p+2}{2N}. \quad (2.4.87)$$

In particular, for the once popular $\lambda\phi^4$ potential this implies $r = 0.32$ and $n_s = 0.94$ if we use $N = 50$, or $r = 0.27$ and $n_s = 0.95$ for $N = 60$. Both of these are far from the 95% confidence level contour in Figure 2.6, and the model is ruled out at more than 99% confidence level.

Depending on foregrounds, a mission like CMBPol would be sensitive roughly to $r > 0.001$, and could conclusively distinguish between small field models of inflation with $\Delta\phi \ll M_P$ and large field models with $\Delta\phi \gtrsim M_P$. While these experiments may not provide a huge amount of detail even in the case of a detection of a tensor signal, they would still provide *some* information about physics at these high energy scales that seem far out of reach for any other experiment.

Chapter 3

Tensor Microwave Background Fluctuations for Large Multipole Order

We now present approximate formulas for the tensor BB, EE, TT, and TE multipole coefficients for large multipole order ℓ . The error in using the approximate formula for the BB multipole coefficients is less than cosmic variance for $\ell > 10$. These approximate formulas make various qualitative properties of the calculated multipole coefficients transparent: specifically, they show that, whatever values are chosen for cosmological parameters, the tensor EE multipole coefficients will always be larger than the BB coefficients for all $\ell > 15$, and that these coefficients will approach each other for $\ell \ll 100$. These approximations also make clear how these multipole coefficients depend on cosmological parameters. These approximations were the subject of [122].

As explained in subsection 2.4.3, tensor fluctuations are a prime target for future observations of the cosmic microwave background, because if detected they can provide a conclusive verification of the theory of inflation and a unique tool for exploring the details of this theory. The contribution of these fluctuations to the correlations of temperature and polarization correlations is well known and was reviewed in subsection 2.3.4.2. Their contributions to the

multipole coefficients are given by¹

$$C_{EE,\ell}^{(T)} = \pi^2 T_0^2 \int_0^\infty q^2 dq \quad (3.0.1)$$

$$\times \left| \int_{t_1}^{t_0} dt P(t) \Psi(q, t) \left\{ \left[12 + 8\rho \frac{\partial}{\partial \rho} - \rho^2 + \rho^2 \frac{\partial^2}{\partial \rho^2} \right] \frac{j_\ell(\rho)}{\rho^2} \right\}_{\rho=q r(t)} \right|^2,$$

$$C_{BB,\ell}^{(T)} = \pi^2 T_0^2 \int_0^\infty q^2 dq \quad (3.0.2)$$

$$\times \left| \int_{t_1}^{t_0} dt P(t) \Psi(q, t) \left\{ \left[8\rho + 2\rho^2 \frac{\partial}{\partial \rho} \right] \frac{j_\ell(\rho)}{\rho^2} \right\}_{\rho=q r(t)} \right|^2,$$

$$C_{TE,\ell}^{(T)} = -2\pi^2 T_0^2 \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \int_0^\infty q^2 dq \quad (3.0.3)$$

$$\times \int_{t_1}^{t_0} dt P(t) \Psi(q, t) \left\{ \left[12 + 8\rho \frac{\partial}{\partial \rho} - \rho^2 + \rho^2 \frac{\partial^2}{\partial \rho^2} \right] \frac{j_\ell(\rho)}{\rho^2} \right\}_{\rho=q r(t)}$$

$$\times \int_{t_1}^{t_0} dt' d(q, t') \left\{ \frac{j_\ell(q r(t'))}{q^2 r^2(t')} \right\},$$

$$C_{TT,\ell}^{(T)} = \frac{4\pi^2(\ell+2)!T_0^2}{(\ell-2)!} \int_0^\infty q^2 dq \left| \int_{t_1}^{t_0} dt d(q, t) \frac{j_\ell(q r(t))}{q^2 r^2(t)} \right|^2. \quad (3.0.4)$$

Once again: T_0 is the microwave background temperature at the present time t_0 ; $P(t) = \omega_c(t) \exp[-\int_t^{t_0} \omega_c(t') dt']$ is the probability distribution of last scattering, with $\omega_c(t)$ the photon collision frequency; t_1 is any time taken early enough before recombination so that any photon present at t_1 would have collided many times before the present; $r(t) = \int_t^{t_0} dt'/a(t')$ is the co-moving radial coordinate of a source from which light emitted at time t would reach

¹These formulas are equivalent to those of Zaldarriaga and Seljak [57] (See also [82]. An equivalent analysis was given in [123],[56]). Their gravitational wave amplitude h and power spectral function $P_h(k)$ are related to our gravitational wave amplitude $\mathcal{D}_q(t)$ by $h\sqrt{P_h} = \mathcal{D}/2$. In consequence, their function $\Psi\sqrt{P_h}$ is 1/4 times our source function Ψ .

us at the origin at the present time t_0 ; and $\Psi(q, t)$ is the “source function,” which is customarily calculated from a hierarchy of equations for partial-wave amplitudes, equations (2.3.97) and (2.3.98) derived originally in [46], [124]. The quantity $d(q, t)$ is given by

$$d(q, t) \equiv \exp \left[- \int_t^{t_0} dt' \omega_c(t') \right] \left(\dot{\mathcal{D}}_q(t) - \frac{1}{2} \omega_c(t) \Psi(q, t) \right), \quad (3.0.5)$$

where $\mathcal{D}_q(t)$ is the gravitational wave amplitude defined through equations (2.3.12) and (2.3.44).

Aside from the treatment of the tensor mode as a first-order perturbation, and the assumption of purely elastic Thomson scattering, Eqs. (3.0.1)–(3.0.4) may be regarded as exact. They serve as the basis of computer programs such as CMBfast and CAMB, that are used to compare observations of microwave background polarization and temperature fluctuations with models that predict values for the gravitational wave amplitude $\mathcal{D}_q(t)$. But they are not very transparent.

For one thing, as shown in Figure 3.1, computer calculations using Eqs. (3.0.1) and (3.0.2) yield results for $C_{EE,\ell}^{(T)}$ and $C_{BB,\ell}^{(T)}$ that are of the same order of magnitude, and nearly equal for $\ell < 100$, while $C_{EE,\ell}^{(T)} > C_{BB,\ell}^{(T)}$ for all $\ell > 15$.

Of course computer calculations can only show this for specific choices of cosmological parameters.² It would be impossible to conclude just by inspection of Eqs. (3.0.1) and (3.0.2) that these are general properties of the multipole coefficients, independent of the choice of cosmological parameters.

²We will use the parameters derived from the five-year WMAP data alone, *i.e.* those of equations (2.0.6)–(2.0.11), except that reionization is ignored and τ set to zero. The tensor to scalar ratio is taken to be $r = 1$, and the tensor spectral index is set to $n_t = 0$.

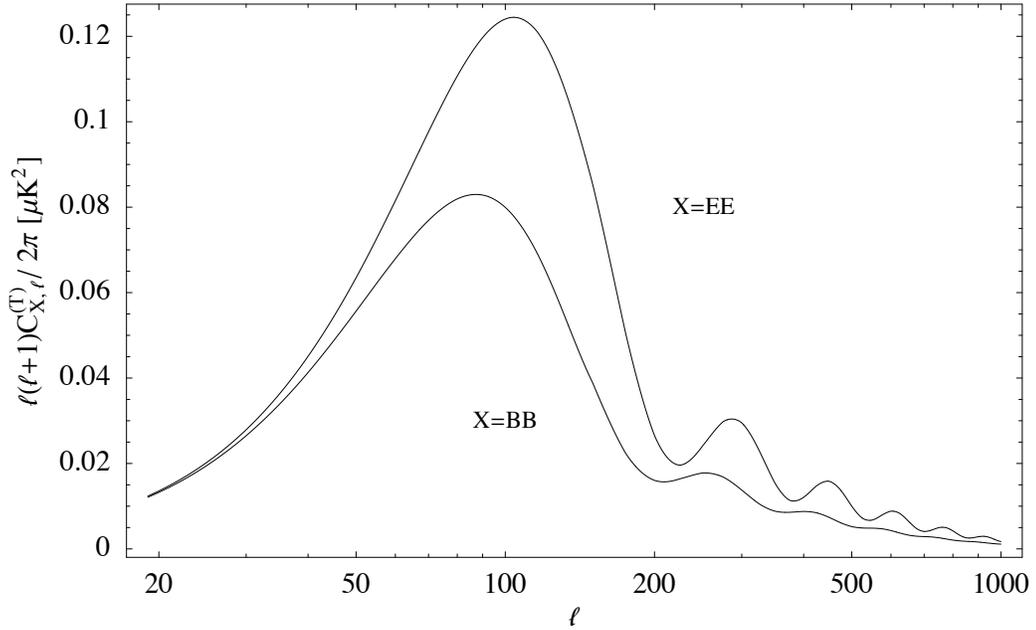


Figure 3.1: Comparison of $C_{EE,\ell}^{(T)}$ and $C_{BB,\ell}^{(T)}$, in $(\mu\text{K})^2$.

In this paper we present successive approximations that make these properties apparent, and that, at the cost of only a small additional loss in accuracy, also clarify how the multipole coefficients depend on various cosmological parameters.

3.1 The Large- ℓ Approximation

We can approximate Eqs. (3.0.1)–(3.0.4) by much simpler and more transparent formulas, by using an asymptotic formula [125] for the spherical

Bessel functions ³:

$$j_\ell(\rho) \rightarrow \begin{cases} \frac{\cos b \cos [\nu(\tan b - b) - \pi/4]}{\nu\sqrt{\sin b}} & \rho > \nu \\ 0 & \rho < \nu \end{cases}, \quad (3.1.1)$$

where $\nu \equiv \ell + 1/2$, and $\cos b \equiv \nu/\rho$, with $0 \leq b \leq \pi/2$. This approximation is valid for $|\nu^2 - \rho^2| \gg \nu^{4/3}$. Hence for $\ell \gg 1$, this formula can be used over most of the ranges of integration in Eqs. (3.0.1)–(3.0.4). Furthermore, for $\rho > \nu \gg 1$ the phase $\nu(\tan b - b)$ in Eq. (3.1.1) is a very rapidly increasing function of ρ , so the derivatives in Eqs. (3.0.1)–(3.0.3) can be taken to act chiefly on this phase:

$$\begin{aligned} \left[12 + 8\rho \frac{\partial}{\partial \rho} - \rho^2 + \rho^2 \frac{\partial^2}{\partial \rho^2} \right] \frac{j_\ell(\rho)}{\rho^2} &\rightarrow -j_\ell(\rho) + j_\ell''(\rho) \\ &\rightarrow -\frac{(1 + \sin^2 b) \cos b}{\nu\sqrt{\sin b}} \cos [\nu(\tan b - b) - \pi/4] \end{aligned} \quad (3.1.2)$$

$$\begin{aligned} \left[8\rho + 2\rho^2 \frac{\partial}{\partial \rho} \right] \frac{j_\ell(\rho)}{\rho^2} &\rightarrow 2j_\ell'(\rho) \\ &\rightarrow -\frac{2\sqrt{\sin b} \cos b}{\nu} \sin [\nu(\tan b - b) - \pi/4] \end{aligned} \quad (3.1.3)$$

³The same approximation is used by J. R. Pritchard and M. Kamionkowski, [126]. However, after making this approximation they make further approximations that are quite different from ours, and that lead to a divergence in the integral over wave number, which must be dealt with by an arbitrary cut-off. The error introduced by their approximation is comparable to the one introduced by our last approximation given by Eqs. (3.2.2) and (3.2.3). Another approximation that consists in averaging over the rapid oscillations in the square of the Bessel functions leading to results similar to our last approximation was proposed by M. Zaldarriaga and D. D. Harari, [54]. An analytic expression for the contribution of the tensor modes to the temperature multipole coefficients approximately valid for $1 \ll \ell < 50$ obtained using a similar average was given by A. A. Starobinsky, [127]

Then Eqs. (3.0.1)–(3.0.4) become, for $\nu = \ell + 1/2 \gg 1$,

$$\ell(\ell + 1)C_{EE,\ell}^{(T)} \rightarrow \pi^2 T_0^2 \int_0^\infty q^2 dq \quad (3.1.4)$$

$$\times \left| \int_{r(t) > \nu/q} dt P(t) \Psi(q, t) \right. \\ \left. \times \left\{ \frac{(1 + \sin^2 b) \cos b}{\sqrt{\sin b}} \cos \left[\nu(\tan b - b) - \pi/4 \right] \right\}_{\cos b = \nu/q r(t)} \right|^2,$$

$$\ell(\ell + 1)C_{BB,\ell}^{(T)} \rightarrow \pi^2 T_0^2 \int_0^\infty q^2 dq \quad (3.1.5)$$

$$\times \left| \int_{r(t) > \nu/q} dt P(t) \Psi(q, t) \right. \\ \left. \times \left\{ 2\sqrt{\sin b} \cos b \sin \left[\nu(\tan b - b) - \pi/4 \right] \right\}_{\cos b = \nu/q r(t)} \right|^2,$$

$$\ell(\ell + 1)C_{TE,\ell}^{(T)} \rightarrow -2\pi^2 T_0^2 \int_0^\infty q^2 dq \quad (3.1.6)$$

$$\times \int_{r(t) > \nu/q} dt P(t) \Psi(q, t) \\ \times \left\{ \frac{(1 + \sin^2 b) \cos b}{\sqrt{\sin b}} \cos \left[\nu(\tan b - b) - \pi/4 \right] \right\}_{\cos b = \nu/q r(t)} \\ \times \int_{r(t') > \nu/q} dt' d(q, t') \left\{ \frac{\cos^3 b}{\sqrt{\sin b}} \cos \left[\nu(\tan b - b) - \pi/4 \right] \right\}_{\cos b = \nu/q r(t')},$$

$$\ell(\ell + 1)C_{TT,\ell}^{(T)} \rightarrow 4\pi^2 T_0^2 \int_0^\infty q^2 dq \quad (3.1.7)$$

$$\times \left| \int_{r(t) > \nu/q} dt d(q, t) \left\{ \frac{\cos^3 b}{\sqrt{\sin b}} \cos \left[\nu(\tan b - b) - \pi/4 \right] \right\}_{\cos b = \nu/q r(t)} \right|^2.$$

In evaluating both the exact and approximate expressions, instead of calculating the source function $\Psi(q, t)$ by truncating the Boltzmann hierarchy (2.3.97), (2.3.98), we use the integral equation (2.3.103) derived in [75],[76]

$$\begin{aligned} \Psi(q, t) &= \frac{3}{2} \int_{t_1}^t dt' \exp \left[- \int_{t'}^t \omega_c(t'') dt'' \right] \\ &\times \left[- 2\dot{\mathcal{D}}_q(t') K \left(q \int_{t'}^t \frac{dt''}{a(t'')} \right) + \omega_c(t') F \left(q \int_{t'}^t \frac{dt''}{a(t'')} \right) \Psi(q, t') \right], \end{aligned} \quad (3.1.8)$$

where $K(v)$ and $F(v)$ are the functions

$$K(v) \equiv j_2(v)/v^2, \quad F(v) \equiv j_0(v) - 2j_1(v)/v + 2j_2(v)/v^2. \quad (3.1.9)$$

(This is not an approximation; in principle it should give the same results as the truncated Boltzmann hierarchy used by CMBfast and CAMB, aside from the supposedly small errors produced by the truncation. In fact, our method gives results that differ by a few percent from both CMBfast and CAMB, but CMBfast and CAMB give results for both $C_{EE,\ell}^T$ and $C_{BB,\ell}^T$ that differ by similar amounts from each other, especially for large ℓ . At this point we are not able to tell which of the three methods is the most reliable.) The specific cosmological model chosen for this and all other numerical calculations in this chapter is again the Λ CDM model with the values for the parameters derived from the five-year WMAP data (2.0.6)-(2.0.11). In calculating the photon collision frequency, we use the recfast recombination code [71],[72], with helium abundance $Y = 0.24$. The gravitational field amplitude outside the horizon is taken as

$$|\mathcal{D}_q|^2 = 4.79 \times 10^{-11} q^{-3}$$

corresponding to $n_T = 0$, $\Delta_{\mathcal{R}}^2(q_* = 0.002 \text{ Mpc}^{-1}) = 2.41 \times 10^{-9}$, and a tensor/scalar ratio $r = 1$. The gravitational wave amplitude $\mathcal{D}_q(t)$ is calculated including the damping due to neutrino anisotropic inertia, as in [84]; the effects of photon anisotropic inertia are negligible. Reionization is ignored and

the optical depth set to $\tau = 0$. To take into account a finite optical depth τ of the reionized plasma or a different value of r , for $\ell \gtrsim 15$ it is only necessary to multiply the multipole coefficients given here by $r \exp(-2\tau)$. The approximate results obtained in this way from Eqs. (3.1.4)–(3.1.7) are compared with the exact formulas (3.0.1)–(3.0.4) in Figures 3.2–3.5.

We can gain further simplicity and transparency in the formulas for the EE and BB multipole coefficients by using another approximation that actually leads to improved accuracy for $C_{BB,\ell}^T$. The last-scattering probability distribution $P(t)$ is concentrated around a time t_L , corresponding to a redshift $z_L \simeq 1088$. For any q of the same order of magnitude as $\nu/r(t_L)$, the quantity $b \equiv \cos^{-1}(\nu/qr(t))$ does not vary appreciably for t within the range in which $P(t)$ is appreciable. Hence we can set $r(t)$ equal to r_L everywhere *except* in the phase $\nu(\tan b - b)$, which for $\nu \gg 1$ does vary over a wide range in this interval. Furthermore, because $\nu(\tan b - b)$ varies over a wide range for $\nu \gg 1$, the difference between $\cos[\nu(\tan b - b) - \pi/4]$ and $\sin[\nu(\tan b - b) - \pi/4]$ is immaterial, and we can replace both with $\cos[\nu(\tan b - b)]$. Making these replacements in Eqs. (3.1.4) and (3.1.5) gives

$$\begin{aligned} \ell(\ell + 1)C_{EE,\ell}^{(T)} &\rightarrow \pi^2 T_0^2 \int_{\nu/r_L}^{\infty} q^2 dq \{(1 + \sin^2 b_L)^2 \cos^2 b_L\}_{\cos b_L = \nu/qr_L} \\ &\times \left| \int_{r(t) > \nu/q} dt P(t) \Psi(q, t) \left\{ \frac{\cos[\nu(\tan b - b)]}{\sqrt{\sin b}} \right\}_{\cos b = \nu/qr(t)} \right|^2, \end{aligned} \quad (3.1.10)$$

$$\begin{aligned} \ell(\ell + 1)C_{BB,\ell}^{(T)} &\rightarrow \pi^2 T_0^2 \int_{\nu/r_L}^{\infty} q^2 dq \{4 \sin^2 b_L \cos^2 b_L\}_{\cos b_L = \nu/qr_L} \\ &\times \left| \int_{r(t) > \nu/q} dt P(t) \Psi(q, t) \left\{ \frac{\cos[\nu(\tan b - b)]}{\sqrt{\sin b}} \right\}_{\cos b = \nu/qr(t)} \right|^2 \end{aligned} \quad (3.1.11)$$

(We have not set $b = b_L$ in the factors $1/\sqrt{\sin b}$ in both integrals over t , in order to avoid a divergence in the integration over q at $q = \nu/r_L$. This factor does not introduce a divergence in the integrals over time, because $dt \propto \sin b db$.)

These approximate formulas are compared with results of the exact formulas (3.0.1) and (3.0.2) in Figures 3.6 and 3.7. The approximate result (3.1.11) for $C_{BB,\ell}^T$ agrees with the exact result (3.0.2) to about 1% for all $\ell > 10$, which is better than cosmic variance. The approximate result (3.1.10) for $C_{EE,\ell}^T$ is not quite as accurate; it agrees with the exact result (3.0.1) to better than about 14% for all $\ell > 10$. These approximations are evidently accurate enough for us to draw qualitative conclusions about the EE and BB multipole coefficients.

One immediate consequence is that, since $(1 + \sin^2 b_L)^2 \geq 4 \sin^2 b_L$ for all real b_L , we expect that $C_{EE,\ell}^{(T)} \geq C_{BB,\ell}^{(T)}$ for all ℓ large enough to justify our approximations. Also, since $\Psi(q, t_L)$ falls off for wave lengths that come into the horizon before matter-radiation equality, we expect that for relatively small ℓ (say, $\ell < 100$) the integrals over q are dominated by values for which $\cos b_L$ is small, so that $(1 + \sin^2 b_L)^2 \simeq 4 \sin^2 b_L$, and hence $C_{EE,\ell}^{(T)} \simeq C_{BB,\ell}^{(T)}$ for such ℓ . As mentioned in Section I, and shown in Figure 3.1, both properties are observed in the output of numerical calculations based on the accurate formulas (3.0.1) and (3.0.2).

3.2 Parameter-Dependence of the EE and BB Correlations

With one further approximation, we can find reasonably accurate formulas for $C_{EE,\ell}^{(T)}$ and $C_{BB,\ell}^{(T)}$ that reveal the way that these coefficients depend on various cosmological parameters. We write the squared time integrals in Eqs. (3.1.10) and (3.1.11) as double integrals over times t and t' , and write

$$\begin{aligned} & \cos \left[\nu(\tan b - b) \right] \cos \left[\nu(\tan b' - b') \right] = \\ & \frac{1}{2} \left[\cos \left[\nu(\tan b - b) - \nu(\tan b' - b') \right] + \cos \left[\nu(\tan b - b) + \nu(\tan b' - b') \right] \right], \end{aligned} \quad (3.2.1)$$

where $\cos b = \nu/qr(t)$ and $\cos b' = \nu/qr(t')$. For $\nu \gg 1$, and $qr(t)$ and $qr(t')$ both of order ν , the second term on the right oscillates very rapidly, and hence may be neglected in the integral over t and t' . On the other hand, because $P(t)$ and $P(t')$ are sharply peaked around the same time t_L , the argument of the first cosine on the right is small where $P(t)$ and $P(t')$ are appreciable, so this cosine may be replaced with unity. Then (now dropping the distinction between ν and ℓ), Eqs. (3.1.10) and (3.1.11) become

$$\begin{aligned} \ell(\ell + 1)C_{EE,\ell}^{(T)} & \rightarrow \frac{\pi^2 T_0^2}{2} \int_{\ell/r_L}^{\infty} q^2 dq \{ (1 + \sin^2 b_L)^2 \cos^2 b_L \}_{\cos b_L = \ell/qr_L} \\ & \times \left| \int_{r(t) > \ell/q} dt P(t) \Psi(q, t) \left(1 - \frac{\ell^2}{q^2 r^2(t)} \right)^{-1/4} \right|^2, \end{aligned} \quad (3.2.2)$$

$$\begin{aligned} \ell(\ell + 1)C_{BB,\ell}^{(T)} & \rightarrow \frac{\pi^2 T_0^2}{2} \int_{\ell/r_L}^{\infty} q^2 dq \{ 4 \sin^2 b_L \cos^2 b_L \}_{\cos b_L = \ell/qr_L} \\ & \times \left| \int_{r(t) > \ell/q} dt P(t) \Psi(q, t) \left(1 - \frac{\ell^2}{q^2 r^2(t)} \right)^{-1/4} \right|^2. \end{aligned} \quad (3.2.3)$$

This approximation is compared with the results of the exact formulas (3.0.1) and (3.0.2) in Figures 3.8 and 3.9. As shown there, the fractional error here is less than about 20% for $10 < \ell < 600$, but it becomes larger for larger values of ℓ , where the multipole coefficients become quite small.

Eqs. (3.2.2) and (3.2.3) are useful in revealing the parameter dependence of these multipole coefficients. Where the last-scattering probability distribution $P(t)$ is appreciable, the only cosmological parameters on which either $P(t)$ or the source function $\Psi(q, t)$ depend are the baryonic and matter density parameters $\Omega_B h^2$ and $\Omega_M h^2$, as well as the present microwave background temperature T_0 . All dependence of the multipole coefficients on H_0 or the curvature $\Omega_K h^2$ or the vacuum energy $\Omega_\Lambda h^2$ is contained in the function $r(t)$. But Eqs. (3.2.2) and (3.2.3) show that $r(t)$ and ℓ enter in the multipole coefficients only in the combination $r(t)/\ell$. Hence, with $\Omega_B h^2$, $\Omega_M h^2$, and T_0 fixed, to a good approximation $C_{EE,\ell}^{(T)}$ and $C_{BB,\ell}^{(T)}$ depend on H_0 , $\Omega_K h^2$, and $\Omega_\Lambda h^2$ only through their effect on the scale of the ℓ -dependence of $C_{EE,\ell}^{(T)}$ and $C_{BB,\ell}^{(T)}$. Furthermore, since $P(t)$ is sharply peaked at the time of last scattering, just as for scalar modes there is a high degree of degeneracy here: for $\ell > 10$ the coefficients $C_{EE,\ell}^{(T)}$ and $C_{BB,\ell}^{(T)}$ depend on H_0 , $\Omega_K h^2$, and $\Omega_\Lambda h^2$ only through a single parameter, the radius $r(t_L)$ of the surface of last scattering. Of course, the degeneracy here is not as important as it is for scalar modes, because tensor modes when discovered will be studied primarily for the purpose of measuring the tensor/scalar ratio r and the tensor slope n_T , rather than other cosmological parameters.

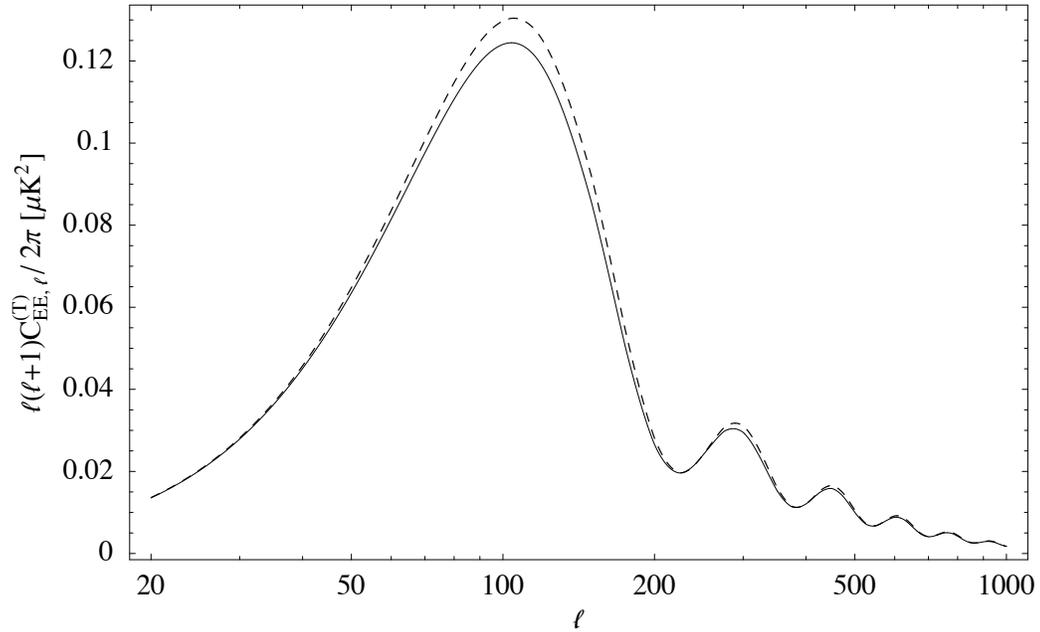


Figure 3.2: Comparison of formulas for $C_{EE,\ell}^T$. The solid line is the result of using the exact expression (3.0.1); the dashed line is the result of using the approximation (3.1.4). Figures 2–5 show the degree of accuracy of the large- ℓ approximation by itself, without further approximations. In this and all other figures, all calculations are done using the cosmological parameters given in Section II, and the units of the vertical axis are square microKelvins.

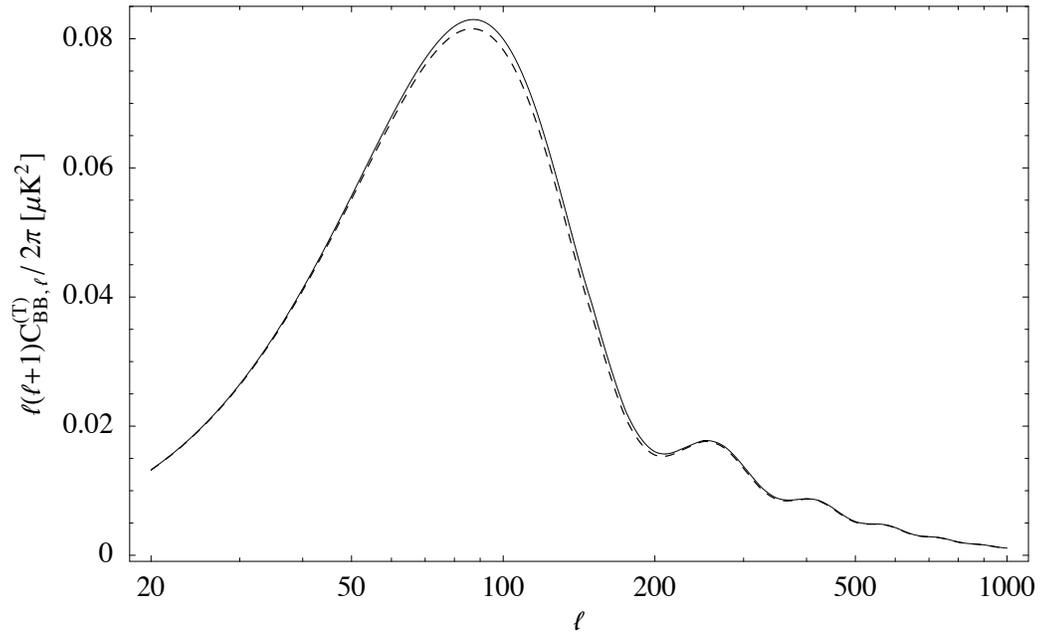


Figure 3.3: Comparison of formulas for $C_{BB,\ell}^T$. The solid line is the result of using the exact expression (3.0.2); the dashed line is the result of using the approximation (3.1.5).

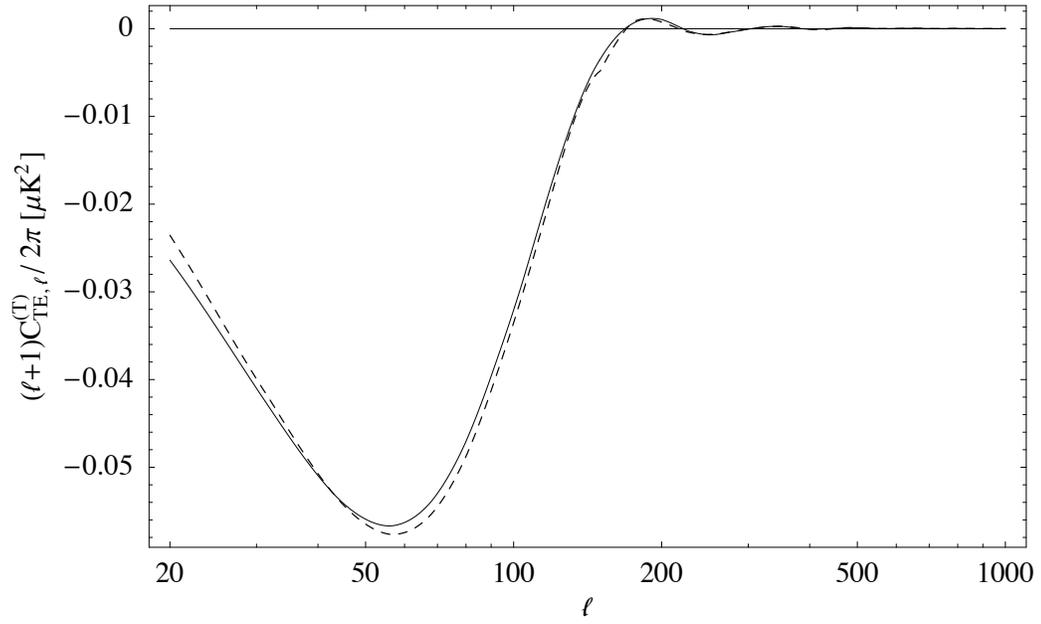


Figure 3.4: Comparison of formulas for $C_{TE, \ell}^T$. The solid line is the result of using the exact expression (3.0.3); the dashed line is the result of using the approximation (3.1.6).

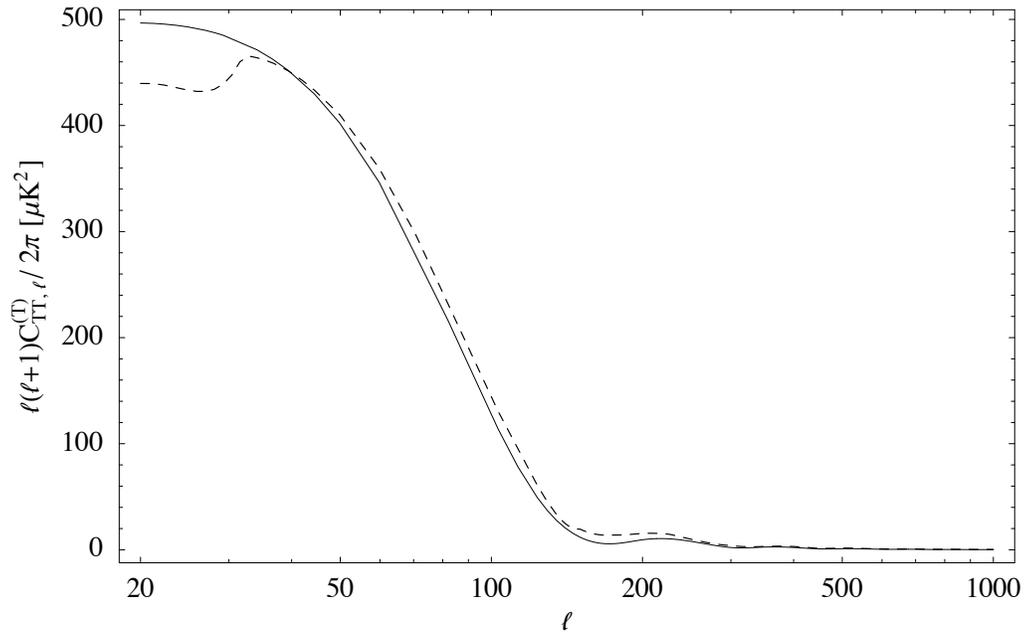


Figure 3.5: Comparison of formulas for $C_{TT, \ell}^T$. The solid line is the result of using the exact expression (3.0.4); the dashed line is the result of using the approximation (3.1.7).

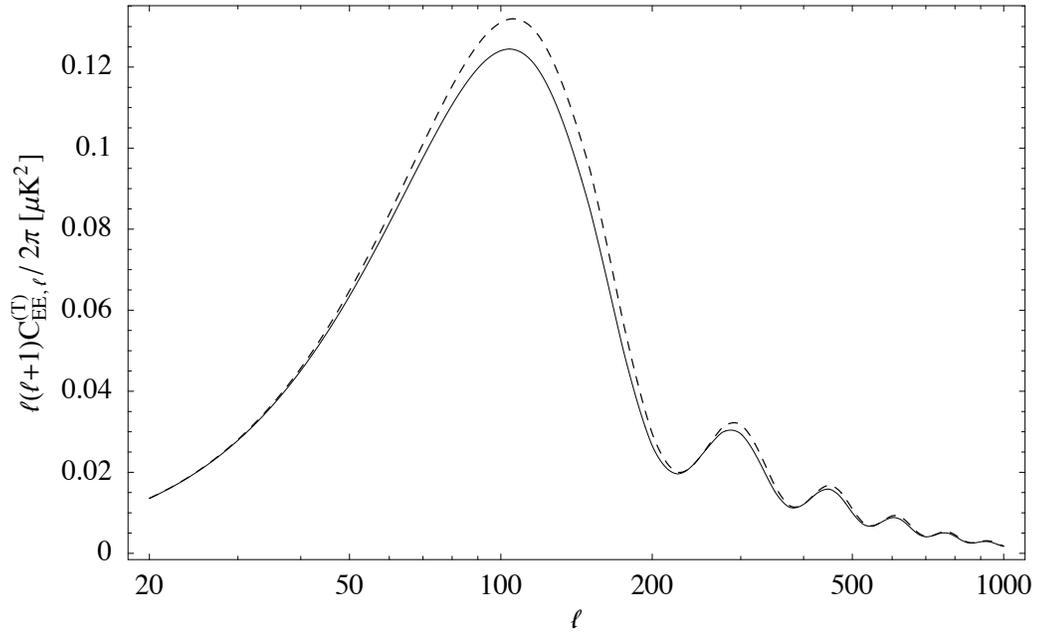


Figure 3.6: Comparison of formulas for $C_{EE,\ell}^T$. The solid line is the result of using the exact expression (3.0.1); the dashed line is the result of using the approximation (3.1.10). Figures 6 and 7 show the degree of accuracy of the combined approximations that we use to show analytically that $C_{EE,\ell}^T > C_{BB,\ell}^T$.

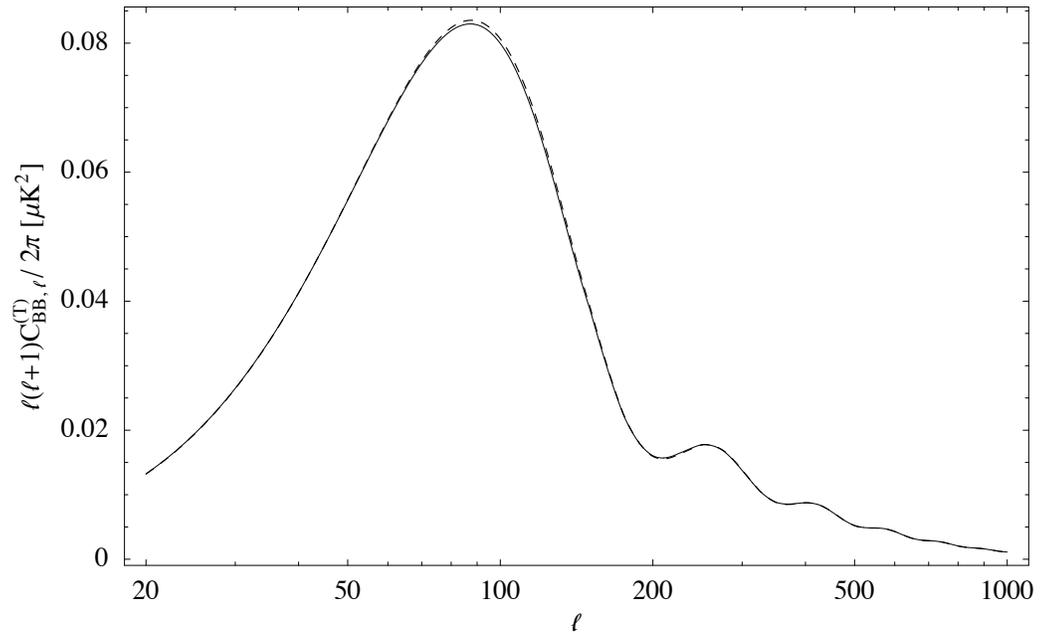


Figure 3.7: Comparison of formulas for $C_{BB,\ell}^T$. The solid line is the result of using the exact expression (3.0.2); the dashed line is the result of using the approximation (3.1.11). This is our best approximation for $C_{BB,\ell}^T$

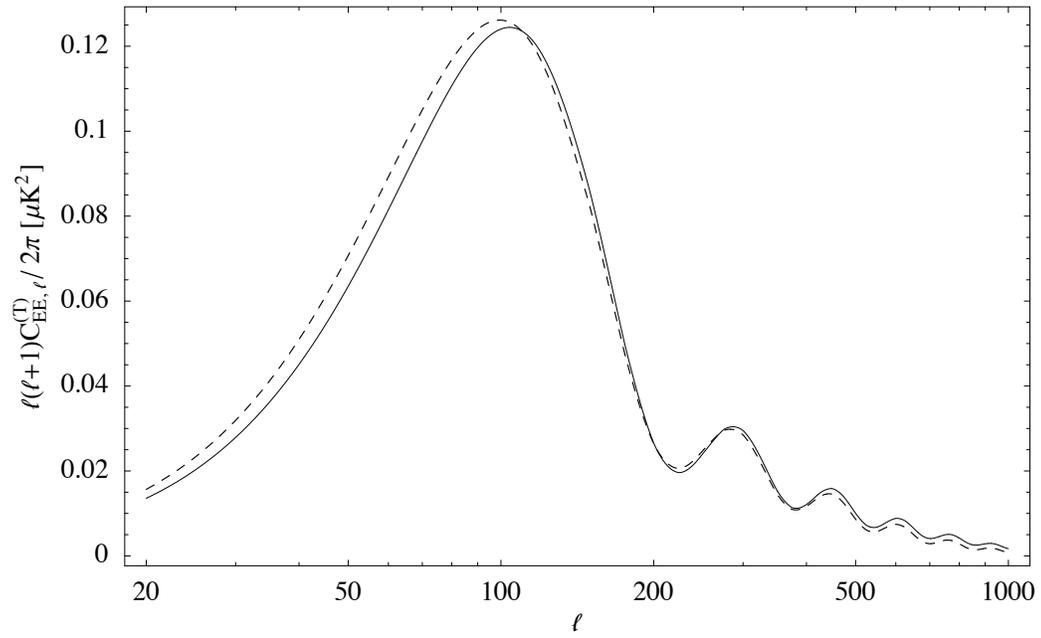


Figure 3.8: Comparison of formulas for $C_{EE,\ell}^T$. The solid line is the result of using the exact expression (3.0.1); the dashed line is the result of using the approximation (3.2.2). This figure shows the degree of accuracy of the further approximations used to explore the parameter-dependence of $C_{EE,\ell}^T$.

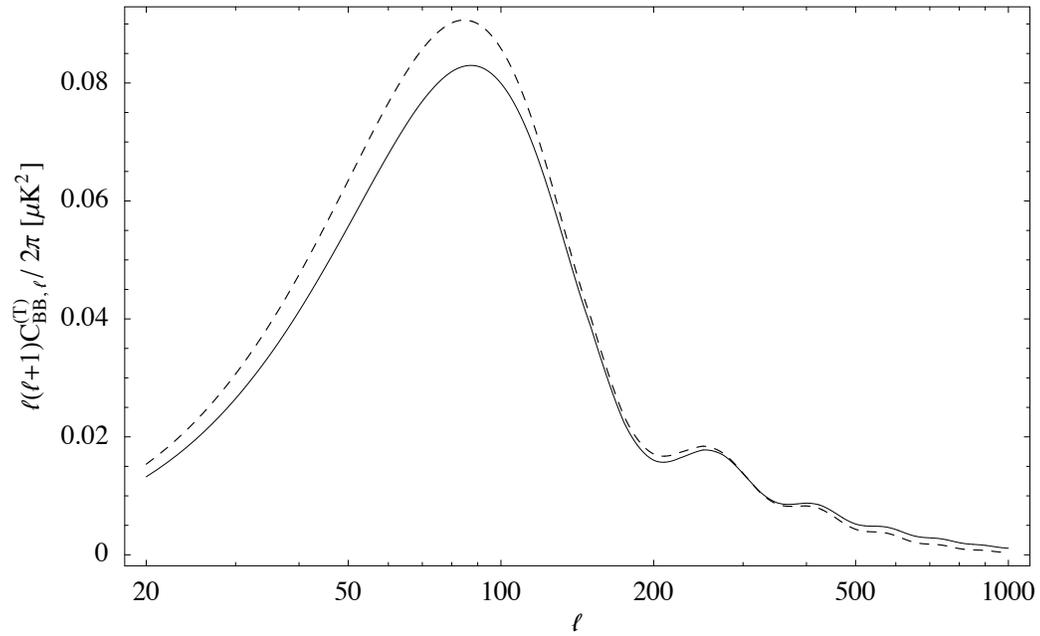


Figure 3.9: Comparison of formulas for $C_{BB,\ell}^T$. The solid line is the result of using the exact expression (3.0.2); the dashed line is the result of using the approximation (3.2.3). This figure shows the degree of accuracy of the further approximations used to explore the parameter-dependence of $C_{BB,\ell}^T$.

Chapter 4

Signatures of a Graviton Mass in the Cosmic Microwave Background

There exist consistent low energy effective field theories describing gravity in the Higgs phase that allow the coexistence of massive gravitons and the conventional $1/r$ potential of gravity. In an effort to constrain the value of the graviton mass in these theories, in this chapter, we study the tensor contribution to the CMB temperature anisotropy and polarization spectra in the presence of a non-vanishing graviton mass. The conclusion is that the observation of a B-mode signal consistent with the spectrum predicted by inflationary models would provide the strongest limit yet on the mass of an elementary particle – a graviton – at a level of $m \lesssim 10^{-30} \text{ eV} \approx (10 \text{ Mpc})^{-1}$. We also find that a graviton mass in the range between $(10 \text{ Mpc})^{-1}$ and $(10 \text{ kpc})^{-1}$ leads to interesting modifications of the polarization spectrum. The characteristic signature of a graviton mass in this range would be a plateau in the B -mode spectrum up to angular multipoles of $\ell \sim 100$. For even larger values of the graviton mass the tensor contribution to the CMB spectra becomes strongly suppressed.

The possibility of a non-zero graviton mass is intriguing and has attracted the attention of theorists for a long time, see e.g. [128],[129],[130],[131],[132],[133] (see [134] for a recent review). Massive gravitons have a number of peculiar properties, such as the van Dam-Veltman-Zakharov (vDVZ) discontinu-

ity [129],[130], ghost instabilities [132] and strong coupling effects at unacceptably low energy scales [133] which complicate the construction of a sensible theory. However, in light of the cosmological constant problem and coincidence problems between baryonic matter, dark matter, and dark energy, there has been increased interest in massive gravity theories, and it was found that models of modified gravity with Lorentz-violating graviton mass terms may avoid all of these problems [135], [136], [137]. The effective field theories remain valid up to reasonably large energy scales provided a large enough subgroup of the full diffeomorphism group is left unbroken by the graviton mass.

From a phenomenological point of view, the class of models characterized by the residual local symmetry $x^i \rightarrow x^i + \xi^i(t)$ is of particular interest. Among the possible choices of residual subgroups required by consistency, this is the only choice giving rise to massive gravitational waves [137].¹ Some phenomenological and cosmological consequences of these models were studied in Refs. [138], [139], [140],[141]. In particular, late time cosmological attractors have been found where an additional dilatation symmetry $t \rightarrow \lambda t$, $x^i \rightarrow \lambda^{-\gamma} x^i$ gets restored, where γ is a real constant. The main properties of the system in the vicinity of these attractors can be summarized as follows.

- (i) The evolution of the background cosmology is described by the usual Friedmann equation. For suitable choices of the Lagrangian in the symmetry breaking sector that leads to the graviton mass there will be an additional “dark energy” component that may contribute to the observed accelerated expansion of the universe.

¹Note that in the Lorentz-violating theories a presence of graviton mass terms does not yet imply that the tensor modes are massive. For instance, the graviton mass term $m_{00}^2 h_{00}^2$ just fixes the gauge $h_{00} = 0$.

- (ii) The equations describing the evolution of the scalar and vector fluctuations coincide with those of General Relativity.
- (iii) The equation describing the evolution of the tensor fluctuations is modified by the presence of a mass term. In other words, the dispersion relation for gravitational waves takes the form $\omega^2 = p^2 + m_g^2$.

As a consequence of (i) and (ii), the strongest bounds on the graviton mass, m_g , in this class of models come from direct or indirect observations of gravitational waves. Since these observations are very limited so far, this opens up the possibility for m_g to be quite large. The best constraint on the graviton mass in these models currently comes from indirect evidence for the emission of gravitational waves from binary pulsars timing [142] yielding an upper limit of

$$m_g \lesssim 10^4 \text{ pc}^{-1} \approx 3 \times 10^{-15} \text{ cm}^{-1} \approx 6 \times 10^{-20} \text{ eV}. \quad (4.0.1)$$

Relatively large graviton masses $m_g \gtrsim 0.1 \text{ pc}^{-1}$ could be detected by observing the characteristic signature of a large graviton mass, a strong monochromatic signal in gravitational wave detectors due to relic gravitons at a frequency equal to the graviton mass [138]. This signal might be observed either by LISA or using millisecond pulsar timing data.

Graviton masses close to the bound (4.0.1) can also be found by using higher frequency gravitational wave detectors to measure a time delay between optical and gravitational wave signals from a distant source.

Finally, these theories may give rise to non-universality of high multipoles of the galactic black hole metric. If present, these may be detected by LISA [141].

The characteristic energy scale Λ of the symmetry breaking sector that leads to a massive graviton is of order $\Lambda \sim \sqrt{M_{Pl}m_g}$. This scale is of the same order as the energy density scale of the Universe now if $m_g \sim H_0$ where $H_0 \approx 0.7 \times (3 \text{ Gpc})^{-1}$ is the current Hubble constant. In other words, if the observed acceleration of the Universe were due to some modification of gravity that gives rise to a graviton mass, one would expect the mass to be of order H_0 , far below the reach of the experiments mentioned above. This motivates us to look for signatures that are sensitive to much smaller values of the graviton mass.

One expects that a small graviton mass will leave an imprint in the temperature anisotropy and polarization spectra of the cosmic microwave background (CMB). We study the contribution of tensor perturbations to the CMB spectra in a modified gravity theory with the properties outlined above and show that this is indeed the case. We treat the graviton mass as a phenomenological parameter. Our results will therefore be valid for any theory in which a modification of gravity amounts to massive gravitational waves with the Λ CDM cosmological background being unchanged.

The chapter is organized as follows. After a summary of the basic equations, we start with an analytic discussion of the CMB spectrum in a massive gravity theory in subsection 4.2.1. After an analytic discussion of the general properties of the B-mode spectrum, we analytically calculate the contribution to the CMB B-type polarization for low multipole coefficients ignoring the effects of reionization and show that the characteristic feature in the low ℓ range is a plateau. For a range of masses above the Hubble rate at the time of recombination this contribution even dominates over the contribution from reionization and is a good approximation to the full numerical spectrum for

low multipole coefficients. We then calculate the contributions to the CMB temperature anisotropy and polarization from a spatially homogeneous tensor mode. The existence of such a contribution is a very unusual property of massive gravity theories. In the massless case, tensor modes are frozen as long as their wavelength is larger than the Hubble scale H^{-1} where $H(t) \equiv a^{-1}(da/dt)$ and $a(t)$ is the scale factor of an isotropic Friedmann-Robertson-Walker (FRW) cosmological model. Since temperature anisotropies get generated only by a time varying tensor mode, this implies that very long wavelengths cannot contribute.² If the graviton is massive, however, when the expansion rate of the universe drops below the graviton mass, these modes acquire an oscillatory time dependence with a frequency set by the graviton mass, and a decreasing amplitude due to Hubble friction. As a consequence, long wavelength modes in a massive gravity theory will generate a temperature anisotropy quadrupole that will get converted into a polarization quadrupole during recombination if the mass is in the right range and more efficiently once the universe becomes reionized. The zero mode does not contribute to the B -mode spectrum, but does contribute to the temperature T , polarization E -mode and TE cross-correlation quadrupoles.

We then proceed to a numerical treatment in subsection 4.2.2. The most interesting result is that a graviton mass would strongly modify the shape of the B -mode spectrum for $\ell < 100$. If such modifications are observed in experiments such as CMBPol, it would provide strong support for massive gravity theories. If on the other hand the observed spectrum is consistent with General Relativity, this would imply an upper bound on the graviton

²The solution of the zero mode equation that does not decay at late times is in fact even a pure gauge mode in the massless case.

mass of $m_g < (10Mpc)^{-1} \approx 10^{-30}eV$. We conclude with a brief summary of the signatures of a graviton mass in the CMB in Section 4.3. This chapter is largely based on [143].

4.1 Tensor Contribution to the CMB for a Massive Graviton

As explained in Chapter 2, the properties of the early universe are encoded in correlations of the temperature anisotropies and polarization patterns at different points in the sky. The quantities most commonly used to represent the two-point correlations are the TT as well as the TE, EE, and BB multipole coefficients. The contribution of the tensor fluctuations to them is given by equations (2.3.178)-(2.3.181). To be explicit, in this chapter we use them in the form

$$C_{BB,\ell}^{(T)} = \pi^2 T_0^2 \int_0^\infty q^2 dq \quad (4.1.1)$$

$$\times \left| \int_{\tau_1}^{\tau_0} d\tau P(\tau) \Psi(q, \tau) \left\{ \left[8\rho + 2\rho^2 \frac{\partial}{\partial \rho} \right] \frac{j_\ell(\rho)}{\rho^2} \right\}_{\rho=q(\tau_0-\tau)} \right|^2,$$

$$C_{EE,\ell}^{(T)} = \pi^2 T_0^2 \int_0^\infty q^2 dq \quad (4.1.2)$$

$$\times \left| \int_{\tau_1}^{\tau_0} d\tau P(\tau) \Psi(q, \tau) \left\{ \left[12 + 8\rho \frac{\partial}{\partial \rho} - \rho^2 + \rho^2 \frac{\partial^2}{\partial \rho^2} \right] \frac{j_\ell(\rho)}{\rho^2} \right\}_{\rho=q(\tau_0-\tau)} \right|^2,$$

$$C_{TE,\ell}^{(T)} = -2\pi^2 T_0^2 \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \int_0^\infty q^2 dq \quad (4.1.3)$$

$$\times \int_{\tau_1}^{\tau_0} d\tau P(\tau) \Psi(q, \tau) \left\{ \left[12 + 8\rho \frac{\partial}{\partial \rho} - \rho^2 + \rho^2 \frac{\partial^2}{\partial \rho^2} \right] \frac{j_\ell(\rho)}{\rho^2} \right\}_{\rho=q(\tau_0-\tau)}$$

$$\begin{aligned}
& \times \int_{\tau_1}^{\tau_0} d\tau' d(q, \tau') \left\{ \frac{j_\ell(q(\tau_0 - \tau'))}{q^2(\tau_0 - \tau')^2} \right\}, \\
C_{TT,\ell}^{(T)} &= \frac{4\pi^2(\ell+2)!T_0^2}{(\ell-2)!} \int_0^\infty q^2 dq \left| \int_{\tau_1}^{\tau_0} d\tau d(q, \tau) \frac{j_\ell(q(\tau_0 - \tau))}{q^2(\tau_0 - \tau)^2} \right|^2. \quad (4.1.4)
\end{aligned}$$

These formulas are equivalent to those of Zaldarriaga and Seljak [82] up to integration by parts.³ While this makes no difference in the massless case, in the massive case the formulas as written here are better suited for numerical calculations.

In these equations, $q = pa(t)$ is the comoving momentum; $\tau = \int dt/a(t)$ is the conformal time; $T_0 = 2.725 K$ is the microwave background temperature at the present conformal time τ_0 ; $P(\tau) = \dot{\kappa} \exp[-\int_\tau^{\tau_0} \dot{\kappa}(\tau') d\tau']$ is the probability distribution of last scattering (or visibility function), with $\dot{\kappa}(\tau)$ the photon collision frequency (or differential optical depth); τ_1 is any time taken early enough before recombination so that any photon present at τ_1 would have collided many times before the present; and $\Psi(q, \tau)$ is the source function, which is customarily calculated from a hierarchy of equations for partial-wave amplitudes (2.3.97) and (2.3.98) derived originally in [46], [124]. With conformal time as independent variable, it takes the form (The dot denotes the derivative with respect to τ)

$$\begin{aligned}
& \dot{\tilde{\Delta}}_{T,\ell}^{(T)}(q, \tau) + \frac{q}{(2\ell+1)} \left((\ell+1) \tilde{\Delta}_{T,\ell+1}^{(T)}(q, \tau) - \ell \tilde{\Delta}_{T,\ell-1}^{(T)}(q, \tau) \right) \\
& = \left(-2\dot{\mathcal{D}}_q(\tau) + \dot{\kappa}(\tau) \Psi(q, \tau) \right) \delta_{\ell,0} - \dot{\kappa}(\tau) \tilde{\Delta}_{T,\ell}^{(T)}(q, \tau), \quad (4.1.5)
\end{aligned}$$

³We also use slightly different conventions. Their gravitational wave amplitude h and power spectral function $P_h(k)$ are related to our gravitational wave amplitude $\mathcal{D}_q(\tau)$ by $h\sqrt{P_h} = \mathcal{D}/2$. In consequence, their function $\Psi\sqrt{P_h}$ is 1/4 times our source function Ψ .

$$\begin{aligned}
& \dot{\tilde{\Delta}}_{P,\ell}^{(T)}(q, \tau) + \frac{q}{(2\ell + 1)} \left((\ell + 1) \tilde{\Delta}_{P,\ell+1}^{(T)}(q, \tau) - \ell \tilde{\Delta}_{P,\ell-1}^{(T)}(q, \tau) \right) \\
& = -\dot{\kappa}(\tau) \Psi(q, \tau) \delta_{\ell,0} - \dot{\kappa}(\tau) \tilde{\Delta}_{P,\ell}^{(T)}(q, \tau) ,
\end{aligned} \tag{4.1.6}$$

with

$$\begin{aligned}
\Psi(q, \tau) &= \frac{1}{10} \tilde{\Delta}_{T,0}^{(T)}(q, \tau) + \frac{1}{7} \tilde{\Delta}_{T,2}^{(T)}(q, \tau) + \frac{3}{70} \tilde{\Delta}_{T,4}^{(T)}(q, \tau) - \frac{3}{5} \tilde{\Delta}_{P,0}^{(T)}(q, \tau) \\
&+ \frac{6}{7} \tilde{\Delta}_{P,2}^{(T)}(q, \tau) - \frac{3}{70} \tilde{\Delta}_{P,4}^{(T)}(q, \tau) .
\end{aligned} \tag{4.1.7}$$

Alternatively, the source function can be calculated from the integral equation (2.3.103) derived in [75],[76]. We have used both approaches.

The quantity $d(q, \tau)$ is given by

$$d(q, \tau) \equiv \exp \left[- \int_{\tau}^{\tau_0} d\tau' \dot{\kappa}(\tau') \right] \left(\dot{\mathcal{D}}_q(\tau) - \frac{1}{2} \dot{\kappa}(\tau) \Psi(q, \tau) \right) , \tag{4.1.8}$$

where $\mathcal{D}_q(t)$ is the gravitational wave amplitude defined through equations (2.3.12) and (2.3.44).

In massive gravity, the evolution of the gravitational wave amplitude is described by the solution to the equation for a minimally coupled *massive* scalar field [138], [139]⁴

$$\ddot{\mathcal{D}}_q(\tau) + 2 \frac{\dot{a}}{a} \dot{\mathcal{D}}_q(\tau) + (q^2 + m_g^2 a^2) \mathcal{D}_q(\tau) = 0 . \tag{4.1.9}$$

In conventional cosmological perturbation theory, the solution of Eq. (4.1.9) which remains finite for $a \rightarrow 0$ and has a wavelength larger than the Hubble scale H^{-1} is not observable locally and does not contribute to the CMB anisotropy and polarization spectra. On the other hand, the other, decaying

⁴In writing this equation, we drop contributions on the right hand side due to anisotropic stress generated by neutrinos and photons [144], [84]. This is done merely for simplicity and we will include these effects in our calculations.

graviton mode produces locally measurable effects even in the $k \rightarrow 0$ limit. Since it does decay, it is usually assumed to be negligible by the time of recombination.

In the massive gravity case, Eq. (4.1.9) implies that a homogeneous metric perturbation starts to oscillate with a frequency equal to m_g when the expansion rate H drops below the graviton mass m_g . This is completely analogous to what happens for light scalar fields (*e.g.* axion, moduli, ...). An important difference, however, is that in the case of the graviton these oscillations may directly affect the CMB spectra (or, if the mass m_g is high enough, may be observed by gravitational wave detectors), similar to a super-Hubble decaying mode or a generic sub-Hubble tensor perturbation in the massless limit. Indeed, the presence of a zero mode implies that, superimposed upon a conventional Hubble expansion, spatial metric components experience anisotropic (but homogeneous) high frequency oscillations with a small amplitude. The effect of such oscillations on the CMB spectra can easily be understood analytically and we will return to this at the end of subsection 2.1.

4.1.1 Analytic results for low multipole coefficients

Before presenting the results of the numerical calculations in the next subsection, let us start with a brief analytic discussion of the spectrum. For the most part, we will limit ourselves to the contribution to the spectrum generated during recombination and ignore the effects of reionization. In the massless case, the effects of reionization give the dominant contribution to the spectrum for $\ell \lesssim 20$ but leave the higher multipole coefficients unchanged (or rather change them trivially by an overall rescaling by $e^{-2\tau_{\text{reion}}}$, where $\tau_{\text{reion}} = 0.087 \pm 0.017$ is the optical depth of the medium due to reionization

and despite the clash of notation should not be confused with conformal time), so that ignoring the effects of reionization is good as long as one is interested in $\ell > 20$. In the massive case, our numerical results indicate that for a range of masses above the Hubble rate at recombination the contribution from recombination provides a good approximation to the spectrum even at low ℓ providing additional motivation for this simplifying assumption.

We will focus on the B-mode spectrum because it is the most interesting one from an experimental point of view. The discussion could be straightforwardly extended to include the TT, TE, and EE spectra, but we limit ourselves to the contribution of the zero mode to those.

Depending on their comoving momentum, the modes fall into one of two classes or one of three classes depending on whether the mass is smaller than the Hubble rate at recombination or larger than that.

For masses below the Hubble rate at recombination, the first possibility is that modes are relativistic at the time they enter the horizon. In this case they will still be relativistic during recombination. These modes are essentially unaffected by the graviton mass, and the spectrum for the values of ℓ these modes contribute to is expected to agree with the one in the massless case. The second possibility is that the modes enter the horizon when they are already non-relativistic. The multipole coefficients these modes contribute to will be different from the ones for the massless case and we will discuss those in more detail below.

For masses above the Hubble rate at recombination there is a third option. The modes can be relativistic as they enter the horizon but become non-relativistic by the time of recombination.

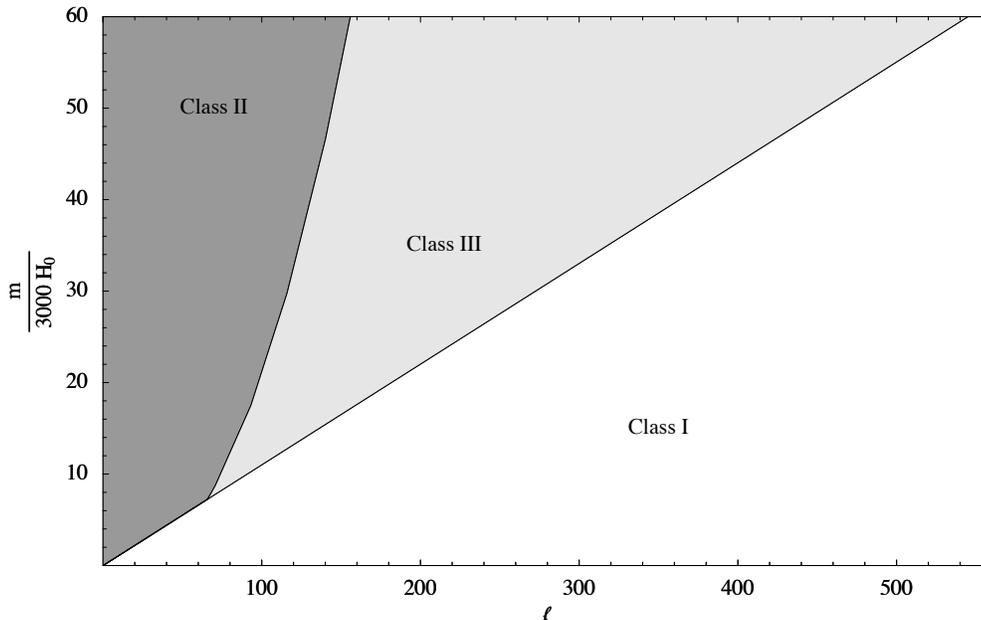


Figure 4.1: This plot summarizes the different behaviors of modes and which range of multipole coefficients they contribute to. Class I corresponds to modes that are relativistic at recombination. Class II corresponds to modes that are non-relativistic as they enter the horizon and during their subsequent evolution. Class III corresponds to modes that enter the horizon when they are relativistic but become non-relativistic before recombination.

The results are summarized in terms of the range of multipole coefficients the different classes affect for a given mass in Figure 4.1.

Let us now discuss these regimes in more detail. Both for masses below and above the expansion rate during recombination, the short-wavelength modes that are in the relativistic regime during recombination,

$$\frac{q}{a(\tau_r)} > m_g , \quad (4.1.10)$$

are not affected by the graviton mass and will lead to the same spectrum as in the massless case. In terms of the multipole number $\ell \approx q(\tau_0 - \tau_{rec})$ this

transition to the massless regime corresponds to

$$\begin{aligned} \ell &\gg \ell_0 \equiv m_g a(\tau_r)(\tau_0 - \tau_r) \\ &\approx \frac{m_g}{H_0} (1 + z_r)^{-1} \int_{(1+z_r)^{-1}}^1 \frac{dx}{\sqrt{\Omega_\Lambda x^4 + \Omega_m x + \Omega_r}} \approx 3.3 (1 + z_r)^{-1} \frac{m_g}{H_0}, \end{aligned} \quad (4.1.11)$$

where $z_r \approx 1088$ is the redshift at recombination, and we have used the five-year WMAP values for the cosmological parameters [78]. In particular, equation (4.1.11) implies that, ignoring the contribution generated during reionization, the B -mode spectrum is not modified for masses smaller than $\sim 300H_0$, which is the scale corresponding to the size of the visible patch of the Universe during recombination.

For larger masses, but still smaller than the expansion rate at recombination, $m_g < H(\tau_r) \approx 2 \times 10^4 H_0$, the modes that are affected by the graviton mass are superhorizon during recombination because they satisfy $\frac{q}{a(\tau_r)} \lesssim m_g < H(\tau_r)$. As a consequence these modes do not oscillate during recombination. Nevertheless, just like in the massless case, the corresponding source term $\dot{\mathcal{D}}_q$ in (4.1.5) is non-zero (though small) and some amount of polarization is being generated. In contrast to the massless case, however, where the value of $\dot{\mathcal{D}}_q$ at recombination is determined by the value of q alone, and goes to zero as $q \rightarrow 0$, in the massive case $\dot{\mathcal{D}}_q$ depends on $(q^2 + m_g^2 a(\tau_r)^2)$ and is independent of q for long wavelength modes leading to an enhancement of the spectrum for $\ell < \ell_0$.

For values of the graviton mass larger than the Hubble rate at recombination, $m_g \gtrsim H(\tau_r)$, all modes start to oscillate before recombination. The modes with long wavelengths start to oscillate as soon as the expansion rate of the universe drops below the graviton mass, *i.e.* at a time $\tau_m < \tau_r$, such

that

$$H(\tau_m) = m_g . \quad (4.1.12)$$

In particular,

$$\mathcal{D}_q \simeq \mathcal{D}_{q_0} \frac{\sin m_g t}{m_g t} \quad (4.1.13)$$

for $(q/a)^2 \ll m_g H$ at the matter dominated stage, where $t \propto \tau^3$. Shorter modes start to oscillate when they enter the horizon just like in the massless case. The transition between these two regimes happens at $q_m = m_g a(\tau_m)$.

To a good approximation, all modes with momenta smaller than q_m have the same evolution—they are frozen until τ_m , and oscillate afterwards with a frequency set by the mass. This value of comoving momentum, q_m , corresponding to the transition between class II and III translates to a value in ℓ -space of

$$\ell_m = 3.3(1 + z_m)^{-1} \frac{m_g}{H_0} . \quad (4.1.14)$$

Here z_m is the redshift corresponding to τ_m and is determined by condition (4.1.12) which can be written more explicitly as

$$H_0 \sqrt{\Omega_m(1 + z_m)^3 + \Omega_r(1 + z_m)^4} = m_g . \quad (4.1.15)$$

At $m_g = H(\tau_r)$ the multipole number ℓ_m coincides with ℓ_0 and takes a value of around $\ell_m \sim 65$. At higher masses these two scale are different. According to (4.1.11) ℓ_0 grows linearly with mass. On the other hand, ℓ_m grows more slowly, $\ell_m \propto m_g^{1/3}$ for masses that become relevant during matter domination (*i.e.*, for $m_g \lesssim 1.5 \times 10^5 H_0$), and $\ell_m \propto m_g^{1/2}$ at higher m_g . For masses much larger than this, all modes oscillate rapidly during recombination. As a consequence the polarization signal gets averaged out and becomes strongly suppressed.

Modes corresponding to angular scales between ℓ_m and ℓ_0 enter the horizon when they are still relativistic, but become non-relativistic before recombination. As a result, they are still expected to exhibit the conventional oscillation pattern in the angular spectrum, but the phase of oscillations is different because the oscillations at late times are driven by the mass rather than the spatial momentum.

After this discussion of various regimes, let us take a more detailed look at the spectrum. As discussed, for the modes referred to as class I in Figure 4.1 that are relativistic during recombination the spectrum to a good approximation agrees with the one in the massless case. While a number of analytic results for the temperature anisotropy and polarization have been found for this case, we will not review those here and refer the interested reader to the literature [54, 145], [126], [122].

As can be seen by inspection of equation (4.1.9), for the modes referred to as class II in Figure 4.1 the dependence of the gravitational wave amplitude on comoving momentum is trivially given by that of the power spectrum because they are frozen as long as they are outside the horizon and the mass already dominates by the time they enter. This does not in general guarantee that the same is true for the source function as it will generically develop its own q -dependence. To a good approximation the momentum dependence of the source function during recombination is the same as that of the gravitational wave amplitude provided the comoving momentum of the mode is less than the duration of recombination in conformal time, *i.e.* $q\Delta\tau_{\text{rec}} \ll 1$. This is satisfied for all modes in class II for the range of masses we are interested in and hence does not provide an additional constraint.

For modes in class II, the dependence of the gravitational wave ampli-

tude and that of the source function on comoving momentum are then trivially given by that of the power spectrum, implying *e.g.* for a standard inflationary scenario that $\Psi(q, \tau)q^{\frac{3}{2} - \frac{n_T}{2}}$ is q -independent. This allows us to evaluate the expression for $C_{BB, \ell}^{(T)}$ given by equation (4.1.1) analytically. Conventionally, one first evaluates the integrals over conformal time and then integrates over momentum. For us it will be more convenient to perform the integral over momentum *first*. This is possible by rewriting the square of the integral over time as an integral in a plane and using the identity:

$$\left(8\rho + 2\rho^2 \frac{\partial}{\partial \rho}\right) \frac{j_\ell(\rho)}{\rho^2} = \frac{\sqrt{2\pi}}{\rho^{\frac{3}{2}}} \left((2 + \ell)J_{\ell+\frac{1}{2}}(\rho) - \rho J_{\ell+\frac{3}{2}}(\rho)\right), \quad (4.1.16)$$

where $J_\nu(\rho)$ is the Bessel function of the first kind. The resulting four integrals over q can then be done exactly using an integral known as the Weber-Schafheitlin integral (see *e.g.* [146]). Dropping terms of order

$$\left\langle \frac{(\tau - \tau_L)^2}{\tau_L^2} \right\rangle \equiv \int_{\tau_1}^{\tau_0} d\tau P(\tau) \Psi(q, \tau) q^{\frac{3}{2} - \frac{n_T}{2}} \frac{(\tau - \tau_L)^2}{\tau_L^2}, \quad (4.1.17)$$

and higher in the terms in the integrals over conformal time arising from the integral over comoving momentum, and setting $n_T = 0$ for simplicity, one finds the following expression for the power spectrum:

$$\frac{\ell(\ell + 1)}{2\pi} C_{BB, \ell}^{(T)} = \frac{2(\ell(\ell + 1) + 16)}{3(\ell + 2)(\ell - 1)} \mathcal{I}^2. \quad (4.1.18)$$

The ℓ -dependence is now explicit and it is easy to see that for $\ell \gtrsim 10$ this becomes independent of ℓ . The ℓ -independent quantity \mathcal{I} is defined as⁵

$$\mathcal{I} = \sqrt{\frac{\pi}{2}} T_0 \int_{\tau_1}^{\tau_0} d\tau P(\tau) \Psi(q, \tau) q^{3/2}, \quad (4.1.19)$$

⁵Recall that $T_0 = 2.725 K$ is the CMB temperature at the present time.

and it encodes the dependence of the spectrum on the mass through the dependence of the source function on the mass. The quantity \mathcal{I}^2 as a function of mass is shown in Figure 4.2 for a scalar amplitude of $\Delta_{\mathcal{R}}^2 = 2.41 \times 10^{-9}$, a tensor-to-scalar ratio $r = 1$.

The oscillatory features seen in the plot can be understood from the fact that the tensor perturbations of the metric take the form given in equation (4.1.13). In particular, they are regular in the limit $t \rightarrow 0$ and the “decaying” mode, which diverges in this limit, is absent. In turn, this property (which also takes place for larger values of q , $(q/a)^2 \geq m_g H$) is a consequence of local isotropy of the Universe at very early times. We assume the latter to be produced by inflation and use the inflationary prediction for the primordial power spectrum, but the existence of the oscillations in \mathcal{I} , as well as the existence of oscillations in the multipole power spectra of polarization and temperature anisotropy seen in Figures 4.3 and 4.4 below (“primordial peaks”, similar to the well known acoustics peaks produced by scalar perturbations but with approximately twice less asymptotic period in ℓ , $T_\ell = \pi(\tau_0 - \tau_r)/\tau_r \approx 140$ [147]), is a more general phenomenon not depending on how this early time isotropy was achieved.

As we will derive shortly, what enters into the source function in a crucial way is the tensor perturbation of the metric evaluated at the time of recombination. This quantity viewed as a function of the graviton mass oscillates around zero, implying that the integral does, too. After squaring, this will give rise just to what is seen in Figure 4.2.

Ignoring terms higher order in the quantity (4.1.17) is typically only a good approximation for $\ell \lesssim 30$, but our numerical calculations show that the plateau persists to higher values of ℓ . For the values of masses where \mathcal{I} is close

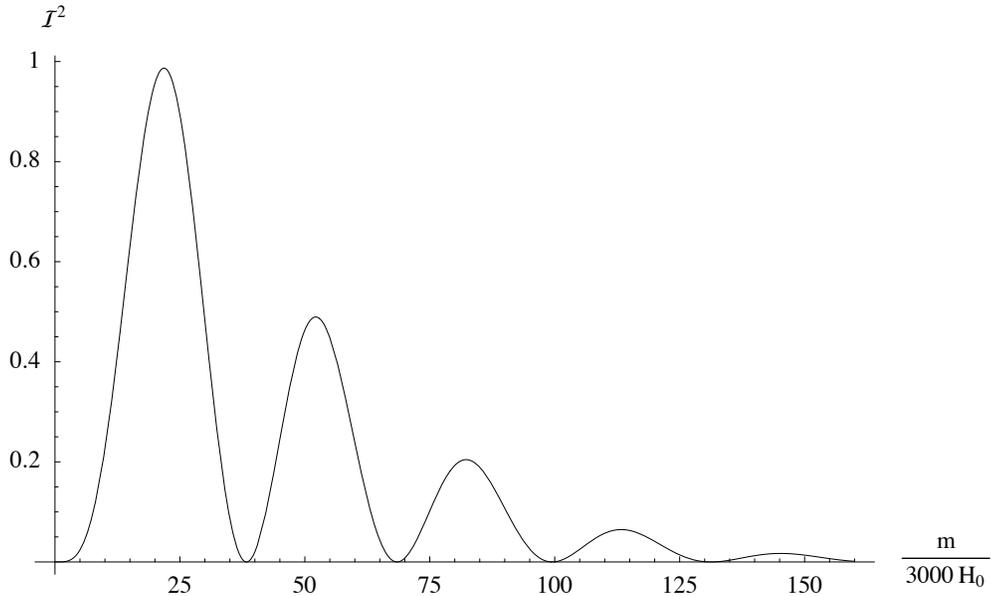


Figure 4.2: This plot shows the quantity \mathcal{I}^2 in $(\mu K)^2$ as a function of mass for a scalar amplitude of $\Delta_{\mathcal{R}}^2 = 2.41 \times 10^{-9}$, a tensor-to-scalar ratio of $r = 1$, and a tensor spectral index $n_T = 0$.

to zero, the higher order terms in this expansion give the leading contribution and have to be included even for low ℓ . Furthermore, for these ranges of masses the effects of reionization become important. Especially for masses below the expansion rate at recombination, reionization is important as it will increase the sensitivity of polarization measurements for masses near the low end of the accessible mass range $m_g \sim 300H_0$ by about one order of magnitude to $m_g \sim 30H_0$.

After discussing the B-mode signal in a massive gravity theory, let us briefly discuss another unusual feature in these theories, a contribution of modes of extremely long wavelength to the TT, EE, and TE quadrupoles. In the limit of vanishing momentum, the Boltzmann hierarchy, *i.e.* equa-

tions (4.1.5), (4.1.6) become very simple. Only $\tilde{\Delta}_{T,0}^{(T)}(q, \tau)$ and $\tilde{\Delta}_{P,0}^{(T)}(q, \tau)$ get generated while all others remain zero. Similar equations were considered in Ref. [44], where CMB polarization and anisotropy in the Kasner universe were studied. To calculate the contribution to the CMB temperature anisotropy and polarization in this limit, it is convenient to write equations (4.1.5), (4.1.6) as

$$\frac{d}{d\tau} \left(e^{-\kappa(\tau)} \tilde{\Delta}_{T,0}^{(T)}(\tau) \right) = -2d(\tau), \quad (4.1.20)$$

$$\frac{d}{d\tau} \left(e^{-\kappa(\tau)} \tilde{\Delta}_{P,0}^{(T)}(\tau) \right) = -P(\tau)\Psi(\tau), \quad (4.1.21)$$

where we have defined the integral optical depth $\kappa(\tau)$ as

$$\kappa(\tau) = \int_{\tau}^{\tau_0} d\tau' \dot{\kappa}(\tau').$$

From equations (4.1.2), (4.1.3), and (4.1.4) we see that the contribution of the zero mode to the TT, EE, and TE quadrupoles is thus given by

$$C_{TT,2}^{(T)} = \frac{2\pi}{75} T_0^2 \tilde{\Delta}_{T,0}^{(T)}(\tau_0)^2, \quad (4.1.22)$$

$$C_{EE,2}^{(T)} = \frac{4\pi}{25} T_0^2 \tilde{\Delta}_{P,0}^{(T)}(\tau_0)^2, \quad (4.1.23)$$

and

$$C_{TE,2}^{(T)} = -\frac{2\pi}{25} \sqrt{\frac{2}{3}} T_0^2 \tilde{\Delta}_{T,0}^{(T)}(\tau_0) \tilde{\Delta}_{P,0}^{(T)}(\tau_0). \quad (4.1.24)$$

It is straightforward to solve equations (4.1.5) and (4.1.6) for $\tilde{\Delta}_{T,0}^{(T)}(\tau_0)$ and $\tilde{\Delta}_{P,0}^{(T)}(\tau_0)$ in this simple case. The result is

$$\tilde{\Delta}_{T,0}^{(T)}(\tau_0) = -\frac{6I_1}{7} - \frac{I_2}{7}, \quad (4.1.25)$$

$$\tilde{\Delta}_{P,0}^{(T)}(\tau_0) = -\frac{I_1}{7} + \frac{I_2}{7}, \quad (4.1.26)$$

where I_1 and I_2 are the following integrals

$$I_1 = 2 \int_0^{\tau_0} d\tau e^{-\kappa(\tau)} \dot{\mathcal{D}}(\tau) , \quad (4.1.27)$$

$$I_2 = 2 \int_0^{\tau_0} d\tau e^{-\frac{3}{10}\kappa(\tau)} \dot{\mathcal{D}}(\tau) . \quad (4.1.28)$$

It is convenient to discuss the behavior of the integrals (4.1.27), (4.1.28) first neglecting the contribution from the reionization epoch. In the absence of reionization, the functions $e^{-\kappa(\tau)}$, $e^{-\frac{3}{10}\kappa(\tau)}$ have a step-like shape and change their values from 0 to 1 at the time of recombination, $\tau = \tau_r$. Let $a(\tau_r)\Delta\tau \sim 35$ kpc be the characteristic width of these step functions (or the duration of the recombination epoch). The relevant parameter which determines the behavior of integrals (4.1.27), (4.1.28) is then

$$\delta = m_g a(\tau_r) \Delta\tau .$$

For small masses, such that $\delta \ll 1$ (but, of course, assuming $m_g > H_0$) one has

$$I_1 = I_2 = -2\mathcal{D}(\tau_r) . \quad (4.1.29)$$

Consequently, in this case polarization is negligible, while $\tilde{\Delta}_{T,0}^{(T)}(\tau_0) = 2\mathcal{D}(\tau_r)$ gives the temperature anisotropy quadrupole. Note that for a fixed initial amplitude of the metric perturbation, the anisotropy is smaller for larger m_g . For large masses, $\delta \gg 1$, the mode oscillates rapidly even during recombination. As a result, both integrals I_1 and I_2 are very small ($\propto e^{-\delta}$) and both temperature and polarization quadrupoles are negligible.

The largest amount of polarization is generated for $\delta \sim 1$. In this case recombination cannot be treated as instantaneous, so that there is no cancellation between the two terms in the expression for the polarization $\tilde{\Delta}_{P,0}^{(T)}(\tau_0)$.

On the other hand, metric oscillations during recombination do not wash out the whole effect yet, and one gets comparable contributions to polarization and temperature anisotropy.

Finally, let us include the effect of reionization. In the presence of reionization Eq. (4.1.29) does not hold even for small masses. Instead, one has⁶

$$I_1 = -2e^{-\tau_{\text{reion}}}\mathcal{D}(\tau_r) , \quad (4.1.30)$$

$$I_2 = -2e^{-\frac{3}{10}\tau_{\text{reion}}}\mathcal{D}(\tau_r) , \quad (4.1.31)$$

$$(4.1.32)$$

where τ_{reion} is again the optical depth of the medium due to reionization. As a result, both integrals get somewhat suppressed. On the other hand, the two terms in Eq. (4.1.26) no longer cancel, and the contributions to polarization and temperature anisotropy can be of the same order.

4.1.2 Numerical results

To calculate the angular power spectra for values of $\ell > 50$, we use CMBfast [82]⁷ with the evolution equation for the tensor perturbations modified according to Eq. (4.1.9). For the low multipole coefficients, $\ell \leq 50$, we use the CMBfast source function, but perform the line of sight integration in Mathematica using equations (4.1.1), (4.1.2), (4.1.3), and (4.1.4) because the CMBfast results become unreliable for $\ell \leq 50$ at least for large graviton masses.

⁶We assume that the graviton mass is still large enough, $m_g \gg H_0$, so that the graviton oscillates rapidly during the reionization.

⁷Available at <http://www.cfa.harvard.edu/mzaldarr/CMBFAST/cmbfast.html>

The issue arises because of rapid oscillations of the source function for all values of comoving momentum. After the integration by parts as implemented in CMBfast, eq. (29) of [74] involves second derivatives of the source function, which are unpleasant to deal with numerically. As a result, the line of sight integration, as implemented in CMBfast, produces unreliable results. The problem exists for the BB spectrum as well but is especially severe for the low- ℓ parts of the EE and TE spectra, where it appears at masses of order $m_g \sim 3000H_0$.

Using independent Mathematica code, we also checked that the source function as produced by the modified CMBfast is reliable. We assume a scale-invariant power spectrum for the tensor perturbations, $n_T = 0$. The results for a range of masses are shown in Figure 4.3.

In all the plots we drop the quadrupole, because its value depends on the IR cutoff at low momenta, as follows from the discussion in section 4.1.1 (see also section 4.2 for more details). We have used $\Delta_{\mathcal{R}}^2 = 2.41 \times 10^{-9}$ and have set the tensor-to-scalar ratio, r , to unity.

In agreement with our estimates, the effect of the mass is rather mild for masses much below the Hubble rate during recombination and is present only for very low ℓ .

For masses approaching the Hubble rate during recombination, the spectrum is significantly modified up to $\ell \sim 100$, and at $m_g = 3 \times 10^4 H_0$ the characteristic plateau at $\ell \lesssim 100$ is fully developed. As we increase the mass further, the height of the plateau increases up to values of $\mu \equiv \frac{m_g}{3000H_0} \approx 25$ but starts to decrease beyond that in agreement with the oscillations we saw in our semi-analytic result in Figure 4.2. The origin of the oscillations is that depending on the mass the metric perturbation enters recombination in different

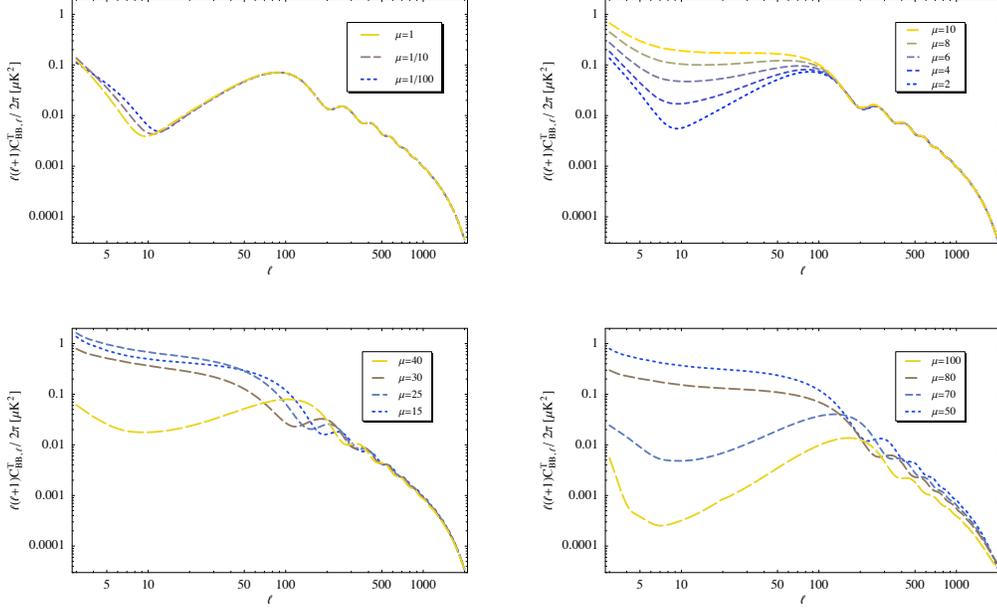


Figure 4.3: These plots show the $C_{BB,\ell}^{(T)}$ multipole coefficients for the range of masses that lead to the most interesting signal in the CMB. The masses are given by $m_g = \mu \times 3000H_0$, where μ is given in the legend. Longer dashes correspond to larger mass. All plots are for a scalar amplitude $\Delta_{\mathcal{R}}^2 = 2.41 \times 10^{-9}$, a tensor-to-scalar ratio, $r = 1$, and a tensor spectral index $n_T = 0$. For the remaining cosmological parameters parameterizing the background, we use the five-year WMAP values [78].

phase.

In agreement with our qualitative arguments summarized in Figure 4.1, for masses $\mu \gtrsim 10$ we see that a transition region appears between the multipole moment ℓ_m where the plateau ends and the multipole moment ℓ_0 where the massless spectrum is approached.

We are not showing the spectra at higher masses, because the polarization signal becomes strongly suppressed for $\mu \gg 150$ because of rapid oscillation.

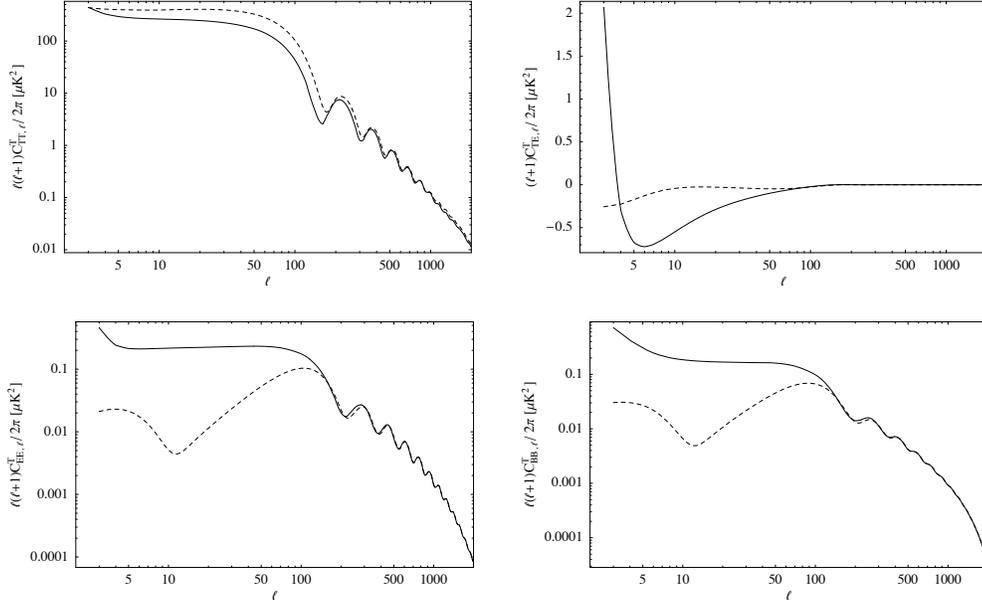


Figure 4.4: This plot shows T (upper left panel), E (lower left), TE (upper right) and B (lower right) spectra for the massive case with $\mu = 10$ (solid line) and for the massless case (dashed line).

tions of the metric during recombination in agreement with Figure 4.2.

So far we have only shown the $C_{BB,\ell}^{(T)}$ multipole coefficients as they are the most interesting from a phenomenological point of view. In Figure 4.4, we show a comparison of all four CMB spectra (temperature anisotropy, E - and B -type polarization and TE cross-correlations) for $\mu = 10$ and the massless case.

We see that the mass affects both E - and B -type polarization in a rather similar way. The effect of the mass on the temperature anisotropy is rather mild and the shape of the spectrum for massive gravity is very similar to the massless case. Unlike polarization, the temperature anisotropy receives

contributions not only from recombination and reionization, but from all times. As a result, the contribution at horizon crossing dominates and one obtains a plateau reflecting the flatness of the primordial spectrum in both the massive and the massless case (see [145] for the analytic expression describing this plateau in the massless case which, as is seen from the upper left plot in Figure 4.4, produces a good approximation to the massive case, too). While the polarization spectra look rather different from the ones in the massless case, one should have in mind that we only show the tensor contribution to the signal. The main component for the temperature anisotropy as well as the TE and EE spectra comes from the scalar perturbations which are identical to the ones in general relativity. One may still wonder how large a tensor signal one could tolerate in the massive case given the existing data and what has already been ruled out. To this end, we perform a Markov chain Monte Carlo study for a single value of mass $m_g = 3 \times 10^4 H_0$, or equivalently $\mu = 10$.

We use the publicly available CosmoMC code [148]⁸ to sample the parameter space together with CAMB [83]⁹ to generate the spectra for a given set of cosmological parameters, and we use a modified version of the WMAP likelihood code that is now available on the LAMBDA website¹⁰ to evaluate the likelihood function for a given spectrum.

In addition to varying the six parameters of the Λ CDM model and marginalizing over the Sunyaev-Zeldovich amplitude, we allow the tensor-to-scalar ratio to vary but keep the mass and the tensor spectral index fixed. We do not implement the slow-roll consistency condition but set $n_T = 0$ as one

⁸CosmoMC is available online at <http://cosmologist.info/cosmomc/>

⁹The code is available online at <http://camb.info/>

¹⁰http://lambda.gsfc.nasa.gov/product/map/current/likelihood_get.cfm

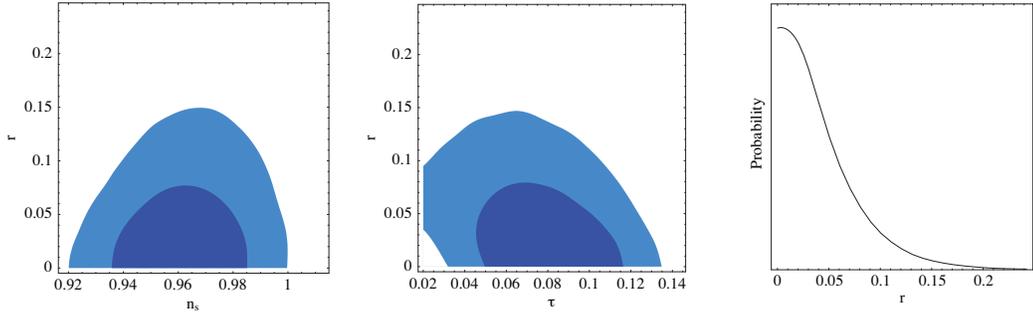


Figure 4.5: These plots show marginalized likelihood plots obtained from a Markov chain Monte Carlo study of a massive gravity model with a mass $m_g = 3 \times 10^4 H_0$, or equivalently $\mu = 10$ using the five-year WMAP data. The dark and light blue contours correspond to 68% and 95% confidence level, respectively.

may expect the consistency condition to be modified in these theories, but this does not significantly change the results.

We find that the tensor-to-scalar ratio for this value of mass is constrained to $r < 0.11$ at 95% confidence level by the five-year WMAP data alone. The results are shown in Figure 4.5.

Different from the massless case [78], where adding additional data sets like BBN, supernovae, or baryon acoustic oscillations significantly lowers the allowed tensor-to-scalar ratio, the results do not change much for the massive case. In the massless case, the reason for the significant strengthening of the bound is that additional data sets constraining the baryon or dark matter abundance break a chain of degeneracies. The main constraint on the tensor-to-scalar signal in the massless case currently comes from the low- ℓ TT spectrum, where the tensor signal has a plateau as can be seen *e.g* in Figure 4.4. Raising n_s lowers the power in the scalar spectrum at low ℓ thus allowing for larger r . On the other hand raising n_s becomes possible only because of a

degeneracy between Ω_B and n_S so that adding the BBN priors on the baryon abundance eliminates this possibility and substantially lowers the bound on the tensor-to-scalar ratio in the massless case. In the massive case, however, the polarization data is starting to constrain the model, eliminating the degeneracy in a different way. Including a BBN prior then does not significantly lower the bound in this case and all in all the bound is roughly a factor of two stronger for $m_g = 3 \times 10^4 H_0$ than it is in the massless case. As one might expect from looking at the spectra in the massive case, there is now a degeneracy between the tensor-to-scalar ratio and the optical depth, however. This is shown in Figure 4.5. We interpret these results as telling us that the model is not in conflict with present data but not much more. In a more serious analysis, the mass should certainly not be taken as fixed but be thought of as an unknown parameter that has to be extracted from the data. We leave a more systematic study for when more data becomes available.

4.2 Summary

To summarize, we see that a detection of the CMB B-mode signal either with Planck or with next generation CMB measurements such as CMBPol [117], in addition of opening a new observational window on inflation, will also provide a sensitive probe of the graviton mass. We showed that the most interesting consequence of the graviton mass for the cosmic microwave background is probably the characteristic plateau in the B -mode spectrum for multipoles with $\ell \lesssim 100$. This plateau is most pronounced for masses a few times the Hubble rate at recombination, but in principle CMB polarization measurements are capable of constraining the graviton mass down to $m_g^{-1} \sim 10$ Mpc. Taking into account that large graviton masses $m_g^{-1} \ll 10$ kpc lead to a strong

suppression of the tensor contribution to the CMB spectra, we conclude that the observation of B -mode with the conventional inflationary spectrum would provide by far the tightest bound on the mass of an elementary particle – a graviton – at a level of $m_g \lesssim 10^{-30}$ eV.

Of course, even more exciting would be to find out that a graviton mass is actually non-zero. It is worth stressing that gravitational waves after being produced during inflation remain practically undisturbed throughout the later evolution of the Universe [100] and secondary sources of the B mode, such as the weak lensing contribution, are negligibly small at $\ell < 100$ [149]. It also appears extremely hard to mock up the effect of the mass by modifying an inflationary model or invoking another mechanism generating gravitational waves such as cosmic strings. Consequently, the detection of a B -mode signal with the shape discussed above would provide an unambiguous signal of a graviton mass.

We have also seen that in a massive gravity theory superhorizon tensor perturbations are physical and contribute to the quadrupole of TT, EE, and TE spectra. As a consequence, the amplitude of these quadrupoles are IR sensitive and can be significantly enhanced over the rest of the spectrum. The ratio between the quadrupole and the rest of the spectrum is model dependent. It provides a probe of the total duration of the inflation provided the tensor perturbations are naturally set to zero at early times in a given model. This is the case, for instance, if the graviton mass is not constant during inflation and in the beginning is bigger than the expansion rate. One way this possibility may be realized is if the dilatation symmetry $t \rightarrow \lambda t$, $x^i \rightarrow \lambda^{-\gamma} x^i$ is not present in the full theory, but only gets restored during inflation as one approaches a cosmological attractor discussed in [139]. In this particular scenario, the

quadrupoles measure the number of e-folds during the period for which the graviton is lighter than the Hubble rate.

Chapter 5

Oscillations in the CMB from Axion Monodromy Inflation

In this chapter, we study the CMB observables in axion monodromy inflation. These well-motivated scenarios for inflation in string theory have monomial potentials over super-Planckian field ranges, with superimposed sinusoidal modulations from instanton effects. Such periodic modulations of the potential can drive resonant enhancements of the correlation functions of cosmological perturbations, with characteristic modulations of the amplitude as a function of wavenumber. We give an analytical result for the scalar power spectrum in this class of models, and we determine the limits that present data places on the amplitude and frequency of modulations. Then, incorporating an improved understanding of the realization of axion monodromy inflation in string theory, we perform a careful study of microphysical constraints in this scenario. We find that detectable modulations of the scalar power spectrum are commonplace in well-controlled examples, while resonant contributions to the bispectrum are undetectable in some classes of examples and detectable in others. We conclude that resonant contributions to the spectrum and bispectrum are a characteristic signature of axion monodromy inflation that, in favorable cases, could be detected in near-future experiments.

As explained in section 2.4, inflation is a successful paradigm for describing the early universe, but it is sensitive to the physics of the ultraviolet

completion of gravity. This motivates pursuing realizations of inflation in string theory, a candidate theory of quantum gravity. Considerable progress has been made on this problem in recent years, so much so that the most pressing task, particularly in view of upcoming CMB experiments, is to learn how to distinguish various incarnations of inflation in string theory from each other and from related models constructed directly in quantum field theory.

Fortunately, the additional constraints inherent in realizing inflation in an ultraviolet-complete framework can leave imprints in the low-energy Lagrangian, and hence ultimately in the cosmological observables. In favorable cases, a given class of models may make distinctive predictions for a variety of correlated observables, allowing one to exclude this class of models given adequate data.

One decisive observable for probing inflation is the tensor-to-scalar ratio, r . A promising class of string inflation models producing a detectable tensor signature are those involving *monodromy* [150], in which the potential energy is not periodic under transport around an angular direction in the configuration space. The first examples [150] involved monodromy under transport of a wrapped D-brane in a nilmanifold, and a subsequent class of examples invoked monodromy in the direction of a closed string axion [151].

The axion monodromy inflation scenario of [151] is falsifiable on the basis of its tensor signature, $r \approx 0.07$. However, primordial tensor perturbations have not been detected at present, while the temperature anisotropies arising from scalar perturbations have been mapped in great detail [78]. One could therefore hope to constrain axion monodromy inflation more effectively by understanding the signatures that it produces in the scalar power spectrum and bispectrum. Characterizing these signatures is the subject of this chapter.

As we shall explain, the potential in axion monodromy inflation is approximately linear, but periodically modulated: each circuit of the loop in configuration space can provide a bump on top of the otherwise linear potential. Modulations of the inflaton potential with suitable frequency and amplitude can yield two striking signatures: periodic undulations in the spectrum of the scalar perturbations, and resonant enhancement [152] of the bispectrum. Let us stress that the presence of some degree of modulations of the potential is automatic, and is an example of the situation described above in which traces of ultraviolet physics remain in the low-energy Lagrangian. We do not introduce modulations in order to make the scalar perturbations more interesting. However, it is important to examine the typical amplitude and frequency of modulations in models that are under good microphysical control, in order to ascertain whether well-motivated models produce signatures that can be detected in practice.

To achieve this, we first investigate in detail the realization of axion monodromy inflation in string theory. We compute the axion decay constants in terms of compactification data, we assess the importance of higher-derivative terms, and we estimate the amplitude of modulations for the case of Euclidean D1-brane contributions to the Kähler potential. We also identify a potentially-important contribution to the inflaton potential, arising from backreaction in the compact space, and we present a model-building solution that suppresses this contribution.

We find that detectable modulations of the scalar power spectrum and bispectrum are possible in models that are consistent with all current data and that are under good microphysical control. In fact, we find substantial parameter ranges that are excluded not by microphysics, but by observational

constraints on modulations of the scalar power spectrum.

The organization of this chapter is as follows. We begin with a very brief review of axion monodromy inflation in section 5.1. In section 5.2 we describe the classical evolution of the homogeneous background in axion monodromy inflation with a modulated linear potential. We then solve, in section 5.3, the Mukhanov-Sasaki equation governing the evolution of scalar perturbations, giving an analytical result for the spectrum in terms of the frequency and amplitude of the modulations of the potential. Next, we briefly discuss the bispectrum and express the amplitude of the non-Gaussianity in terms of the model parameters. We then present, in section 5.4, an analysis of the constraints imposed on axion monodromy inflation by the WMAP5 data (for prior work constraining similar oscillatory power spectra, see *e.g.* [153],[154],[155],[156],[157],[158],[159]). In sections 5.5 and 5.6, we present a comprehensive analysis of the constraints imposed by the requirements of computability and of microphysical consistency, including validity of the string loop and α' perturbation expansions, successful moduli stabilization, and bounds on higher-derivative terms. In section 5.7 we combine the observational and theoretical constraints, with results presented in figure 5.7.

The chapter is based on [160].

5.1 Review of axion monodromy inflation

In this section we will briefly review the motivation for axion monodromy inflation, as well as the most salient phenomenological features. We will postpone until section 5.5 a more comprehensive discussion of the realization of this model in string theory.

Inflation is sensitive to Planck-scale physics: contributions to the ef-

fective action arising from integrating out degrees of freedom with masses as large as the Planck scale play a critical role in determining the background evolution, and hence the observable spectrum of perturbations (see [161] for a review of this issue). A central problem in inflationary model-building is establishing knowledge of Planck-suppressed terms in the effective action with accuracy sufficient for making predictions. The most elegant solution to this problem is to provide a symmetry that forbids such Planck-suppressed contributions. Because invoking such a symmetry amounts to forbidding couplings of the inflaton to Planck-scale degrees of freedom, it is important to understand this issue in an ultraviolet-complete theory, such as string theory.

One promising mechanism for inflation in string theory involves the shift symmetry of an axion. Axions are numerous in string compactifications and generally enjoy continuous shift symmetries $a \rightarrow a + \text{constant}$ that are valid to all orders in perturbation theory, but are broken by nonperturbative effects to discrete shifts $a \rightarrow a + 1$. As noted in [151], the shift symmetries of axions descending from two-forms are also broken by suitable space-filling five-branes (D5-branes or NS5-branes) wrapping two-cycles in the compact space.

In axion monodromy inflation [151], an NS5-brane wrapped on a two-cycle Σ breaks the shift symmetry of the Ramond-Ramond two-form potential C_2 , inducing a potential that is asymptotically linear in the corresponding canonically normalized field ϕ ,

$$V = \mu^3 \phi, \tag{5.1.1}$$

with μ a constant mass scale. Inflation begins with a large expectation value for the inflaton, $\phi \propto \int_{\Sigma} C_2 \gg 1$, and proceeds as this expectation value diminishes; note that the NS5-brane, like any D-branes that may be present in the

compactification, remains fixed in place during inflation. As argued in [151], this gives rise to a natural model of inflation, with the residual shift symmetry of the axion protecting the potential from problematic corrections that are endemic in string inflation scenarios.

In this chapter we perform a careful analysis of the consequences of nonperturbative effects for the axion monodromy scenario. Such effects are generically present: specifically, Euclidean D-branes make periodic contributions to the potential in most realizations of axion monodromy inflation. However, the size of these contributions is model-dependent. It was shown in [151] that there exist classes of examples in which nonperturbative effects are practically negligible, but we expect – as explained in detail in section 5.6.5 – that in generic configurations, periodic terms in the potential make small, but not necessarily negligible, contributions to the slow roll parameters.

Therefore, it is of interest to understand the consequences of small periodic modulations of the inflaton potential in axion monodromy inflation. In this chapter we address this question in two ways: first, in subsections 5.2-5.4, by studying a phenomenological potential that captures the essential effects; and second, in section 5.5 and section 5.6, by investigating the ranges of the phenomenological parameters that satisfy all known microphysical consistency requirements dictated by the structure of string compactifications in which axion monodromy inflation can be realized.

In the absence of oscillations, axion monodromy gives rise to a model of large field inflation with a linear potential that is easily studied using the slow roll expansion. The slow roll parameters ϵ_V and η_V are given by

$$\epsilon_V = \frac{1}{2\phi^2} \quad \text{and} \quad \eta_V = 0, \quad (5.1.2)$$

so that the tensor-to-scalar ratio and the scalar spectral index are

$$r = \frac{8}{\phi^2} \quad \text{and} \quad n_s = 1 - \frac{3}{\phi^2}. \quad (5.1.3)$$

Assuming that the pivot scale left the horizon 60 e-foldings before the end of inflation, we should evaluate these at $\phi = \phi_* \simeq 11M_{pl}$. This leads to a tensor-to-scalar ratio $r \simeq 0.07$ and a scalar spectral index $n_s \simeq 0.975$. After determining μ from the observed value of the scalar amplitude, one finds that the energy scale of inflation around the time when the modes we observe in the CMB exit the horizon is given by $V^{1/4} \simeq 7 \times 10^{-3}M_{pl} \simeq 1.7 \times 10^{16}$ GeV. For completeness, the Hubble constant during inflation is then $H \simeq 2.8 \times 10^{-5}M_{pl} \simeq 6.8 \times 10^{13}$ GeV.

5.2 Background Evolution

In this section we will study the background evolution of the inflaton in the presence of small periodic modulations of the potential. We will focus on modulations in axion monodromy inflation with a linear potential, but our derivations are easily modified to account for other models with a modulated potential. We will denote the size of the modulation by Λ^4 , and write our potential as in [151],

$$V(\phi) = \mu^3 \phi + \Lambda^4 \cos\left(\frac{\phi}{f}\right) = \mu^3 \left[\phi + bf \cos\left(\frac{\phi}{f}\right) \right], \quad (5.2.1)$$

where we defined the parameter $b \equiv \frac{\Lambda^4}{\mu^3 f}$. The equation of motion for the inflaton is then¹

$$\ddot{\phi} + 3H\dot{\phi} + \mu^3 - \mu^3 b \sin\left(\frac{\phi}{f}\right) = 0. \quad (5.2.2)$$

To solve (5.2.2), we begin with two approximations. Monotonicity of the potential requires² $b < 1$, and as we will see in section 5.4, for the case $b < 1$ observational constraints in fact imply $b \ll 1$. This suggests treating the oscillatory term in the potential as a perturbation. Furthermore, the COBE normalization implies that $\phi \gg M_p$ during the era when the modes that are observable in the cosmic microwave background exit the horizon. This allows us to drop terms of higher order in M_p/ϕ .

Under these conditions, it is straightforward to solve for the evolution of the homogeneous background. Expanding the field as $\phi = \phi_0 + b\phi_1 + \mathcal{O}(b^2)$, the equations of motion of zeroth and first order in b become

$$\dot{\phi}_0 = -\sqrt{\frac{\mu^3}{3\phi_0}}, \quad (5.2.3)$$

$$\ddot{\phi}_1 + \sqrt{3\mu^3\phi_0}\dot{\phi}_1 - \frac{\mu^3}{2\phi_0}\phi_1 = \mu^3 \sin\left(\frac{\phi_0}{f}\right), \quad (5.2.4)$$

¹In section 2.4, we denoted the background field by $\bar{\phi}$. In this chapter we will always work in a gauge such that $\delta\phi = 0$, so that $\phi = \bar{\phi}$. We will thus drop the bar and use ϕ to denote the background scalar field.

²The case of non-monotonic potentials may also be interesting. On the one hand, for sufficiently large $b > 1$, it may be possible to realize chain inflation [162],[163],[164] in our model. In this scenario, the inflaton would tunnel from minimum to minimum, with the universe expanding by less than one third of an e-fold per tunneling event. This requires a more careful analysis, and we will leave this for future studies. On the other hand, for $b \gtrsim 1$ the model essentially turns into a small-field model of inflation because the inflaton gets trapped at the peaks for a large number of e-folds. It seems hard to distinguish this from other models of small field inflation, but it may be interesting to take a closer look at this as well.

where we have neglected terms of higher order in M_p/ϕ and we have made use of the slow roll approximation for ϕ_0 .³ Using equation (5.2.3), we can rewrite equation (5.2.4) with ϕ_0 as an independent variable instead of t , yielding

$$\phi_1'' - 3\phi_0\phi_1' - \frac{3}{2}\phi_1 = 3\phi_0 \sin\left(\frac{\phi_0}{f}\right). \quad (5.2.5)$$

where primes denote derivatives with respect to ϕ_0 . For the period of interest, in which the modes now visible in the CMB exit the horizon, it is a good approximation to neglect the motion of ϕ_0 everywhere except in the driving term. The inhomogeneous solution is then given by

$$\phi_1(t) = f \frac{6f\phi_*}{(2+3f^2)^2 + 36f^2\phi_*^2} \times \left[-(2+3f^2) \sin\left(\frac{\phi_0(t)}{f}\right) + 6f\phi_* \cos\left(\frac{\phi_0(t)}{f}\right) \right], \quad (5.2.6)$$

where ϕ_* denotes the value of the field ϕ_0 at the time at which the pivot scale k_* exits the horizon. Assuming 60 e-foldings of inflation, this happens around $\phi_* \simeq 11M_{pl}$. For decay constants f obeying $f \gtrsim M_p/10$, there is less than one oscillation in the range of modes that are observable in the cosmic microwave background, leading to an uninteresting modulation with very long wavelength. We will thus make the additional assumption that $f \ll M_p$. Assuming that $\phi_0 \gg M_p$ and $f \ll 1$, using the slow roll approximation for $\phi_0(t)$, and working to first order in b , the solution thus becomes

$$\begin{aligned} \phi(t) &= \phi_0(t) + b\phi_1(t) \\ &= \phi_0(t) + bf \frac{3f\phi_*}{1+(3f\phi_*)^2} \left[-\sin\left(\frac{\phi_0(t)}{f}\right) + 3f\phi_* \cos\left(\frac{\phi_0(t)}{f}\right) \right], \quad (5.2.7) \end{aligned}$$

³In approximating $\sin(\phi/f) \simeq \sin(\phi_0/f)$ on the right hand side of (5.2.4), we have assumed not only that $b \ll 1$ but also that $b\phi_1/f \ll 1$. As we will see from the solution (5.2.7), ϕ_1 is of order $f^2\phi_*$. Hence the mild assumption $bf\phi_* \ll 1$ justifies this approximation.

with $\phi_0(t)$ given by

$$\phi_0(t) = \left[\phi_*^{3/2} - \frac{\sqrt{3}}{2} \mu^{3/2} (t - t_*) \right]^{2/3}. \quad (5.2.8)$$

5.3 Spectrum of Scalar Perturbations

Having understood the background evolution, we are now in a position to calculate the power spectrum in axion monodromy inflation. One might be tempted to do this by brute-force numerical calculation, but we find it more instructive to have an analytic result. We will show that under the same assumptions made in calculating the background evolution, *i.e.* slow roll for $\phi_0(t)$, $\phi_0 \gg M_p$, $f \ll M_p$, and to first order in b , the scalar power spectrum is of the form

$$\begin{aligned} \Delta_{\mathcal{R}}^2(q) &= \Delta_{\mathcal{R}}^2(q_*) \left(\frac{q}{q_*} \right)^{n_s-1} \left[1 + \delta n_s \cos \left(\frac{\phi_q}{f} \right) \right] \\ &\approx \Delta_{\mathcal{R}}^2 \left(\frac{q}{q_*} \right)^{n_s-1 + \frac{\delta n_s}{\ln(q/q_*)} \cos \left(\frac{\phi_q}{f} \right)}, \end{aligned} \quad (5.3.1)$$

where the quantity $\Delta_{\mathcal{R}}^2(q_*)$ parameterizes the strength of the scalar perturbations and will be introduced in detail in the next subsection. The second equality is valid as long as $\delta n_s \ll 1$, and δn_s is given by

$$\delta n_s = \frac{12b}{\sqrt{(1 + (3f\phi_*)^2)}} \sqrt{\frac{\pi}{8} \coth \left(\frac{\pi}{2f\phi_*} \right) f\phi_*}, \quad (5.3.2)$$

where

$$\phi_q = \sqrt{\phi_*^2 - 2 \ln q/q_*} \simeq \phi_* - \frac{\ln q/q_*}{\phi_*} \quad (5.3.3)$$

is the value of the scalar field at the time when the mode with comoving momentum q exits the horizon.

In subsection 5.3.1 we will give a derivation of this result that makes no further approximations. In subsection 5.3.2, we will present another derivations of (5.3.1) that is valid only as long as $f\phi_* \ll 1$ but that leads to a better understanding of the relevant physical effects behind the power spectrum (5.3.1). Let us at this point briefly summarize the scales that will be relevant for our discussion in the next subsections.

Given the potential (5.2.1), the time frequency of the oscillations of the inflaton is $\omega = \dot{\phi}/f$. This is also the time frequency of the oscillations of the background. Perturbations around this background can be quantized in terms of the solutions of the Mukhanov-Sasaki equation, assuming an asymptotic Bunch-Davies vacuum. Every perturbation mode with comoving momentum q oscillates with a time frequency q/a that is redshifted by the expansion of the universe until the mode exits the horizon and freezes when $q = aH$.

Then, if $H < \omega < M_{pl}$, every mode will at a certain time resonate with the background, as stressed by Chen, Easther, and Lim in [152]. Using the slow roll equation of motion and the COBE normalization,

$$3H\dot{\phi} \simeq -V'(\phi), \quad \dot{\phi}^2 \simeq \frac{2}{3}\epsilon V, \quad V \simeq 5 \times 10^{-7} \epsilon M_{pl}^4, \quad (5.3.4)$$

the requirement $H < \omega < M_{pl}$ can be re-expressed as

$$\frac{\omega}{H} \simeq \frac{M_{pl}^2}{\phi f} \simeq \sqrt{2\epsilon} \frac{M_{pl}}{f} > 1, \quad (5.3.5)$$

$$\frac{\omega}{M_{pl}} \simeq \sqrt{\frac{2\epsilon V}{3}} \frac{1}{f M_{pl}} < 1, \quad (5.3.6)$$

hence defining a range of values for the axion decay constant f for which resonances occur. Using $\sqrt{2\epsilon} \simeq M_{pl}/\phi_* \simeq .09$, we obtain $2.4 \times 10^{-6} < \frac{f}{M_{pl}} < 0.09$. We will show in section 5.5 and section 5.6 that f falls in this range in a class of microphysically well-controlled examples.

Going beyond our approximations, the model also predicts a small amount of running of the scalar spectral index, of order 10^{-4} , from terms of higher order in the M_p/ϕ expansion. Furthermore, δn_s develops a very mild momentum dependence. We will neglect these effects because these will most likely not be observable in current or near-future CMB experiments.

5.3.1 Analytic solution of the Mukhanov-Sasaki equation

We begin our study of the spectrum by choosing a gauge such that the scalar field is unperturbed, $\delta\phi(\mathbf{x}, t) = 0$, and the scalar perturbations in the spatial part of the metric take the form

$$\delta g_{ij}(\mathbf{x}, t) = 2a(t)^2 \mathcal{R}(\mathbf{x}, t) \delta_{ij}. \quad (5.3.7)$$

The quantity $\mathcal{R}(\mathbf{x}, t)$ is a gauge-invariant quantity and in the case of single-field inflation is conserved outside the horizon. It is closely related to the scalar curvature of the spatial slices, but we will not need its precise geometric interpretation at this point.

The translational invariance of the background and thus the equations of motion governing the time evolution of the perturbations make it convenient to look for solutions of the linearized Einstein equations in Fourier space. As explained in Chapter 2, one defines

$$\mathcal{R}(\mathbf{x}, t) = \int \frac{d^3q}{(2\pi)^{3/2}} [\mathcal{R}_q(t) e^{i\mathbf{q}\cdot\mathbf{x}} \alpha(\mathbf{q}) + \mathcal{R}_q(t)^* e^{-i\mathbf{q}\cdot\mathbf{x}} \alpha^*(\mathbf{q})], \quad (5.3.8)$$

where \mathbf{q} is the comoving momentum, and q is its magnitude. The rotational invariance of the background ensures that $\mathcal{R}_q(t)$ can depend only on the magnitude of the comoving momentum but not on its direction. Directional dependence can only be contained in the stochastic parameter $\alpha(\mathbf{q})$ that param-

eterizes the initial conditions and is normalized so that

$$\langle \alpha(\mathbf{q})\alpha^*(\mathbf{q}') \rangle = \delta(\mathbf{q} - \mathbf{q}'), \quad (5.3.9)$$

where the average denotes the average over all possible histories. With this ansatz, the Einstein equations turn into an ordinary differential equation, the Mukhanov-Sasaki equation, governing the time evolution of $\mathcal{R}_q(t)$. We will use it in the form

$$\frac{d^2\mathcal{R}_q}{dx^2} - \frac{2(1+2\epsilon+\delta)}{x} \frac{d\mathcal{R}_q}{dx} + \mathcal{R}_q = 0, \quad (5.3.10)$$

where $x \equiv -q\tau$, with the conformal time τ given as usual by $\tau \equiv \int^t \frac{dt'}{a(t')}$. To evaluate \mathcal{R}_q^o , it will again turn out to be sufficient to solve to first order in b .

We therefore expand the slow roll parameters,

$$\epsilon = \epsilon_0 + \epsilon_1 + \mathcal{O}(b^2), \quad (5.3.11)$$

$$\delta = \delta_0 + \delta_1 + \mathcal{O}(b^2). \quad (5.3.12)$$

For the background solution (5.2.7), the first-order terms are given by

$$\epsilon_1 = -\frac{3bf}{\phi_*[1+(3f\phi_*)^2]} \left[\cos\left(\frac{\phi_0}{f}\right) + (3f\phi_*) \sin\left(\frac{\phi_0}{f}\right) \right], \quad (5.3.13)$$

$$\delta_1 = -\frac{3b}{[1+(3f\phi_*)^2]} \left[\sin\left(\frac{\phi_0}{f}\right) - (3f\phi_*) \cos\left(\frac{\phi_0}{f}\right) \right]. \quad (5.3.14)$$

We now consider an ansatz of the form

$$\mathcal{R}_q = \mathcal{R}_{q,0}^{(o)} \left[i\sqrt{\frac{\pi}{2}} x^{\nu_0} H_{\nu_0}^{(1)}(x) + g(x) \right]. \quad (5.3.15)$$

Here the index ν_0 on the Hankel function, $H_{\nu_0}^{(1)}(x)$, is given by $\nu_0 = \frac{3}{2} + 2\epsilon_0 + \delta_0$, $g(x)$ is a perturbation of order b , and $\mathcal{R}_{q,0}^{(o)}$ is the value of $\mathcal{R}_q(t)$ outside the

horizon in the absence of modulations, *i.e.* for $b = 0$. To be explicit, it is given by⁴

$$\mathcal{R}_{q,0}^{(o)} = \mp i \sqrt{\frac{\mu^3 \phi_q^3}{6}} \frac{1}{q^{3/2}}, \quad (5.3.16)$$

where $\phi_q \approx \phi_* - \frac{\ln q/q_*}{\phi_*}$ once again is the value of the scalar field at the time the mode with comoving momentum q exits the horizon. The quantity of interest to first order in b is then

$$|\mathcal{R}_q^{(o)}|^2 = |\mathcal{R}_{q,0}^{(o)}|^2 [1 + 2 \operatorname{Re} g(0)] \approx |\mathcal{R}_{q,0}^{(o)}|^2 e^{2 \operatorname{Re} g(0)} = |\mathcal{R}_{q,0}^{(o)}|^2 \left(\frac{q}{q_*}\right)^{\frac{2 \operatorname{Re} g(0)}{\ln(q/q_*)}} \quad (5.3.17)$$

Our ansatz automatically solves the equation of order b^0 . To first order in b and in the slow roll parameters, the Mukhanov-Sasaki equation leads to an equation for $g(x)$ of the form

$$\frac{d^2 g}{dx^2} - \frac{2}{x} \frac{dg}{dx} + g = 2e^{ix}(2\epsilon_1 + \delta_1). \quad (5.3.18)$$

In writing this equation, we have dropped terms of order $\mathcal{O}(b\epsilon_0, b\delta_0)$, which amounts to setting $\nu_0 = 3/2$. Next, we notice that ϵ_1 is suppressed relative to δ_1 by a factor $\frac{f}{\phi_*}$. Since we are interested in the regime $\frac{f}{\phi_*} \ll 1$, we can thus drop the term proportional to ϵ_1 on the right hand side of equation (5.3.18). Furthermore, it turns out to be convenient to rewrite δ_1 using trigonometric identities. Ignoring an unimportant phase, one finds

$$\delta_1 = -\frac{3b}{\sqrt{1 + (3f\phi_*)^2}} \cos\left(\frac{\phi_0}{f}\right). \quad (5.3.19)$$

It will be convenient to write $\phi_0(x)$ as $\phi_0(x) = \phi_* - \frac{\ln(q/q_*)}{\phi_*} + \frac{\ln x}{\phi_*} = \phi_q + \frac{\ln x}{\phi_*}$. Introducing $r(x) \equiv \operatorname{Re}(g(x))$, equation (5.3.18) becomes

$$\frac{d^2 r}{dx^2} - \frac{2}{x} \frac{dr}{dx} + r = -\frac{6b}{\sqrt{1 + (3f\phi_*)^2}} \cos(x) \cos\left(\frac{\phi_q}{f} + \frac{\ln x}{f\phi_*}\right). \quad (5.3.20)$$

⁴As mentioned earlier, we will ignore the running of the scalar spectral index, but it may be worth pointing out that the information about the running is contained in this formula.

The solution to this equation can be found *e.g.* using Green's functions. We are particularly interested in the inhomogeneous solution at late times, *i.e.* in the limit of vanishing x . Using more trigonometric identities, we find that the solution in this limit can be brought into the form

$$r(0) = \frac{6b|\mathcal{I}(f\phi_*)|}{\sqrt{1+(3f\phi_*)^2}} \cos\left(\frac{\phi_q}{f} + \beta(f\phi_*)\right), \quad (5.3.21)$$

where $\beta(f\phi_*)$ is an unimportant phase that we will ignore, and \mathcal{I} is the integral

$$\mathcal{I}(f\phi_*) = \frac{\pi}{2} \int_0^\infty dx J_{\frac{3}{2}}(x) J_{-\frac{1}{2}}(x) x^{\frac{i}{f\phi_*}}. \quad (5.3.22)$$

Written in this form, the integral can be recognized as a Weber-Schafheitlin integral and can be done analytically (see *e.g.* [146]). One finds

$$|\mathcal{I}| = \sqrt{\frac{\pi}{8} \coth\left(\frac{\pi}{2f\phi_*}\right) f\phi_*}. \quad (5.3.23)$$

Combining equations (5.3.17), (5.3.21) and (5.3.23), we finally obtain an expression for δn_s ,

$$\delta n_s = \frac{2r(0)}{\cos\left(\frac{\phi_q}{f}\right)} = \frac{12b}{\sqrt{1+(3f\phi_*)^2}} \sqrt{\frac{\pi}{8} \coth\left(\frac{\pi}{2f\phi_*}\right) f\phi_*}. \quad (5.3.24)$$

Once again, this derivation is valid to first order in b and assumes slow roll for $\phi_0(t)$, $\phi_0 \gg M_p$, and $f \ll M_p$. In particular, it makes no use of an $f\phi_* \ll 1$ expansion, although this approximation will be needed in the derivation in subsection 5.3.2. A comparison between our analytical result for δn_s as a function of $f\phi_*$ for a fixed value of b and the result of a numerical calculation using a slight modification of the code described in [165] is shown in Figure 5.1.

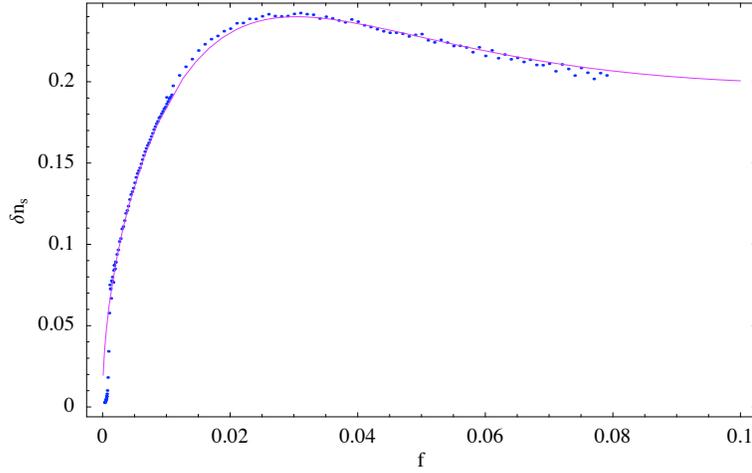


Figure 5.1: The solid line is the analytical result for δn_s as a function of f , for $b = 0.08$, while the dots are the numerical result obtained from an adaptation of the code used in [165].

5.3.2 Saddle-point approximation

As we have seen in the last subsection, it is possible to calculate the power spectrum analytically to first order in b , assuming slow roll for $\phi_0(t)$, $\phi_0 \gg M_p$, and $f \ll M_p$, but the derivation sheds little light on the physics behind the results. To get a better understanding, it is instructive to look at the integral (5.3.22) more explicitly. For this purpose, it is convenient to separate \mathcal{I} into its real and imaginary parts, $\mathcal{I} = \mathcal{I}_c + i\mathcal{I}_s$, with

$$\mathcal{I}_c = \int_0^\infty dx \frac{(\sin x - x \cos x) \cos x}{x^2} \cos\left(\frac{\ln x}{f\phi_*}\right), \quad (5.3.25)$$

$$\mathcal{I}_s = \int_0^\infty dx \frac{(\sin x - x \cos x) \cos x}{x^2} \sin\left(\frac{\ln x}{f\phi_*}\right). \quad (5.3.26)$$

For ranges of the axion decay constant such that $f\phi_* \ll 1$, these integrals can be done in a stationary phase approximation. Using trigonometric identities to rewrite the products of trigonometric functions appearing in the integrands

into sums of trigonometric functions with combined arguments, one finds that the stationary phase occurs at $\bar{x} = \frac{1}{2f\phi_*}$. Expanding around the stationary point and performing the integral as usual, one finds to leading order in $f\phi_*$

$$\mathcal{I}_c = \sqrt{\frac{\pi}{8}f\phi_*} \sin \left[\frac{1 + \ln(2f\phi_*)}{f\phi_*} - \frac{\pi}{4} \right], \quad (5.3.27)$$

$$\mathcal{I}_s = \sqrt{\frac{\pi}{8}f\phi_*} \cos \left[\frac{1 + \ln(2f\phi_*)}{f\phi_*} - \frac{\pi}{4} \right], \quad (5.3.28)$$

which leads to

$$|\mathcal{I}| = \sqrt{\mathcal{I}_c^2 + \mathcal{I}_s^2} = \sqrt{\frac{\pi}{8}f\phi_*}. \quad (5.3.29)$$

This agrees with our previous result, equation (5.3.23), as long as $f\phi_* \ll 1$. We have not only reproduced our earlier results, however: we also learn that at least for small $f\phi_*$, the integral is dominated by a period of time around $\bar{\tau} = -\frac{1}{2qf\phi_*}$. Up to the factor of two in the denominator, this corresponds to the period when the frequency of the oscillations of the scalar field background equals the frequency of the oscillations of a mode with comoving momentum q .⁵ The stationary phase approximation thus captures a resonance between the oscillations of the background and the oscillations of the fluctuations, and is good as long as $f\phi_* \ll 1$, *i.e.* as long as the resonance occurs while the mode is still well inside the horizon. One might suspect that this has an interpretation in terms of particle production, and we shall make this more precise in what follows.

Recall that our ansatz for \mathcal{R}_q was given in (5.3.15), where $g(x)$ is the solution of the equation

$$\frac{d^2g}{dx^2} - \frac{2}{x} \frac{dg}{dx} + g = 2e^{ix}\delta_1, \quad (5.3.30)$$

⁵This factor of two can be understood from momentum conservation.

with δ_1 again given by

$$\delta_1 = -\frac{3b}{\sqrt{1 + (3f\phi_*)^2}} \cos\left(\frac{\phi_q}{f} + \frac{\ln x}{f\phi_*}\right), \quad (5.3.31)$$

and initial conditions given by $\lim_{x \rightarrow \infty} g(x) = 0$ and $\lim_{x \rightarrow \infty} g'(x) = 0$. As we have just learned, the effect of the driving term can be ignored long after the resonance has occurred, *i.e.* for $x \ll \frac{1}{2f\phi_*}$.⁶ This implies that at late times, $g(x)$ must be a solution of the homogeneous equation which can be written as

$$g(x) = c_q^{(+)} \left(i\sqrt{\frac{\pi}{2}} x^{\frac{3}{2}} H_{3/2}^{(1)}(x) \right) + c_q^{(-)} \left(-i\sqrt{\frac{\pi}{2}} x^{\frac{3}{2}} H_{3/2}^{(2)}(x) \right), \quad (5.3.32)$$

where $c_q^{(\pm)}$ are momentum dependent coefficients. The solution for equation (5.3.30) can also be written explicitly as

$$\begin{aligned} g(x) = & (x \cos x - \sin x) \int_x^\infty \frac{2e^{iy}(\cos y + y \sin y)}{y^2} \delta_1 \\ & + (\cos x + x \sin x) \int_x^\infty \frac{2e^{iy}(\sin y - y \cos y)}{y^2} \delta_1. \end{aligned} \quad (5.3.33)$$

For $x \ll \frac{1}{2f\phi_*}$ we can take the lower limit in the integrals to zero and this can be brought into the form

$$\begin{aligned} g(x) = & \frac{1}{2}(\mathcal{I}^2 + i\mathcal{I}^1) \left(i\sqrt{\frac{\pi}{2}} x^{\frac{3}{2}} H_{3/2}^{(1)}(x) \right) \\ & + \frac{1}{2}(\mathcal{I}^2 - i\mathcal{I}^1) \left(-i\sqrt{\frac{\pi}{2}} x^{\frac{3}{2}} H_{3/2}^{(2)}(x) \right), \end{aligned} \quad (5.3.34)$$

⁶One should note that this is not because the driving term goes to zero, but because its frequency becomes too high for the system to keep up with it.

where the integrals \mathcal{I}^1 and \mathcal{I}^2 are given by

$$\mathcal{I}^1 = -\frac{6b}{\sqrt{1+(3f\phi_*)^2}} \int_0^\infty \frac{e^{iy}(\cos y + y \sin y)}{y^2} \cos\left(\frac{\phi_q}{f} + \frac{\ln x}{f\phi_*}\right) dy \quad (5.3.35)$$

$$\mathcal{I}^2 = -\frac{6b}{\sqrt{1+(3f\phi_*)^2}} \int_0^\infty \frac{e^{iy}(\sin y - y \cos y)}{y^2} \cos\left(\frac{\phi_q}{f} + \frac{\ln x}{f\phi_*}\right) dy \quad (5.3.36)$$

In the saddle point approximation these evaluate to

$$\mathcal{I}^1 = i\mathcal{I}^2 = -\frac{6b}{\sqrt{(1+(3f\phi_*)^2)}} \sqrt{\frac{\pi}{8} f\phi_*} e^{-i\left(\frac{\phi_q}{f} - \frac{1+\ln 2f\phi_*}{f\phi_*} + \frac{\pi}{4}\right)}. \quad (5.3.37)$$

Combining equations (5.3.15), (5.3.34), and (5.3.37), we finally find that the curvature perturbation for $x \ll \frac{1}{2f\phi_*}$ takes the form

$$\mathcal{R}_q = \mathcal{R}_{q,0}^{(o)} \left(i\sqrt{\frac{\pi}{2}} x^{\nu_0} H_{\nu_0}^{(1)}(x) - c_q^{(-)} i\sqrt{\frac{\pi}{2}} x^{\nu_0} H_{\nu_0}^{(2)}(x) \right), \quad (5.3.38)$$

with $c_q^{(-)}$ given, up to an unimportant momentum-independent overall phase, by

$$c_q^{(-)} = \frac{6b}{\sqrt{(1+(3f\phi_*)^2)}} \sqrt{\frac{\pi}{8} f\phi_*} e^{-i\left(\frac{\phi_q}{f}\right)}. \quad (5.3.39)$$

One might now interpret the coefficient $c_q^{(-)}$ of the negative frequency mode as a Bogoliubov coefficient that measures the amount of particles with co-moving momentum q being produced while this mode is in resonance with the background. It seems hard to make this precise as one really is comparing mode solutions of different backgrounds rather than mode solutions of different asymptotically Minkowski regions in the same background.

Equation (5.3.38) also shows that instead of starting in the Bunch-Davies state and then following the mode through the resonance, one may

start the evolution after the resonance has occurred but use a state that is different from the Bunch-Davies state, which is similar to what is considered in [166],[167],[153],[154],[155]. The departure from the Bunch-Davies state is of course quantified by $c_q^{(-)}$.

5.3.3 Bispectrum of scalar perturbations

We start by reviewing how resonance can drive the production of large non-Gaussianity during inflation, as proposed in [152]. We then present an estimate for the size of the non-Gaussianity for the model (5.2.1).

The three-point function can be calculated as [67]

$$\begin{aligned} \langle \mathcal{R}(\tau, \mathbf{q}_1) \mathcal{R}(\tau, \mathbf{q}_2) \mathcal{R}(\tau, \mathbf{q}_3) \rangle = \\ - i \int_{\tau_0}^{\tau} \langle [\mathcal{R}(\tau, \mathbf{q}_1) \mathcal{R}(\tau, \mathbf{q}_2) \mathcal{R}(\tau, \mathbf{q}_3), H_I(\tau')] \rangle a d\tau', \end{aligned} \quad (5.3.40)$$

where H_I is the interacting part of the Hamiltonian. H_I was calculated for a generic potential (see e.g. [67],[152]) at cubic order in the perturbations; it takes the form

$$\begin{aligned} H_I = & - \int d^3x \left[a\epsilon^2 \mathcal{R} \mathcal{R}^{\prime 2} + a\epsilon^2 \mathcal{R} (\partial \mathcal{R})^2 - 2\epsilon \mathcal{R}' (\partial \mathcal{R}) (\partial \chi) \right. \\ & \left. + \frac{a}{2} \epsilon \eta' \mathcal{R}^2 \mathcal{R}' + \frac{\epsilon}{2a} (\partial \mathcal{R}) (\partial \chi) (\partial^2 \chi) + \frac{\epsilon}{4a} (\partial^2 \mathcal{R}) (\partial \chi)^2 \right], \end{aligned} \quad (5.3.41)$$

where ∂ denote space derivatives,

$$\chi \equiv a^2 \epsilon \partial^{-2} \dot{\mathcal{R}}, \quad (5.3.42)$$

and we used the Hubble slow-roll parameter $\eta \equiv \dot{\epsilon}/(\epsilon H) = 2(\epsilon + \delta)$ because formulas in this subsection are simpler in terms of η than in terms of δ .

We would like to stress that (5.3.41) is exact for arbitrary values of the slow roll parameters ϵ and η . Substituting H_I into (5.3.40) produces six terms,

plus an additional term coming from a field redefinition. For the modulated linear potential (5.2.1), ϵ is small, as in standard slow roll inflation. On the other hand, contrary to the standard slow-roll approximation, $\dot{\eta}$ can be much larger than ϵ^2 . This suggests that the leading term comes from the $\epsilon\dot{\eta}$ term in the Hamiltonian.⁷ Hence we have [156],[152]

$$\langle \mathcal{R}(t, \mathbf{q}_1) \mathcal{R}(t, \mathbf{q}_2) \mathcal{R}(t, \mathbf{q}_3) \rangle \simeq i \left(\prod_i u_i(\tau_{end}) \right) \times \int_{-\infty}^{\tau_{end}} d\tau \epsilon \dot{\eta}' a^2 \left(u_1^*(\tau) u_2^*(\tau) \frac{d}{d\tau} u_3^*(\tau) + \text{sym} \right) \delta^3(\mathbf{K}) (2\pi)^3 + c.c.. \quad (5.3.44)$$

As in [152], we parameterize the non-Gaussianity as

$$\langle \mathcal{R}(\tau, \mathbf{q}_1) \mathcal{R}(\tau, \mathbf{q}_2) \mathcal{R}(\tau, \mathbf{q}_3) \rangle \equiv \frac{G(q_1, q_2, q_3)}{(q_1 q_2 q_3)^3} \delta^3(\mathbf{K}) \Delta_{\mathcal{R}}^4 (2\pi)^7, \quad (5.3.45)$$

where $\mathbf{K} = \mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3$. We take as an ansatz for the shape of the non-Gaussianity for our modulated linear potential

$$\frac{G(q_1, q_2, q_3)}{q_1 q_2 q_3} = f_{res} \sin \left(\frac{2}{\phi f} \ln K + \text{phase} \right) \quad (5.3.46)$$

Following [152] and comparing (5.3.44), (5.3.45) and (5.3.46), we obtain the estimate

$$f_{res} \simeq \frac{3 \dot{\eta}_1}{8H \sqrt{\phi f}}, \quad (5.3.47)$$

⁷In (3.9) of [67] this term was written as

$$\frac{\dot{\phi}^2}{\dot{\rho}^2} e^{3\rho} \dot{\mathcal{R}} \mathcal{R}^2 \frac{d}{dt} \left(\frac{\ddot{\phi}}{2\dot{\phi}\dot{\rho}} + \frac{\dot{\phi}^2}{4\dot{\rho}^2} \right), \quad (5.3.43)$$

which can be reduced to the term in (5.3.41) using $H' = -\dot{\phi}^2/2$.

where we have again used the notation $\eta = \eta_0 + b\eta_1 + \dots$. Using the background solution obtained in section 5.2, it is straightforward to find

$$\dot{\eta}_1 \simeq 2\dot{\delta}_1 \simeq -\sqrt{\frac{\mu^3}{3\phi_*}} \frac{6b}{f[1 + (3f\phi_*)^2]} \left[\cos\left(\frac{\phi_0}{f}\right) + (3f\phi_*) \sin\left(\frac{\phi_0}{f}\right) \right] \quad (5.3.48)$$

It is not hard to convince oneself that in the region of parameter space where $f_{res} > 1$ and $b \ll 1$, the second term in (5.3.48) is always negligible, *i.e.* $3f\phi \ll 1$. Hence our estimate for the non-Gaussianity is

$$f_{res} \simeq \frac{9b}{4(f\phi)^{3/2}} = \frac{9}{4} b \left(\frac{\omega}{H}\right)^{3/2}. \quad (5.3.49)$$

where we remind the reader that $\omega = \dot{\phi}/f$. As we will often refer to this equation, let us pause and comment on it. The resonant non-Gaussianity vanishes when the modulation is switched off, *i.e.* for $b = 0$. It is inversely proportional to some power of f (depending on which quantity is held fixed). Hence the smaller the axion decay constant f , the larger the non-Gaussianity. On the other hand, as we will see in section 5.5, there are theoretical lower (as well as upper) bounds on f , so that the non-Gaussian signal cannot be made arbitrarily large.

No complete analysis of the observational constraints on resonant non-Gaussianity has been performed to date (however, see [168]), and such an analysis is beyond the scope of the present work. Based on a rough comparison with known shapes of non-Gaussianity, we estimate that $f_{res} \gtrsim 200$ might be at the borderline of being excluded by the current data, while $f_{res} \lesssim 1$ would be difficult to detect in the next generation of experiments. A comprehensive analysis of the detectability of resonant non-Gaussianity is a very interesting topic for future research.

5.4 Observational Constraints

In the last section, we derived the theoretical predictions of axion monodromy inflation for the primordial power spectrum. We will now use these predictions to compare the model with the five-year WMAP data [78]. While the data in principle allows for a variety of statistics to be extracted, we will limit ourselves to the most fundamental one, the angular power spectrum. The reason for this is that the data is not now adequate for the polarization data or the three-point correlations to place meaningful additional constraints on the model. This will change as soon as the Planck data becomes available, and will be an interesting problem especially given the unusual shape of the non-Gaussianities the model predicts.

We work on a grid of model parameters. For each point on the grid, we compute the theoretical angular power spectrum with the publicly-available CAMB code [83]⁸, using the primordial power spectrum derived in the previous section in the form

$$\Delta_{\mathcal{R}}^2(q) = \Delta_{\mathcal{R}}^2(q_*) \left(\frac{q}{q_*} \right)^{n_s - 1 + \frac{\delta n_s}{\ln(q/q_*)} \cos\left(\frac{\phi q}{f} + \Delta\varphi\right)}. \quad (5.4.1)$$

The likelihood for a given theoretical power spectrum is calculated with a modified version of the WMAP five-year likelihood code that is now available on the LAMBDA webpage.⁹ The power spectrum in our model contains additional parameters beyond those of the WMAP five-year Λ CDM fit (namely, $\{\Omega_b h^2, \Omega_c h^2, \Omega_\Lambda, \tau, n_s, \Delta_{\mathcal{R}}^2\}$ and the marginalization parameter $\{A_{SZ}\}$). The additional parameters are δn_s , f and a phase $\Delta\varphi$. This phase parameterizes

⁸The code is available at <http://camb.info/>. Of course, we modify it to calculate all the multipole coefficients rather than calculating some and interpolating.

⁹http://lambda.gsfc.nasa.gov/product/map/dr3/likelihood_get.cfm

both our uncertainty in the number of e-folds needed, which originates in our poor understanding of reheating, and a microscopically determined phase offset in the sinusoidal modulation of the scalar potential arising in the string theory construction.

We fix the value of the scalar spectral index $n_s = 0.975$. As in any model of large-field inflation, the spectral index is a prediction of the model that depends only on the physics of reheating and, correspondingly, on the total amount of inflation since the observable modes exited the horizon. The value we choose corresponds to the situation in which the pivot scale exits the horizon 60 e-folds before the end of inflation. The results turn out to be fairly independent of the precise value chosen for the scalar spectral index and we could have chosen the value corresponding to any number of e-folds between 50 and 60. We fix $\{\Omega_c h^2, \Omega_\Lambda, \tau, A_{SZ}\}$ to the WMAP five-year best-fit values for the Λ CDM fit. We allow $f, \delta n_s, \Omega_b h^2, \Delta\varphi$ to vary on the grid, and we also marginalize over the scalar amplitude $\{\Delta_{\mathcal{R}}^2\}$ in the likelihood code. To obtain Figure 5.2, we thus marginalize over $\{\Omega_b h^2, \Delta_{\mathcal{R}}^2\}$ and over the unknown phase $\Delta\varphi$, while we fix $\{\Omega_c h^2, \Omega_\Lambda, \tau, A_{SZ}\}$, as we expect at most mild degeneracies between these parameters and the primordial ones.

The grid consists of 16 equidistantly spaced points in $\Omega_b h^2$ between $\Omega_b h^2 \approx 0.0212$ and $\Omega_b h^2 \approx 0.0266$, 128 equidistantly spaced points in δn_s between $\delta n_s = 0$ and $\delta n_s = 0.44$, 512 logarithmically spaced points in the axion decay constant f between $f = 9 \times 10^{-5}$ and $f = 10^{-1}$, as well as 32 points for the phase $\Delta\varphi$ between $\Delta\varphi = -\pi$ and $\Delta\varphi = \pi$. This leads to a grid with a total of 33,554,432 points. The analysis was run on 64 of the compute nodes of the Ranger supercomputer at the Texas Advanced Computing Center. The compute nodes are SunBlade x6420 blades, and each of the nodes provides four

AMD Opteron Quad-Core 64-bit processors with a core frequency of 2.3 GHz.

The resulting 68% and 95% contours in the $\delta n_s - f$ plane are shown in the left plot of Figure 5.2. To convert the resulting observational constraints on δn_s as a function of f into constraints on the microscopic parameter bf as a function of f , we make use of equation (5.3.24). The resulting 68% and 95% contours in the $bf-f$ plane are shown in the right plot of Figure 5.2. Roughly, the results can be summarized as $bf \lesssim 10^{-4}$ for $f \lesssim 0.01$ at 95% confidence level. Our best fit point is at a rather small value of the axion decay constant, $f = 6.67 \times 10^{-4}$, and a rather large amplitude for the oscillations, $\delta n_s = 0.17$. The fit improves by $\Delta\chi^2 \simeq 11$ over the fit in the absence of oscillations. The corresponding angular power spectrum is shown in Figure 5.3.

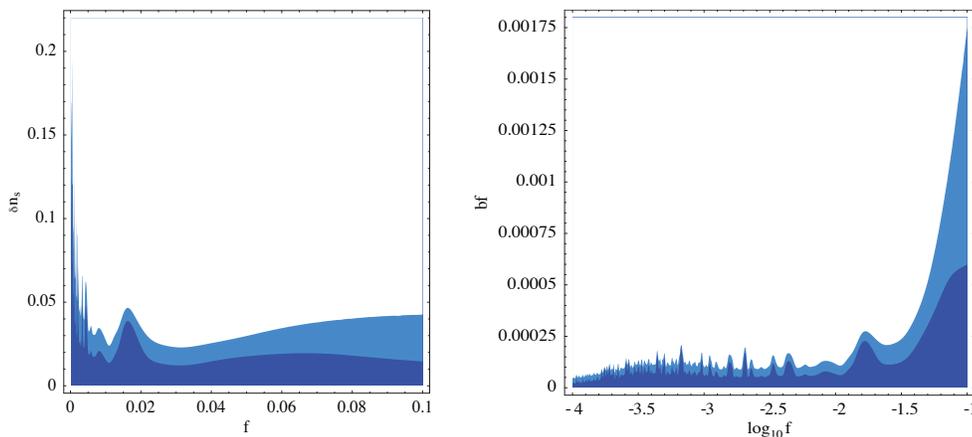


Figure 5.2: This plot shows the 68% and 95% likelihood contours in the $\delta n_s - f$, and $bf - \log_{10} f$ plane, respectively, from the five-year WMAP data on the temperature angular power spectrum.

The improvement can be traced to a better fit to the data around the first peak. We would like to stress, however, that we do not take this as an indication of oscillations in the observed angular power spectrum. Similar spikes

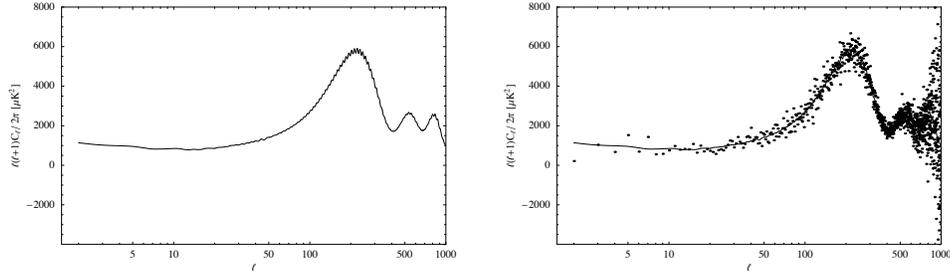


Figure 5.3: The left plot shows the angular power spectrum for the best fit point $f = 6.67 \times 10^{-4}$, and $\delta n_s = 0.17$. The right plot shows the angular power spectrum for the best fit point together with the unbinned WMAP five-year data.

in the likelihood function occur quite generally when fitting an oscillatory model to toy data generated with the conventional power spectrum without any oscillations, because the oscillations fit some features in the noise. The polarization data could provide a cross check, but we find that it is presently not good enough to do so in a meaningful way.

Let us say a few words motivating the necessity of marginalizing over $\Omega_b h^2$ and $\Delta\varphi$. There is a known degeneracy in the angular power spectrum between $\Omega_b h^2$ and n_s , as changing $\Omega_b h^2$ changes the ratio of the power in the first and second acoustic peaks, which to some extent can be undone by changing the spectral tilt n_s . In our case we do not vary n_s , but we add a sinusoidal contribution to the standard power spectrum. It is intuitively clear that by doing so we can change the ratio of power in the first and second acoustic peak by choosing the right oscillation frequency (controlled by f) and phase $\Delta\varphi$, leading to a degeneracy between $\Omega_b h^2$ and δn_s at least for a certain range of f .

The most straightforward way to demonstrate this degeneracy between

$\Omega_b h^2$ and δn_s arising for certain ‘resonant’ values of f is to present a likelihood plot in the Ω_b - δn_s plane for a value of f for which the degeneracy is clearly visible. An example is shown in the plot on the left side of Figure 5.4. It shows that marginalizing over $\Omega_b h^2$ is necessary to obtain correct exclusion contours on δn_s and f . That marginalization over the phase is necessary can easily be

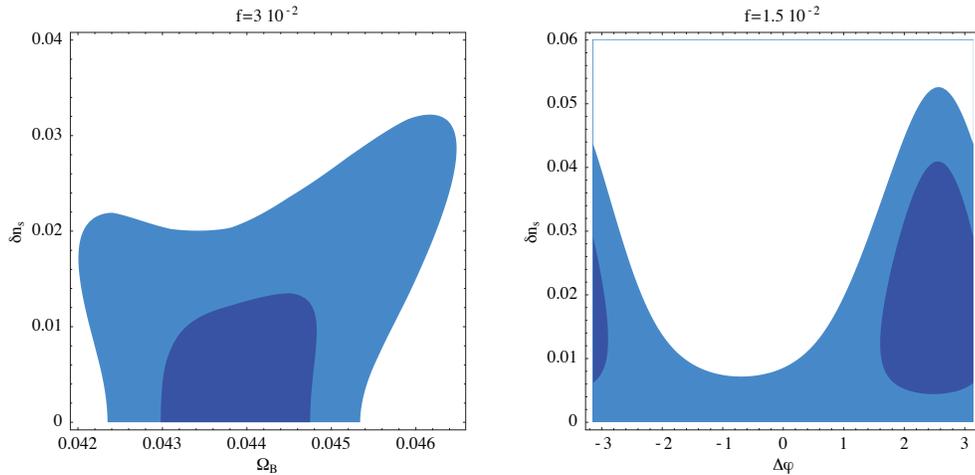


Figure 5.4: These plots show the 68% and 95% likelihood contours for the five-year WMAP data on the temperature angular power spectrum in the δn_s - $\Omega_b h^2$ plane and δn_s - $\Delta\phi$ plane for an axion decay constant of $f = 3 \times 10^{-2}$ and $f = 1.5 \times 10^{-2}$, respectively.

seen from a likelihood plot in the δn_s - $\Delta\phi$ plane. This is shown in the plot on the right side of Figure 5.4.

We have also performed a Markov chain Monte Carlo analysis for the model using the publicly available CosmoMC code [148].¹⁰ While the Monte Carlo has the advantage that it is less computationally intensive than a grid

¹⁰<http://cosmologist.info/cosmomc/>

when varying all cosmological parameters, the likelihood function for oscillatory models turns out to be rather spiky, making the Monte Carlo hard to set up, because the chains tend to get trapped in the spikes.

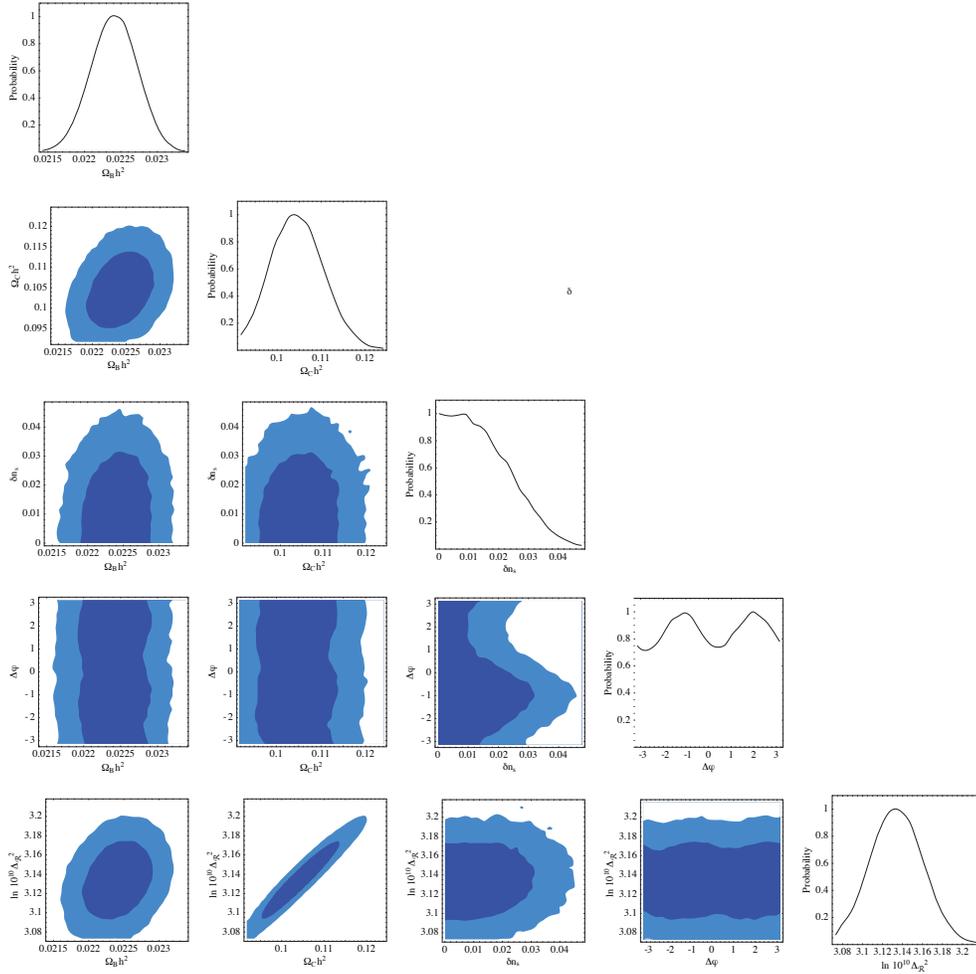


Figure 5.5: This figure shows a triangle plot for some of the parameters that were sampled in a Markov chain Monte Carlo for an axion decay constant of $f = 10^{-2}$. The contours again represent 68% and 95% confidence levels.

To some extent this can be overcome by taking out the problematic re-

gions or increasing the temperature of the Monte Carlo. When run on parts of the parameter space where the Monte Carlo runs reliably, we found agreement with the grid-based results shown above.

The most problematic direction to sample is that of the axion decay constant, f . We show the result of one of our chains for $f = 0.01$ in Figure 5.5. The plot shows marginalized one-dimensional distributions and two-dimensional 68% and 95% confidence level limits for the most important Λ CDM parameters as well as δn_s and $\Delta\phi$. In the Monte Carlo, we sampled the parameters δn_s , $\Delta\phi$, all parameters of the Λ CDM except the scalar spectral index, as well as the Sunyaev-Zel'dovich amplitude.

5.5 Microphysics of Axion Monodromy Inflation

In section 5.1 we briefly reviewed the properties of axion monodromy inflation, focusing on the description in effective field theory. For a general characterization of the signatures of the scenario, the phenomenological model of section 5.1 was sufficient. However, the phenomenological parameters f, μ, Λ are in principle derivable from the data of a string compactification, and as such they obey nontrivial microscopic constraints: the ranges and correlations of these parameters are restricted by microphysics.

We should therefore determine the values of the phenomenological parameters allowed in consistent, computable string compactifications. We will begin by reviewing the string theory origin of axion monodromy inflation, both to set notation and to highlight the properties most relevant in constraining the parameters f, μ, Λ . For concreteness we will restrict our attention to a specific realization of the scenario, in O3-O7 orientifolds of type IIB string theory, with the Kähler moduli stabilized by nonperturbative effects. Our considera-

tions could be generalized to other compactifications, but the numerical results would differ.

5.5.1 Axions in string theory

Let us first review the origin of the relevant axions. Our conventions and notation are summarized in appendix A.1. Consider type IIB string theory compactified on an orientifold of a Calabi-Yau threefold X . Let the forms ω^I be a basis of the cohomology $H^2(X, \mathbb{Z})$, normalized such that $\int_{\Sigma_I} \omega^J = \delta_I^J (2\pi)^2 \alpha'$, where Σ_I are a basis of the dual homology $H_2(X, \mathbb{Z})$. The RR two-form C_2 gives rise to a four-dimensional axion via the ansatz¹¹

$$C_2 = \frac{1}{2\pi} c_I(x) \omega^I, \quad (5.5.1)$$

where x is a four-dimensional spacetime coordinate. The ten-dimensional Einstein-frame action [169] that follows is

$$\int d^{10}x \frac{g_s \sqrt{-g_E}}{2(2\pi)^7 \alpha^4} |dC_2|^2 = \int d^{10}x \frac{g_s \sqrt{-g_E}}{12(2\pi)^9 \alpha^4} g_E^{\mu\nu} \partial_\mu c_I \partial_\nu c_J \omega_{ij}^I \omega_{i'j'}^J g_E^{ii'} g_E^{jj'}. \quad (5.5.2)$$

Notice that the axions only have derivative couplings, and hence enjoy a continuous shift symmetry at the level of the classical action. In subsection 5.5.2.3 we will recall the origin of this symmetry and explain how it persists to all orders in perturbation theory and is broken by nonperturbative effects.

Upon dimensional reduction, one finds a relation between the four-dimensional reduced Planck mass M_{pl} and α ,

$$\alpha M_{pl}^2 = \frac{\mathcal{V}_E}{\pi}, \quad (5.5.3)$$

¹¹The factor of 2π is introduced so that the four-dimensional axions c_I have periodicity 2π , as can be seen via S-duality from the world-sheet coupling $i \int B_2 / (2\pi\alpha)$. Notice that in our conventions C_2 and ω^I have the dimensions of length-squared, while c_I are dimensionless.

where \mathcal{V}_E is the Einstein-frame (dimensionless) volume of the Calabi-Yau X measured in units of $l_s \equiv 2\pi\sqrt{\alpha'}$. The decay constant of the canonically normalized axion is then

$$\frac{f^2}{M_{pl}^2} = \frac{g_s}{48\pi^2\mathcal{V}_E} \left[\frac{\int \omega \wedge *\omega}{(2\pi)^6\alpha'^3} \right]. \quad (5.5.4)$$

The present definition of the axion decay constant differs by a factor of 2π from that in [151], *i.e.* $f^{here} = 2\pi f^{there}$. As a consequence our canonically normalized axion has periodicity $2\pi f$, consistent with (5.2.1).

5.5.2 Dimensional reduction and moduli stabilization

5.5.2.1 Four-dimensional data of O3-O7 orientifolds

Now we consider how to stabilize the compactification in a setup that will allow inflation. We focus on the KKLT scenario for moduli stabilization [170]. We assume that the complex structure moduli, the dilaton, and any open string moduli have been stabilized at a higher scale, and we concentrate on the remaining closed string moduli (specifically, the remaining moduli are those descending from hypermultiplets). The $\mathcal{N} = 1$ supersymmetric four-dimensional theory resulting from dimensional reduction of type IIB orientifolds was worked out in detail in [171]. We are interested in orientifold actions under which the holomorphic three-form Ω of the Calabi-Yau manifold is odd, so that the fixed-point loci are O3-planes and O7-planes. The cohomology decomposes into eigenspaces of the orientifold action,

$$H^{(r,s)} = H_+^{r,s} \oplus H_-^{r,s}. \quad (5.5.5)$$

We therefore divide the basis $\omega_A, A = 1, \dots, h^{1,1}$ into $\omega_\alpha, \alpha = 1, \dots, h_+^{1,1}$ and $\omega_a, a = 1, \dots, h_-^{1,1}$. Working out the sign of the orientifold action on the physical

fields, one finds that the two-forms C_2 and B_2 are odd, and should be expanded in terms of the ω_a . Grimm and Louis [171] have derived the Kähler coordinates on the corresponding moduli space, *i.e.* the proper complex combinations of fields that appear as the lowest components of chiral multiplets:¹²

$$G^a \equiv \frac{1}{2\pi} \left(c^a - i \frac{b^a}{g_s} \right) \quad (5.5.6)$$

$$T_\alpha \equiv i\rho_\alpha + \frac{1}{2}c_{\alpha\beta\gamma}v^\beta v^\gamma + \frac{g_s}{4}c_{abc}G^b(G - \bar{G})^c \quad (5.5.7)$$

where ρ_α comes from the RR four-form C_4 integrated over some orientifold-even four-cycle Σ_α , with $\alpha = 1, \dots, h_+^{1,1}$; and c^a and b^a come from the RR and NS-NS two-forms C_2 and B_2 integrated over some orientifold odd two-cycle Σ_a with $a = 1, \dots, h_-^{1,1}$. The tree-level Kähler potential is given by¹³

$$K = \log \left(\frac{g_s}{2} \right) - 2 \log \mathcal{V}_E \quad (5.5.8)$$

where the (dimensionless) Einstein-frame volume \mathcal{V}_E of the Calabi-Yau manifold is defined in (A.1.4). The dependence of this Kähler potential on the multiplets (5.5.6) and (5.5.7) cannot be written down explicitly for a generic choice of the intersection numbers c_{IJK} . The implicit dependence is given by writing the (Einstein-frame) volume in terms of two-cycle volumes v^α

$$K = \log \left(\frac{g_s}{2} \right) - 2 \log \left[\frac{1}{6} c_{\alpha\beta\gamma} v^\alpha(T, G) v^\beta(T, G) v^\gamma(T, G) \right]. \quad (5.5.9)$$

Then one has to solve (5.5.7) for v^α and substitute the result into the above Kähler potential. The Kähler potential is a function of \mathcal{V}_E , and hence is a

¹²We use the same notation as [171] with two exceptions: we rescale T_α as $T_\alpha^{here} = (2/3)T_\alpha^{there}$, and we add a factor of $(2\pi)^{-1}$ in the definition of G^a such that the fields c^a and b^a have periodicity 2π . See appendix A.1 for more details on our conventions.

¹³We assume that the axio-dilaton $\tau = C_0 + ie^{-\phi}$ is already stabilized by fluxes at $\tau = i/g_s$ and we write down the dilaton-dependent part of the Kähler potential only to keep track of factors of g_s .

function of v^α , and in turn of $\tau_\alpha \equiv \text{Re} T_\alpha$ and $\text{Im} G$, but does not depend on $\text{Re} G$ and $\text{Im} T_\alpha$ (as can be seen by taking the real part of (5.5.7)). One might be tempted to conclude that c enjoys a shift symmetry but that b does not, but, as we will explain in subsection 5.5.2.3, both fields have shift symmetries.

The tree-level superpotential W_0 does not depend on the multiplets (5.5.6) and (5.5.7). In fact it depends on the complex structure moduli and the dilaton, which we assume have already been stabilized by fluxes. Therefore we will take W_0 to be a discretely tunable constant.

5.5.2.2 Nonperturbative stabilization of the Kähler moduli

Let us now proceed to consider nonperturbative effects. We follow the KKLT strategy [170] for the construction of a de Sitter vacuum. We assume that each four-cycle T_α is wrapped either by a Euclidean D3-brane or by a stack of D7-branes giving rise to a four-dimensional gauge theory that undergoes gaugino condensation.¹⁴ This results in the following four-dimensional superpotential:

$$W = W_0 + \sum_{\alpha=1}^{h_+^{1,1}} A_\alpha e^{-a_\alpha T_\alpha}, \quad (5.5.10)$$

where A_α will be treated as constants, as they depend on the complex structure moduli, which we have assumed to be stabilized; $a_\alpha \equiv 2\pi/N_\alpha$, with N_α the number of D7-branes in the stack; and $N_\alpha = 1$ for the case of a Euclidean D3-brane. We can find a supersymmetric minimum by solving for the vanishing

¹⁴In general, Euclidean D3-branes or D7-branes will wrap some linear combinations \tilde{T}_α of the cycles appearing in (5.5.7), rather than the basis cycles T_α themselves, but for simplicity we will suppress this issue.

of all the F-terms: for the $h_+^{1,1}$ even Kähler moduli via

$$\begin{aligned} 0 &= D_\alpha W \equiv \partial_{T_\alpha} W + W \partial_{T_\alpha} K = -A_\alpha a_\alpha e^{-a_\alpha T_\alpha} - 2W \frac{\partial_{T_\alpha} \mathcal{V}_E}{\mathcal{V}_E} \\ &= -A_\alpha a_\alpha e^{-a_\alpha T_\alpha} - W \frac{v^\alpha}{2\mathcal{V}_E}, \end{aligned} \quad (5.5.11)$$

and for the $h_-^{1,1}$ odd moduli via

$$0 = D_a W \equiv \partial_{G^a} W + W \partial_{G^a} K = -iW \frac{c_{\alpha a c} v^\alpha b^c}{4\pi \mathcal{V}_E} \quad (5.5.12)$$

where in both cases in the last step we used the chain rule and the definitions of G^a and T_α in terms of two-cycle volumes v^α . The condition (5.5.11) is simplified if we first solve for $\text{Im } T_\alpha$, which gives

$$a_\alpha \text{Im } T_\alpha = \theta_{A_\alpha} - \theta_{W_0} + k_\alpha \pi, \quad k_\alpha \in \mathbb{Z}. \quad (5.5.13)$$

Then we are left with the set of real equations for each α ,

$$(\pm 1)_\alpha |A_\alpha| a_\alpha e^{-a_\alpha \tau_\alpha} = \partial_{T_\alpha} K \left(|W_0| + \sum_\beta (\pm 1)_\beta |A_\beta| e^{-a_\beta \tau_\beta} \right), \quad (5.5.14)$$

where (± 1) depends on the value of k in (5.5.13). As long as the orientifold-even four-cycle Kähler moduli are defined as in (5.5.7), then

$$\partial_{T_\alpha} K = -v^\alpha / (2\mathcal{V}_E) < 0, \quad (5.5.15)$$

for every α . Now we prove that in (5.5.14) the minus sign has to be chosen for every α in order to have a supersymmetric solution. First we notice that the sign of the right hand side does not depend on α , so k_α and hence $(\pm 1)_\alpha$ have to be the same for every α . If we choose the positive sign in (5.5.14), the quantity in brackets in the right hand side is manifestly positive. Then the two sides of the equation have opposite signs and no (compact) solution

exists. To summarize, the minimization of $\text{Im } T_\alpha$ boils down to taking all A_α real and negative and W_0 real and positive or the other way around.¹⁵

Concerning (5.5.12), an obvious solution is given by $b^a = 0$ for every a . As argued in [151], an inflationary model with a b -type axion as the inflaton will generically suffer from an eta problem, and we will therefore focus on a c -type axion.

5.5.2.3 Nonperturbative breaking of axionic shift symmetries

Axionic shift symmetries are central to this chapter, so we will now explain how they originate and how they are ultimately broken by nonperturbative effects. First, let us recall the classic result [172],[173] establishing the shift symmetry to all orders in perturbation theory. Consider the axion $b = \int_\Sigma B/(2\pi\alpha)$, where B is the NS-NS two-form potential and Σ is a two-cycle in the Calabi-Yau manifold. The vertex operator representing the coupling of b to the string worldsheet is [172]

$$V(k) = \frac{1}{2\pi\alpha'} \int_\Sigma d^2\xi \exp\left(ik \cdot X(\xi)\right) \epsilon^{\alpha\beta} \partial_\alpha X^\mu \partial_\beta X^\nu B_{\mu\nu}(X). \quad (5.5.16)$$

At zero momentum, this coupling is seen to be a total derivative in the worldsheet theory. Therefore, the axion b can only have derivative couplings (which vanish at zero momentum), to any order in sigma-model perturbation theory. Notice that the genus of the worldsheet did not enter in this argument, so the axion shift symmetry is also valid to all orders in string perturbation theory.

¹⁵We notice that if one chooses as Kähler variable a linear combination of the T_α defined in (5.5.7), as is done *e.g.* in the large volume scenario with Swiss-cheese Calabi-Yau manifolds, then the sign of $\partial_{T_\alpha} K$ can depend on α . In this case, the minimization of $\text{Im } T_\alpha$ boils down to taking W_0 real and positive and A_α real with $\text{Sign}(A_\alpha) = \text{Sign}(\partial_{T_\alpha} K)$, up to multiplying W by an overall phase.

This argument fails in the presence of worldsheet boundaries (*i.e.*, D-branes), and also fails once worldsheet instantons, or D-brane instantons, are included. In axion monodromy inflation, both sorts of breaking play an important role, as we shall now explain.

First, the introduction of an NS5-brane wrapping a curve Σ_a creates a monodromy for the axion c_a , spoiling its shift symmetry and inducing an asymptotically linear potential [151]. Specifically, the potential induced by the Born-Infeld action of the NS5-brane (obtained by S-dualizing the Born-Infeld action of a D5-brane) is

$$V(c_a) = \frac{\epsilon}{g_s(2\pi)^5\alpha^2} \sqrt{\ell^4 + (2\pi g_s c_a)^2}, \quad (5.5.17)$$

where $\ell\sqrt{\alpha}$ is the size of Σ_a and ϵ captures the possibility of suppression due to warping. For $c_a \gg 1$, this potential is linear in c_a , or in the corresponding canonically normalized field, which we denoted by ϕ in the preceding sections. Let us remark that the square root form of the potential can be important at the end of inflation and also makes a small change in the number of e-foldings produced for given parameter values, so that in a model that includes a specific scenario for reheating, the square root structure should be incorporated as well. As we have not invoked a concrete reheating scenario, for our purposes the linear potential suffices, but one must still bear in mind that this form is not valid for small ϕ .

As we will explain in detail, the D-brane instantons involved in moduli stabilization introduce sinusoidal modulations to the linear potential. We will work exclusively in a regime in which the breaking by wrapped branes dominates over the nonperturbative breaking, although we remark in passing that the complementary regime might be interesting for realizing models involving repeated tunneling.

The breaking of the b shift symmetry by Euclidean D-branes (or by gaugino condensation on D7-branes) is slightly subtle, so we will address it briefly. As we remarked above, b appears quadratically in the classical Kähler potential, which seems to contradict the statement that it enjoys a shift symmetry at the perturbative level in the absence of boundaries. However, there is no contradiction: the shift symmetry of b is true at constant two-cycle volumes v and *not* at constant four-cycle volumes T . To see this, suppose that there is a single Kähler modulus T , so that the superpotential is of the form (5.5.10) with $h_+^{1,1} = 1$. The Kähler potential is then [171]

$$K = -3M_p^2 \log (T + \bar{T} - db^2), \quad (5.5.18)$$

with d a constant. In the absence of a nonperturbative superpotential, a suitable simultaneous shift of $T + \bar{T}$ and b is a symmetry of the scalar potential of this system; under such a shift, the two-cycle volumes v are invariant. However, this symmetry is spoiled by the nonperturbative term in W , because the superpotential and the scalar potential are no longer invariant. Therefore, in a scenario in which the four-cycle volumes are stabilized nonperturbatively, the b axion receives a mass in a stabilized vacuum.

At this stage the mass-squared m_b^2 of b is proportional to the vacuum energy and hence is negative in the supersymmetric AdS minimum. The minimum of the potential will be the final point of the inflationary dynamics, and hence we would like it to have a very small *positive* cosmological constant to be consistent with the current accelerated expansion of the universe. Thus, we need to include an uplifting term. In the uplifted minimum, $m_b^2 \propto V_{dS} > 0$. This relation is the origin of the eta problem that was found in [151] choosing

b as an axion: for a generic uplifting,¹⁶ $V''(b) \sim V(b)$, so that $\eta \sim \mathcal{O}(1)$ and slow roll inflation does not take place. This is completely analogous to the eta problem of D-brane inflation found in [175] and can be intuitively understood in the same way. Here we will take $b = 0$ as the stabilized value¹⁷ of b and concentrate on c as a candidate inflaton.

Let us now turn to consider c , which does not appear in the Kähler potential or superpotential at any order in perturbation theory. To assess c as an inflaton, one should determine the leading nonperturbative effects, either in the superpotential or in the Kähler potential, that do introduce a potential for c , *i.e.* one should identify the leading breaking of the shift symmetry. Euclidean D3-branes carrying vanishing D1-brane charge do not induce a potential for c , but Euclidean D3-branes supporting worldvolume fluxes (and hence nonvanishing D1-brane charge) give rise to a dependence on c , via the Chern-Simons coupling $\int F_2 \wedge C_2$. As observed in [151], it follows that when the Kähler moduli are stabilized by Euclidean D3-branes, c receives a mass in the stabilized vacuum: one must sum over Euclidean D-brane contributions to the superpotential, including summing over the amount $n = \int F_2$ of magnetization, and this generically introduces an eta problem for c . The solution, as explained in [151], is to stabilize the Kähler moduli via gaugino condensation on D7-branes, which leads to an exponentially smaller (and hence negligible) mass for c .

¹⁶Notice that the proportionality constant in $m_b^2 \propto V_{dS}$ depends on the volume-dependence of the uplifting term and could be made small for particular choices of the latter as proposed in [174].

¹⁷It is easy to check that $b = 0$ is still the stabilized value after the inclusion of nonperturbative corrections to the Kähler potential, *cf.* subsection 5.6.5.

5.5.3 Axion decay constants in string theory

We now turn to the important task of expressing the axion decay constant, f , in terms of the data of a compactification. As we reviewed in subsection 5.5.1, the decay constant of an axion $C_2 = c(x)\omega/(2\pi)$ is given by

$$\frac{f^2}{M_{pl}^2} = \frac{g_s}{48\pi^2\mathcal{V}_E} \left[\frac{\int \omega \wedge *\omega}{(2\pi)^6\alpha^3} \right], \quad (5.5.19)$$

so that the primary task is to compute the norm $\int \omega \wedge *\omega$. (This problem has been studied in a wide range of examples in [176].) We will first recall, in subsection 5.5.3.1, how to express the axion kinetic term, and hence also the axion decay constant, in terms of $\mathcal{N} = 1$ data. This will lead us to a simple expression for the decay constant in terms of intersection numbers of the Calabi-Yau. We will then propose a class of models in which the decay constant is rather small, motivated by the fact that with other parameters held fixed, decreasing f increases the amplitude of the resonant non-Gaussianity. Next, in subsection 5.5.3.2, we will present a concrete example that illustrates the geometry of a configuration that leads to small f .

5.5.3.1 Decay constants in terms of $\mathcal{N} = 1$ data

In subsection 5.5.2.1 we have reviewed, following [171], the four-dimensional $\mathcal{N} = 1$ description of Type IIB O3-O7 orientifolds. The multiplets relevant for us are the orientifold-odd chiral multiplets G^a and the orientifold-even chiral multiplets T_α . The tree-level Kähler potential given in (5.5.9) determines the kinetic terms for G^a and hence the decay constants of the axions b^a and c^a . First let us notice that the Kähler metric in the space of the chiral multiplets T_α and G^a factorizes in two blocks, $K_{T_\alpha\bar{T}_\beta}$ and $K_{G^a\bar{G}^b}$. The reason is that off-diagonal terms such as $K_{T_\alpha\bar{G}^a}$ are proportional to intersection

numbers $c_{\alpha\beta a}$ with one odd index and two even indices, which are forbidden by the orientifold action [171]. We are interested in one particular mode from among the G^a , which we will denote by G^- ; Σ_- is then the orientifold-odd two-cycle that supports our candidate inflaton c^- . We now choose a basis for G^a such that $K_{G^a\bar{G}^b}$ is block diagonal with a 1×1 block $K_{G^-\bar{G}^-}$. The kinetic term for c^- is then given by

$$-\frac{1}{2}f^2 (\partial c^-)^2 = M_{pl}^2 K_{G^-\bar{G}^-} \frac{1}{(2\pi)^2} (\partial c^-)^2 \subset M_{pl}^2 K_{G^-\bar{G}^-} |\partial G^-|^2, \quad (5.5.20)$$

where

$$K_{G^-\bar{G}^-} = \frac{\partial^2 K(G, T)}{\partial G^- \partial \bar{G}^-} = -g_s \frac{c_{\alpha--} v^\alpha}{4\mathcal{V}_E}, \quad (5.5.21)$$

and we used

$$c_{\alpha\beta\gamma} v^\beta v^\gamma = 2\tau_\alpha + g_s c_{\alpha bc} \text{Im } G^b \text{Im } G^c. \quad (5.5.22)$$

Hence we can express the decay constant of the axion c^- as

$$\frac{f^2}{M_{pl}^2} = \frac{g_s}{8\pi^2} \frac{c_{\alpha--} v^\alpha}{\mathcal{V}_E}. \quad (5.5.23)$$

As promised, we have expressed the norm $\int \omega \wedge * \omega$ in terms of the intersection numbers

$$\frac{\int \omega \wedge * \omega}{(2\pi)^6 \alpha^3} = \frac{2}{3} c_{\alpha--} v^\alpha. \quad (5.5.24)$$

In subsection 5.6.4 we will discuss the constraints that follow from the result (5.5.23). First, in the following subsection we provide some geometrical intuition for (5.5.23).

5.5.3.2 An example: a complex plane of fixed points

An instructive example arises from considering an orbifold that is locally $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$, *i.e.* an Eguchi-Hanson space fibered over a base Σ of complex dimension one. Let ω be the two-form dual to the blowup cycle of the orbifold, and let Σ be the two-manifold of fixed points, *i.e.* the base over which the Eguchi-Hanson space is fibered.¹⁸ We are interested in the decay constant of the axion $C_2 = \frac{1}{2\pi}c(x)\omega$, so we must compute $\int \omega \wedge *\omega$. In the local approximation, this is straightforward, as we shall see. However, far from the fixed-point locus, the fiber may deviate substantially from the Eguchi-Hanson geometry, in a complicated and model-dependent way, and moreover the fixed-point locus Σ may be embedded in the compact space in a nontrivial manner. Happily, the integral $\int \omega \wedge *\omega$ has its primary support near the fixed-point locus, where the local approximation is excellent.

We recall, following the useful summary in Appendix B of [177], that the Eguchi-Hanson space has a unique homology two-cycle of radius $a/2$, where $r = a$ defines the location of the coordinate singularity; here r is the standard radial coordinate. The two-form ω corresponding to this cycle may be written

$$\omega = 2\pi\alpha \frac{a^2}{r^2} \left(\frac{dr}{r} \wedge d\psi + \cos\theta \frac{dr}{r} \wedge d\phi + \frac{1}{2} \sin\theta d\theta \wedge d\phi \right) \quad (5.5.25)$$

in terms of r and the angular coordinates ψ, θ, ϕ . By observing that $*_4\omega = -\omega$ and that $\int \omega \wedge \omega = -(2\pi)^4 \alpha^2/2$, one finds

$$\frac{\int_{EH} \omega \wedge *_4\omega}{(2\pi)^4 \alpha^2} = \frac{1}{2}. \quad (5.5.26)$$

¹⁸Concretely, we are imagining that Σ extends into a warped throat region, and that an NS5-brane wraps the blowup cycle at a particular location in the throat. The warping is invoked in order to suppress the energy density of the wrapped NS5-brane. See [151] for further details, and for an example of a suitable orbifold action in a Klebanov-Strassler throat.

Clearly, given the form of ω , this integral has its support in a region $a \leq r \lesssim few \times a$. This justifies the local approximation as long as the compact space has a radius that is large compared to a . Next, we observe that

$$\frac{\int \omega \wedge *_6 \omega}{(2\pi)^6 \alpha^3} = \frac{\int_{EH} d^4 x \omega \wedge *_4 \omega}{(2\pi)^4 \alpha^2} \times \frac{\int_{\Sigma} \sqrt{g}}{(2\pi)^2 \alpha} = \frac{1}{2} \text{Vol}(\Sigma) \quad (5.5.27)$$

Substituting this in (5.5.19), we recover the parametric scaling of subsection 5.5.3.1.

5.6 Microscopic Constraints

We now turn to determining the ranges of our phenomenological parameters that are allowed in a consistent and computable microphysical model.

Let us first remark that, as usual in string theory model building, computability imposes stringent constraints on the compactification parameters. Because large-field inflation involves substantial energy densities and requires correspondingly steep moduli barriers, the compact space needs to be reasonably small, so that the Kaluza-Klein scale and the (necessarily lower) scale of moduli masses can be large enough to prevent runaway moduli evolution. Clearly, one must then carefully check that the compactification is still large enough for the supergravity approximation to be valid; furthermore, back-reaction of the inflationary energy on the compact space is a serious issue, particularly when this space is not large in string units. Incorporating these requirements then leads to severe restrictions on the allowed values of the decay constant f .

We will begin in subsection 5.6.1 by considering the constraints from computability, then give, in subsection 5.6.2, a qualitative description of the

constraints from backreaction, deferring details to appendix B. Next, in subsection 5.6.3, we will verify that a two-derivative action suffices to describe this system. This is not obvious, as rapid oscillations in the potential could enhance the importance of generic higher-derivative terms; however, we will show that the specific terms emerging from string theory are negligible in our solution. We then apply these constraints in subsection 5.6.4 to determine the range of the decay constant f . Finally, we estimate the size bf of the modulations; as this is rather model-dependent, in subsection 5.6.5, we will restrict our attention to a specific example in which a periodic contribution is generated by Euclidean D1-brane corrections to the Kähler potential.

5.6.1 Constraints from computability

In this subsection we will list several constraints coming from the consistency of the string theory setup. We will first require the validity of the string and α perturbation expansions, and the validity of neglecting higher-order corrections to the nonperturbative superpotential, and then we will require that the inflaton potential does not destabilize the compactification.

First of all we require the validity of string perturbation theory, *i.e.* we require $g_s \ll 1$. We must also ensure the validity of the α expansion. To do this including a reasonable *estimate* of numerical factors such as 2π , it is convenient to use worldsheet instantons as a proxy for perturbative α corrections, because the normalization is easily determined. To get the correct coefficient, we start from the string-frame ten-dimensional metric g_{string} and impose that the worldsheet instanton action obeys $e^{-S_{ws}} \lesssim e^{-2}$, or

$$\frac{1}{2\pi\alpha'} \int \sqrt{g_{string}} \gtrsim 2, \quad (5.6.1)$$

which using $g_{string} = g_{Einstein}\sqrt{g_s}$ is converted to Einstein frame

$$2 < \frac{\sqrt{g_s}}{2\pi\alpha} \int_{\Sigma^\alpha} J = \sqrt{g_s} v^\alpha 2\pi \quad \Rightarrow \quad v^\alpha > \frac{1}{\pi\sqrt{g_s}}, \quad (5.6.2)$$

and we used that $\int_{\Sigma^\alpha} \omega^\beta = (2\pi)^2 \alpha \delta_\alpha^\beta$.

As we invoked nonperturbative corrections to the superpotential, we must also require that any further superpotential corrections, *e.g.* from multi-instantons, are negligible. For this purpose it suffices to impose

$$e^{-a_\alpha T_\alpha} < e^{-2} \ll 1 \quad \Rightarrow \quad \tau_\alpha > \frac{N_\alpha}{\pi}. \quad (5.6.3)$$

Additional constraints come from the moduli stabilization process. To use the single-field inflationary analysis we have developed in section 5.2 and section 5.3, we need to require that the uplifted minimum is only slightly perturbed by the inflationary dynamics. In particular, the linear potential that we have represented as $\mu^3\phi$ actually depends on the compactification volume, and hence shifts the minimized value of the volume. In four-dimensional Einstein frame, the leading term in the inflaton potential is

$$V(\phi, \mathcal{V}_E) \approx \left(\frac{\langle \mathcal{V}_E \rangle}{\mathcal{V}_E} \right)^2 \mu^3 \phi \quad (5.6.4)$$

where $\langle \mathcal{V}_E \rangle$ is the expectation value of the volume. To ensure that the resulting contribution to the potential for the volume is unimportant, we will insist that the inflaton potential induced by the NS5-brane, $V(\phi)$, is smaller than the moduli potential \mathcal{U}_{mod} .

At the supersymmetric minimum we have

$$V_{AdS} = -\frac{g_s}{2} \frac{3|W|^2}{\mathcal{V}_E^2}. \quad (5.6.5)$$

Without specifying the details of the uplifting mechanism, we assume that an uplifting to a small and positive cosmological constant is possible, and that the height of the potential barrier \mathcal{U}_{mod} that separates the uplifted minimum from decompactification is of the same order as $\mathcal{U}_{mod} \sim |V_{AdS}|$. Now, the COBE normalization tells us that

$$V(\phi_{CMB}) = \epsilon (0.027 M_{pl})^4 \simeq 2.4 \times 10^{-9} M_{pl}^4. \quad (5.6.6)$$

Hence we obtain the constraint

$$\frac{g_s}{2} \frac{3|W|^2}{\mathcal{V}_E^2} = |V_{AdS}| \simeq \mathcal{U}_{mod} \gg 2.4 \times 10^{-9} M_{pl}^4. \quad (5.6.7)$$

To extract a useful form of the above constraints, let us substitute for W the solution of any of the equations (5.5.11)

$$W = +|A_\alpha| a_\alpha e^{-a_\alpha \tau_\alpha} \frac{2\mathcal{V}_E}{v^\alpha}, \quad (5.6.8)$$

with no sum over α . We will also assume $|A_\alpha| \sim 1$ (see [151] for a discussion of this point). After some manipulations we find

$$\tau_\alpha \ll -\frac{N_\alpha}{2\pi} \log \left(N_\alpha 10^{-5} \frac{v^\alpha}{\pi \sqrt{g_s}} \right), \quad (5.6.9)$$

again with no summation over α . Finally, we should limit the number of D7-branes in each stack; although there plausibly exist examples with N_α quite large, we will impose $N_\alpha \leq 50$. This gives us

$$\tau_\alpha \ll 73 - 8 \log \left(\frac{v^\alpha \pi \sqrt{g_s}}{2g_s} \right). \quad (5.6.10)$$

We notice that $v^\alpha (\pi \sqrt{g_s}) > 1$ was the condition in (5.6.2) that enabled us to neglect α corrections, so that as long as $g_s \leq 0.5$ the second term on the right hand side of (5.6.10) is negative.

5.6.2 Constraints from backreaction on the geometry

Another important constraint comes from the requirement that the backreaction of the inflationary energy density on the compact space is small. In this section we will give a qualitative description of the problem and will briefly sketch a model-building solution; the interested reader is referred to appendix B for a more complete treatment.

At the time that the CMB perturbations are produced, the inflaton has a large vev in Planck units, $\phi \sim 11M_{pl}$, corresponding to a configuration of the two-form potential threading the two-cycle Σ_- of the form

$$\frac{1}{(2\pi)^2\alpha'} \int_{\Sigma_-} C_2 \equiv N_w = \frac{\phi}{2\pi f} \gg 1 \quad (5.6.11)$$

In the absence of an NS5-brane wrapping Σ , there would be no energy stored in this configuration, as C_2 enjoys a shift symmetry. However, inflation is driven by the substantial energy stored in this system by the Born-Infeld action of the wrapped NS5-brane. Moreover, there is a corresponding D3-brane charge induced by the Chern-Simons coupling $\int C_2 \wedge C_4$. Note that the net induced D3-brane charge in the total compactification is zero, as required by Gauss's law, because we have arranged for an additional, tadpole-canceling NS5-brane that wraps a distant cycle Σ'_- homologous to Σ_- , but does so with opposite orientation. Therefore, the Chern-Simons coupling induces a dipole configuration of D3-brane charge, with F_5 flux lines stretching from Σ_- to Σ'_- .

It is essential to ensure that the inflationary energy, which is effectively localized in the compact space in the vicinity of the wrapped NS5-brane, does not substantially correct the remainder of the compact geometry. Heuristically, one can imagine that the increased tension of the NS5-brane, as well as the induced charge, is represented by N_w D3-branes dissolved in the NS5-brane.

We must therefore estimate the effect of N_w D3-branes in a warped throat (recall that we have situated each wrapped NS5-brane in a warped region in order to suppress its energy density below the string scale, as required *e.g.* by the COBE normalization). Clearly, this backreaction will be reasonably small if $N_w \ll N$, with N the D3-brane charge of the background throat.

However, we must be careful about the effect of even a modest distortion of the geometry on the moduli stabilization and therefore on the four-dimensional potential. Let us first recall that in scenarios of D3-brane inflation in nonperturbatively-stabilized vacua, even a single D3-brane moving slowly in a throat can affect the warp factor, and correspondingly the warped volumes of four-cycles bearing nonperturbative effects, to such a degree that this interaction is the leading contribution to the inflaton potential [178],[179].

This sensitivity originates in two facts: first, D3-branes perturb the warped metric in a manner that is not suppressed by the background warp factor at the location of the D3-branes, because D3-branes are BPS with respect to a throat generated by D3-brane charge, and hence their contributions to the metric may simply be superposed on the background. Second, nonperturbative effects on a four-cycle are exponentially sensitive to changes in the four-cycle volume. Both these facts appear threatening for a situation such as ours in which the moduli are stabilized nonperturbatively and substantial D3-brane charge is induced in a throat: one can anticipate that as inflation proceeds and the D3-brane charge diminishes, the four-cycle volume changes, leading to an unanticipated, and possibly steep, contribution to the inflaton potential.

To understand this concretely, we will first consider a simpler system: an anti-D3-brane in a warped throat generated by N D3-branes, or equivalently

a warped throat generated by $N-1$ D3-branes, together with a brane-antibrane pair. Furthermore, from the result of [180] one learns that at long distances, the effect of the brane-antibrane pair on the supergravity solution is strongly suppressed by the warp factor at the location of the pair, *i.e.* at the tip of the throat. In contrast, the effects of D3-branes are not suppressed in this manner. Therefore, for the purpose of computing perturbations to the bulk compact space, we may replace an anti-D3-brane in a warped throat generated by N D3-branes with a warped throat generated by $N-1$ D3-branes, up to exponentially small corrections.

Equipped with this approximation, we may represent the configuration of interest as follows: two warped throats, carrying the charge of N_1, N_2 D3-branes respectively, are perturbed to $N_1 + N_w, N_2 - N_w$ by the inclusion of the NS5-brane in (say) the first throat, and the anti-NS5-brane in the second throat. Here we are ignoring the warping-suppressed correction indicated above, and we are approximating the NS5-branes by the D3-brane charge and tension that they carry, which is an excellent approximation for $N_w \gg 1$. Other effects due to the NS5-brane that do not depend on its induced D3-brane charge, *i.e.* on its world-volume flux, are independent of the inflaton and hence do not correct its potential. One can now easily see that the volume of a four-cycle at a generic location in the compact space will be corrected by the inclusion of the NS5-branes. If the four-cycle happens to enter one or both throats, the change in the volume is easily computed, and is seen to be substantial (*cf.* appendix B).

To control this problem, we situate the NS5-brane and the anti-NS5-brane, together with the family of homologous cycles connecting them, in a single warped region. The idea is that from the bulk of the compact space, the

NS5-brane configuration will appear to be a distant dipole whose net effect, integrated over a four-cycle, averages out to be small. This setup allows us to parametrically suppress the backreaction by a small factor given by the ratio of the dipole length, *i.e.* the distance between two NS5-branes, to the distance between the NS5-branes and the four-cycle in question. This small factor comes in addition to the suppression by the small ratio N_w/N .¹⁹

In appendix C we give more details about the above setup. We show, through two explicit models of increasing complexity, the robustness of the above suppression mechanism.

5.6.3 Constraints from higher-derivative terms

The analysis presented thus far has used the two-derivative action, which is an approximation with a limited range of validity. In general, one expects an infinite series of higher-derivative terms, possibly including multiple derivatives as well as powers of the first derivative. Our background solution involves rapid oscillations, so it is reasonable to ask whether these high frequencies enhance the role of higher-derivative terms and render the two-derivative approximation invalid. To check this, one should evaluate the higher-derivative terms on the solution and compare to the two-derivative action. We will now show that the two-derivative approximation is valid in the scalar sector; analogous considerations apply to the gravitational action.

Rather than write down the most general higher-derivative corrections to the scalar sector, we give here the terms that end up being present in the

¹⁹A further suppression can be achieved with a carefully-chosen embedding of the four-cycle, *e.g.* one that is symmetric with respect to the two NS5-branes. However, this requires fine-tuning, whereas the dipole suppression on which we have focused is parametric.

string theory examples. In string theory, we can directly compute the leading higher-derivative terms in the action for b , extending the result to c using S-duality. To get the leading terms, one considers the α'^3 corrections to the effective action due to Gross and Sloan [181] (at the four-point level) and Kehagias and Partouche [182] (up to the eight-point level). These corrections are of the same lineage as the famous Riemann⁴ term, but involve NS-NS three-form flux. This yields corrections to the axion kinetic terms. Following [182], the ten-dimensional Einstein-frame action including the leading (α'^3) corrections is

$$S_{10D,E} = \frac{1}{(2\pi)^7 \alpha'^4} \int d^{10}x \sqrt{g_E} \left(R_E - \frac{1}{12g_s} H_{KMN} H^{KMN} + \frac{\zeta(3)}{3 \times 2^6} g_s^{-3/2} \alpha'^3 \bar{R}^4 + \dots \right). \quad (5.6.12)$$

where

$$\bar{R}_{MN}{}^{PQ} = R_{MN}{}^{PQ} + \frac{1}{2} g_s^{-1/2} \nabla_{[M} H_{N]}{}^{PQ} - \frac{1}{4} g_s^{-1} H_{[M}{}^{C[P} H_{N]C}{}^{Q]} + \dots \quad (5.6.13)$$

and the square brackets are defined without the combinatorial factor 1/2 in front. Hence, the terms that are relevant for our axion at order α'^3 are proportional to H^8 and $(\nabla H)^4$. To estimate the importance of these terms, we will consider a special case in which the internal space is a $T^2 \times T^4$, with the NS-NS two-form field only along the T^2 directions 8 and 9, *i.e.* $B_{89} = -B_{98} = b$. Furthermore, since the background dynamics involves large frequencies but not large spatial gradients, we are primarily interested in terms containing only time derivatives, and can therefore take b to be homogeneous in the noncompact spatial directions. In this special case, making use of (2.13) in [181], and using S-duality to determine the action for c from that for b , we find that after

dimensional reduction the corrected action for c is

$$S_{4D} = \int d^4x \left[-M_p^2 \frac{g_s}{2} \dot{c}^2 g^{88} g^{99} + \frac{\zeta(3)}{2^6 g_s^{3/2}} \frac{\mathcal{V}_E^3}{\pi^3 M_p^4} \left(\frac{1}{2} \dot{c}^8 g_s^4 (g^{88} g^{99})^4 + \frac{1}{2^4} \dot{c}^4 g_s^2 (g^{88} g^{99})^2 \right) \right] \quad (5.6.14)$$

Now we use $\phi = cf$ to make the kinetic term canonical, yielding the action in terms of ϕ ,

$$\begin{aligned} S_{4D} &= \int d^4x \left[-\frac{1}{2} \dot{\phi}^2 + \frac{\zeta(3)}{2^6 g_s^{3/2}} \frac{\mathcal{V}_E^3}{\pi^3} \left(\frac{1}{2} \frac{\dot{\phi}^8}{M_p^{12}} + \frac{1}{2^4} \frac{\ddot{\phi}^4}{M_p^8} \right) \right] \\ &\equiv \int d^4x \left[-\frac{1}{2} \dot{\phi}^2 + \frac{\dot{\phi}^8}{M_I^{12}} + \frac{\ddot{\phi}^4}{M_{II}^8} \right], \end{aligned} \quad (5.6.15)$$

where we can now calculate the scale of the higher derivative terms M_I and M_{II} to be

$$M_I = M_p \frac{g_s^{1/8}}{\mathcal{V}_E^{1/4}} \left(\frac{\pi^3 2^7}{\zeta(3)} \right)^{1/12} \quad (5.6.16)$$

and

$$M_{II} = M_p \frac{g_s^{3/16}}{\mathcal{V}_E^{3/8}} \left(\frac{\pi^3 2^{10}}{\zeta(3)} \right)^{1/8} \quad (5.6.17)$$

To determine whether these higher-derivative terms will become important, we compute the dimensionless quantity $\frac{\omega}{M_{I,II}}$, where $\omega = \frac{\dot{\phi}}{f}$ (5.3.5) is the physical frequency of oscillations; we obtain

$$\frac{\omega}{M_I} \simeq 5 \times 10^{-3} \left(\frac{f}{10^{-3}} \right)^{-1} \left(\frac{g_s}{0.2} \right)^{-1/8} \left(\frac{\mathcal{V}_E}{120} \right)^{1/4}, \quad (5.6.18)$$

$$\frac{\omega}{M_{II}} \simeq 6 \times 10^{-3} \left(\frac{f}{10^{-3}} \right)^{-1} \left(\frac{g_s}{0.2} \right)^{-3/16} \left(\frac{\mathcal{V}_E}{120} \right)^{3/8}. \quad (5.6.19)$$

For the ranges of f and \mathcal{V}_E that will be of interest to us (*cf.* section 5.7), the higher-derivative terms are not important and our two-derivative approximation is justified.

5.6.4 Constraints on the axion decay constant

In this subsection, we discuss direct constraints on the axion decay constant f . We first recall a rather general (conjectured) upper bound $f < M_{pl}$ [183], and we then describe and incorporate a novel lower bound, specific to our setup, that arises from combining the requirements that α perturbation theory should be valid and that the inflationary energy should not drive decompactification.

Despite many attempts, at the time of writing there is no known, controllable string theory construction that provides $f > M_{pl}$. In particular, the authors of [183] have scanned several classes of string theory models and found sub-Planckian axion decay constants in every case. However, this upper bound on f is of relatively little importance for the phenomenological signatures we are considering in this chapter.

On the other hand, a potential lower bound on f is of considerable importance for our analysis. Considering oscillations in the CMB spectrum, in the regime $f \ll M_{pl}$ one can easily find models that range from being observationally excluded to giving undetectably small modifications, depending on the amplitude of the ripples in the inflationary potential. Furthermore, the resonant non-Gaussianity becomes large only for small f (*e.g.* we will find that $f < 3 \times 10^{-3}$ is a necessary condition to give a reasonable prospect of detectability). Hence we will move on to consider possible lower bounds on f .

As discussed in [151] and in the preceding section, a direct lower bound

on f comes from the requirement of small backreaction. In particular, the radius of curvature induced by the energy localized on the wrapped NS5-brane should be smaller than the smallest radius of curvature R_\perp in a direction transverse to the NS5-brane in the compactification. This requires

$$N_w \ll \frac{R_\perp^4 X}{4\pi g_s} \quad \Rightarrow \quad \frac{f}{M_{pl}} \gg \frac{2\phi g_s}{R_\perp^4 X}, \quad (5.6.20)$$

where we have defined $X \equiv \text{Vol}(X_5)/\pi^3$, with X_5 the base of the cone forming the warped throat. We remark that $X \leq 1$, as S^5 is the Sasaki-Einstein manifold with the largest volume, in the sense defined above. We can estimate R_\perp as being comparable to the AdS radius R of the throat containing the NS5-brane. Given that the volume²⁰ \mathcal{V} of the Calabi-Yau has to be larger than the volume of any throat it includes, one finds that

$$\mathcal{V} > V_{throat} = \frac{\pi^3}{2} X R^6, \quad (5.6.21)$$

where for simplicity we have assumed that the UV cutoff of the throat is at $r \sim R$ where the warp factor becomes of order unity. Putting together (5.6.20) and (5.6.21), we find

$$\frac{f}{M_{pl}} > \frac{\pi^2 2^{1/3} \phi g_s}{X^{1/3} \mathcal{V}^{2/3}} \simeq \frac{137 g_s}{X^{1/3} \mathcal{V}^{2/3}} = \frac{0.09}{X^{1/3} \mathcal{V}_E^{2/3}}. \quad (5.6.22)$$

Although the above constraint substantially restricts our parameter space, an even stronger constraint comes from demanding the validity of α perturbation theory: using (5.5.23) for f and combining this with the lower bound on two-cycle volumes given in (5.6.2), we obtain

$$\frac{f^2}{M_{pl}^2} = \frac{\sqrt{g_s}}{(2\pi)^3 \mathcal{V}_E} (c_{\alpha--} v^\alpha \sqrt{g_s} \pi) > \frac{\sqrt{g_s}}{(2\pi)^3 \mathcal{V}_E}, \quad (5.6.23)$$

²⁰We always refer to the warped volume, calculated with the whole warped metric.

where we have assumed that $c_{\alpha--} \geq 1$. (5.6.23) turns out to give the strongest microphysical lower bound on f . An upper bound is harder to determine from this formula. Assuming again that $c_{\alpha--} \geq 1$, assuming that no precise cancellations occur, and using (5.6.2), we find

$$\frac{f}{M_{pl}} < g_s \frac{\sqrt{3}}{2}. \quad (5.6.24)$$

5.6.5 Constraints on the amplitude of the modulations

So far we have seen that with the Kähler potential and superpotential given in (5.5.9) and (5.5.10), the axion c persists as a flat direction after moduli stabilization.²¹ As explained in [151], the presence of an NS5-brane wrapping the two-cycle that defines c introduces a monodromy and results, for large c , in the linear potential in (5.5.17). In this subsection we will consider further non-perturbative corrections that will in general induce small modulations of this linear potential. These are precisely the modulations whose phenomenology we have studied in the first part of this chapter.

Nonperturbative corrections could appear both in the Kähler potential and in the superpotential. We focus on the first possibility and comment at the end of this section on the second. Consider the type IIB orientifolds with O3-planes and O7-planes. As we have remarked, the RR two-form C_2 is odd under the orientifold projection and therefore a four-dimensional axion that survives projection comes from integrating C_2 over an odd two-cycle v^- . Such an odd cycle can be thought of as $v^- = v^1 - v^2$, where v^1 and v^2 represent two two-cycles in the parent Calabi-Yau manifold that are mapped into each other by

²¹As we have remarked, the axion b has its flat direction lifted by nonperturbative stabilization of the Kähler moduli.

the orientifold action. Now consider a Euclidean D1-brane wrapping the even cycle $v^+ = v^1 + v^2$. Such an instanton feels the local $\mathcal{N} = 1$ supersymmetry of the orientifolded theory, and it breaks this supersymmetry completely.²² Hence this is a non-BPS instanton with four universal fermionic zero modes, namely the goldstini of the broken $\mathcal{N} = 1$ supersymmetry. If the Euclidean D1-brane wraps a minimum-volume cycle in the homology class v^+ then it has the right total number of fermionic zero modes (four) to contribute to a D-term and in particular to the Kähler potential.

More specifically, in [185] it was argued that nonperturbative contributions from worldsheet instantons and their $SL(2, \mathbb{Z})$ images, Euclidean (p, q) strings, give rise to corrections to the prepotential of the $\mathcal{N} = 2$ theory of the parent Calabi-Yau compactification. Such corrections are most naturally expressed inside the logarithm of the Kähler potential,

$$K = -2 \log [\mathcal{V}_E + g(G, \bar{G})] , \quad (5.6.25)$$

where g is an appropriate function. Invariance under $SL(2, \mathbb{Z})$, or more generally under a subgroup $\Gamma \subset SL(2, \mathbb{Z})$, is naturally achieved if g is the sum of some individual correction \tilde{g} over an orbit of Γ .

At the time of writing, the nonperturbative correction g is not known explicitly, but a modular-invariant result has been conjectured in [186]. Inspired by the structure of this result (which we will not reproduce here), we will make a simple educated guess based on the following criteria: the nonperturbative correction should go to zero exponentially for large two-cycle

²²To see this, note that (*cf.* [184]) the instanton action depends on a two-cycle volume, but the proper Kähler coordinates are four-cycle volumes. Therefore, the instanton action cannot be holomorphic, so the instanton cannot contribute to a superpotential, and must instead be non-BPS.

volume v^+ ; it should break the continuous shift-symmetry of c to a discrete shift-symmetry $c \rightarrow c + 2\pi$; and it should be invariant under whatever discrete subgroup $\Gamma \subset SL(2, \mathbb{Z})$ of the ten-dimensional $SL(2, \mathbb{Z})$ symmetry is preserved by the compactification. The subgroup Γ may well be trivial, and we will assume this for simplicity; note, however, that one can plausibly obtain a more constrained result when some or all of the symmetry is preserved, as in [186]. Moreover, notice that along the orbits of Γ , the instanton action generally increases compared to that of a single worldsheet instanton or Euclidean D1-brane; thus, when the volume v^+ is not too small, only a few terms make an important contribution, with the remainder enjoying further exponential suppression.

A reasonable guess satisfying these criteria, for Γ trivial, is

$$K = -2 \log [\mathcal{V}_E + e^{-S_{ED1}} \cos(c)] = -2 \log \left[\mathcal{V}_E + e^{-\frac{2\pi v^+}{\sqrt{g_s}}} \cos(c) \right]. \quad (5.6.26)$$

In light of this corrected Kähler potential, we should revisit the moduli stabilization before proceeding to calculate the size b of the periodic contribution to the scalar potential.

We begin by noticing the following implication

$$\left\{ \begin{array}{l} D_{T_\alpha} W = \mathcal{O}(e^{-2S_{ED1}}) \\ D_{G^a} W = \mathcal{O}(e^{-S_{ED1}}) \\ W_{G^a} = 0 \end{array} \right\} \longrightarrow \left\{ \begin{array}{l} \partial_{T_\alpha} V = 0 + \mathcal{O}(e^{-2S_{ED1}}) \\ \partial_{G^a} V = -2 e^K |W|^2 K_{G^a} \\ \quad \quad \quad + \mathcal{O}(e^{-2S_{ED1}}) \end{array} \right\} \quad (5.6.27)$$

which can be verified by direct computation. This allows us to use the F-flatness condition to find the minimum in the T_α -directions even when one of the F-terms, namely $D_{G^-} W$, does not vanish. Equipped with this knowledge we repeat, *mutatis mutandis*, the steps of section 5.5.

First, the phases of the T_α are stabilized as in (5.5.13), with k being odd as explained below (5.5.14). The reason is that the sign of $\partial_{T_\alpha} K$ is not

changed by the small nonperturbative correction $e^{-S_{ED1}}$. Second, $\text{Im } G$ is again stabilized at 0. Given (5.6.27), the equation one needs to solve is $D_{G^a}W = 0$, which reduces to

$$0 = W \partial_{G^a} K \propto \partial_{G^a} \mathcal{V}_E - e^{-S_{ED1}} \left[\pi \sin(c) + \cos(c) \frac{2\pi}{\sqrt{g_s}} \partial_{G^a} v^+ \right] = 0, \quad (5.6.28)$$

where we made use of (5.5.6) to perform the derivative on c . Since \mathcal{V}_E and v^+ only depend on $\text{Im } G$ implicitly as in (5.5.22), we can take the imaginary part of (5.6.28),

$$\frac{1}{2} (\partial_{\text{Im } G^-} v^\alpha) c_{\alpha\beta\gamma} v^\beta v^\gamma - e^{-S_{ED1}} \cos(c) \frac{2\pi}{\sqrt{g_s}} \partial_{\text{Im } G^-} v^+ = 0. \quad (5.6.29)$$

But from (5.5.22) we know that

$$c_{\alpha\beta\gamma} (\partial_{\text{Im } G^-} v^\beta) v^\gamma = g_s c_{\alpha a} \text{Im } G^a, \quad (5.6.30)$$

which means that $\text{Im } G^a = 0$ (for every a) is a solution to (5.6.29). The real part of (5.6.28) is nonvanishing and of order $e^{-S_{ED1}}$. Again because of (5.6.27), the minimization in the τ_α is obtained by imposing $D_{T_\alpha}W = 0$. These equations depend on the inflaton c , appearing explicitly in (5.6.26), and hence the minimum in the T_α directions will be a function of c . Integrating out the T_α leads to a contribution in the effective potential $V[T(c), c]$ for c which is of the same order as the contribution coming from the explicit c -dependence in the Kähler potential. Therefore this effect cannot be neglected. To take it into account, we solve the $D_{T_\alpha}W = 0$ equations perturbatively in $e^{-S_{ED1}}$.

We define the coefficients of the minimum in the τ_α directions in a perturbative expansion in $e^{-S_{ED1}}$ by

$$\tau_{\alpha, \min} \equiv \tau_{\alpha, (0)} + \cos(c) e^{-S} \tau_{\alpha, (1)} + \dots, \quad (5.6.31)$$

and so on for all other variables. The zeroth-order equations are

$$(D_{T_\alpha} W)_{(0)} = (\partial_{T_\alpha} W)_{(0)} + W_{(0)}(\partial_{T_\alpha} K)_{(0)} = 0, \quad (5.6.32)$$

which can be solved numerically once the model is specified. The first-order equations are

$$(D_{T_\alpha} W)_{(1)} = (\partial_{T_\alpha} W)_{(1)} + W_{(1)}(\partial_{T_\alpha} K)_{(0)} + W_{(0)}(\partial_{T_\alpha} K)_{(1)} = 0, \quad (5.6.33)$$

which again can be solved numerically using the solutions of (5.6.32). We turn now to estimate the parameter b defined in (5.2.1). One finds

$$bf \equiv V_{(1)}\mu^{-3}e^{-S_{ED1}} = \frac{\mathcal{U}_{mod}\phi}{\mu^3\phi}e^{-S_{ED1}} \left(K_{(1)} + 2\text{Re} \frac{W_{(1)}}{W_{(0)}} \right), \quad (5.6.34)$$

where we have defined \mathcal{U}_{mod} as the moduli stabilization barrier at zeroth order in $e^{-S_{ED1}}$, *i.e.*

$$\mathcal{U}_{mod} = \frac{g_s}{2} \left(\frac{3|W|^2}{\mathcal{V}_E^2} \right)_{(0)}. \quad (5.6.35)$$

More explicitly, using (5.6.26) and (5.5.10),

$$\begin{aligned} bf &= \frac{\mathcal{U}_{mod}\phi}{2.4 \times 10^{-9} M_{pl}^4} e^{-S_{ED1}} \left[\frac{8\pi}{\sqrt{g_s}} \frac{(\partial_{T_\alpha} v^+)_{(0)}}{v_{(0)}^\alpha} - 2a_\alpha \tau_{\alpha,(1)} - \frac{2v_{(1)}^\alpha}{v_{(0)}^\alpha} \right] \\ &= \frac{\mathcal{U}_{mod}\phi}{2.4 \times 10^{-9} M_{pl}^4} 2e^{-S_{ED1}} \left[\frac{\sum_\beta (\partial_{T_\beta} W)_{(0)} \tau_{\beta,(1)}}{W_{(0)}} - \frac{\mathcal{V}_{E,(1)} + 1}{\mathcal{V}_{E,(0)}} \right] \end{aligned} \quad (5.6.36)$$

where the first line is valid for any α and the second line (obtained using (5.6.33)) shows that the expression for b is independent of α . Notice that $\partial_{T_\alpha} v^+ = \frac{1}{2} \partial_{\tau_\alpha} v^+$ is given implicitly by

$$c_{\alpha\beta\gamma} (\partial_{\tau_\rho} v^\gamma) v^\beta = \partial_{\tau_\rho} \tau_\alpha = \delta_\alpha^\rho. \quad (5.6.37)$$

Some comments are in order. The size of the ripples in the potential is proportional to the ratio of the moduli stabilization barrier to the scale of inflation,

which has to be large for the self-consistency of the estimate. We have used the value of the potential at the would-be AdS minimum to estimate the moduli stabilization barrier once an uplifting term is included. Due to the exponential suppression $e^{-S_{ED1}}$, the size of bf is extremely sensitive to g_s and v^+ .

An upper bound can be derived from (5.6.36) using the following considerations. In the KKLT construction, perturbative corrections to the Kähler potential can be neglected as long as $W_0 \ll 1$, and generically $W \sim W_0$. For larger values of W_0 , perturbative corrections have to be included, as in the large volume scenario [187]. In the present work, we focused on the former setup and we leave an investigation of the latter for the future. The exponential suppression in (5.6.36) can be bounded by (5.6.2). Finally, we denote the model-dependent term in square brackets in (5.6.36) by c_0 . Putting things together leads to the bound

$$bf < 2c_0 \times 10^7 \frac{g_s}{\mathcal{V}_E^2} e^{-2/g_s} \left(\frac{W}{0.1} \right)^2. \quad (5.6.38)$$

Even imposing all the model-independent constraints we have described in the previous sections, one can still have $bf > 10^{-4}$, which, as shown in section 5.4, is roughly the upper bound imposed by measurements of the scalar power spectrum. Therefore, in certain parameter ranges the primary constraint on modulations of the potential comes from the data, not from microphysics.

5.7 Combined Theoretical and Observational Constraints

We now summarize our results, combining the observational constraints from section 5.4 with the theoretical constraints from section 5.6. As an aid to the reader, we will now briefly recall the qualitative properties of those results.

Axion monodromy can produce characteristic signatures in the CMB: the oscillations in the axion potential generated by nonperturbative effects source resonant contributions to the scalar power spectrum and bispectrum. The amplitude and frequency of the oscillations in the potential can therefore be bounded by comparison to observations. We recall from section 5.4 that the observational constraints take the form of exclusion contours in the space of the phenomenological parameters, after marginalization over additional model parameters that have important degeneracies with those displayed. For convenience, we have chosen to display constraints in terms of the parameters f and bf defined in (5.2.1), marginalizing over the phase $\Delta\phi$ and over $\Omega_b h^2$.

The first new step is to combine the exclusion contours based on the temperature two-point function with estimates of constraints from the three-point function. Based on the rough estimates described in section 5.3.3, we present, in figure 5.6, three contours at $f_{res} = 200, 20, 2$, with the expectation that the gray region ($f_{res} > 200$) might plausibly be excluded, while the colored, lighter regions ($20 < f_{res} < 200$ and $2 < f_{res} < 20$, respectively) are possibly within detectability. A careful study of the constraints on resonant non-Gaussianity would be a worthwhile topic for future research.

Next, we recall that in section 5.6, we found that the requirements of consistency and computability in the string compactifications giving rise to axion monodromy models led to constraints on the parameters f and bf . Let us remark that as these constraints are not rooted in deep principles of string theory or of quantum field theory, but rather originate in practical limitations in our present ability to construct computable models, they may well be loosened in further work. As such, the theoretical constraints we present here should be understood as designating *included* rather than excluded

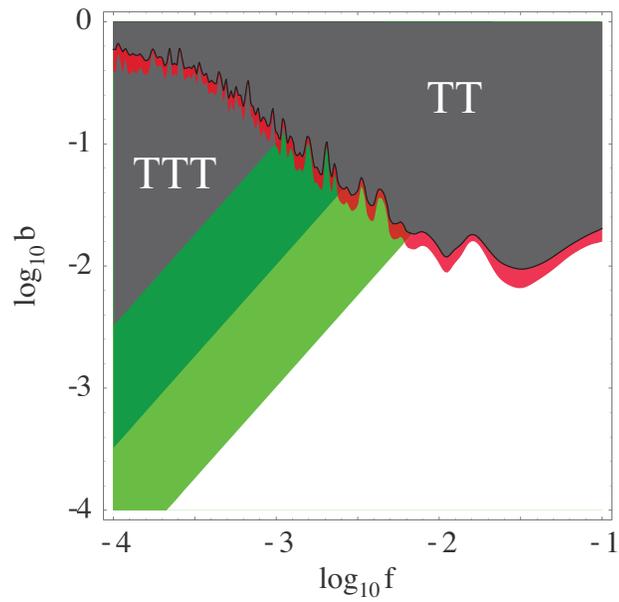


Figure 5.6: We show the (one- and two-sigma) likelihood contours for the temperature two-point function together with three contours that characterize the amplitude of the three point function, for $f_{res} = 200, 20, 2$.

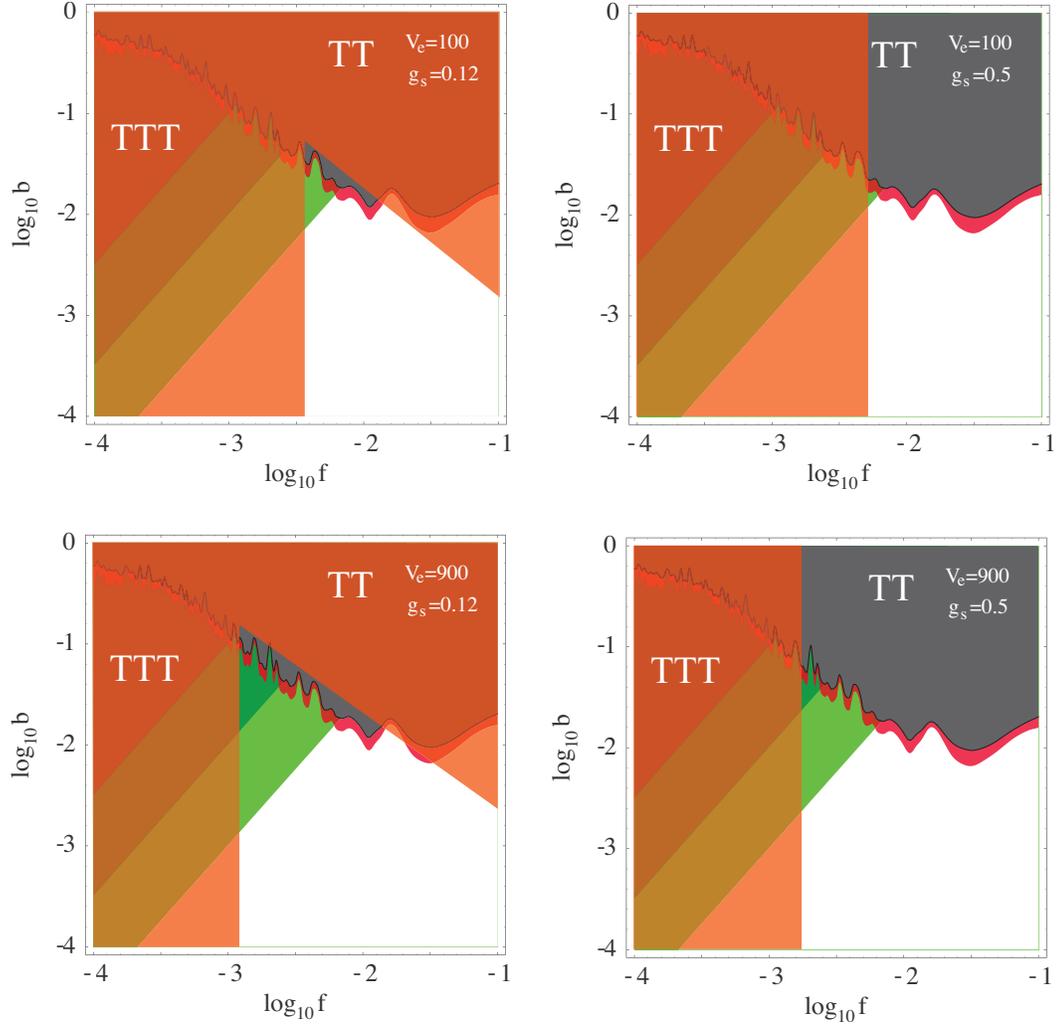


Figure 5.7: We superimpose the theoretical constraints, summarized in (5.7.1) and (5.7.2), on the constraints imposed by observations, which are shown in figure 5.6. The orange overlay indicates regions of the parameter space that are difficult to reach in the class of models considered in the present work. The theoretical constraints are shown for $g_s = 0.12, 0.5$ and $\mathcal{V}_E = 100, 900$.

regions: in contrast to experimental contours, theoretical contours of this sort may expand rather than contract given improved understanding.

Because the parameter b measures the amplitude of a nonperturbative effect, it is exponential in the natural input parameters, and can therefore be made small without substantial fine-tuning. We therefore do not present a lower bound on b . However, we found the theoretical upper bound (5.6.38)²³

$$bf < 2c_0 \times 10^9 M_{pl}^4 \frac{g_s}{\mathcal{V}_E^2} e^{-2/g_s}, \quad (5.7.1)$$

with a model-dependent constant c_0 that can be estimated in explicit examples, and which we find to be typically of order 10^{-2} .

Next, we obtained a lower bound for f in (5.6.23).²⁴ A precise upper bound, however, is highly model-dependent. We estimate an upper limit by assuming that the intersection numbers are of order one²⁵ and that no precise cancellations occur. From (5.6.23) and (5.6.24), the complement of the theoretically excluded range for f is then

$$\frac{g_s^{1/4}}{(2\pi)^{3/2} \sqrt{\mathcal{V}_E}} < f < g_s \frac{\sqrt{3}}{2}. \quad (5.7.2)$$

Notice that the theoretical constraints depend mainly on two quantities: the string coupling g_s and the volume \mathcal{V}_E of the compactification. The former appears in the exponential suppression of the nonperturbative effect

²³We stress that this ‘bound’ is *not* universal and depends on the assumptions enumerated in section 5.6. We include it here as a representative example of the constraints that arise in particular scenarios.

²⁴The constraint from the backreaction described in subsection 5.6.2 is weaker than (5.6.23).

²⁵Larger intersection numbers are an interesting possibility that we will not investigate here.

generating the modulations of the potential. Hence, $g_s \lesssim 0.1$ suppresses any possible signature of the modulations. For $g_s \gtrsim 0.1$, there is always a theoretically allowed region in which the oscillations in the inflaton potential lead to observable ripples in the two-point function of the CMB. On the other hand, the size of the non-Gaussianity depends critically on \mathcal{V}_E as well. Assuming $g_s \gtrsim 0.1$, larger \mathcal{V}_E allows for a larger range of f and therefore larger non-Gaussianity (see (5.7.2)). A way to quantify this is to use the estimate obtained in section 5.3.3,

$$f_{res} \simeq \frac{9}{4} \frac{b}{(\phi f)^{3/2}}, \quad (5.7.3)$$

and the lower bound in (5.7.2). The result is

$$\mathcal{V}_E > 170 \left(\frac{g_s}{0.2}\right)^{1/2} \left(\frac{f_{res}}{10}\right)^{4/5} \left(\frac{10^{-4}}{bf}\right)^{4/5}. \quad (5.7.4)$$

We now combine the theoretical and observational constraints, presenting them in the plane $\{\log_{10} f, \log_{10} b\}$. We choose as boundaries $10^{-4} < f \ll M_{pl}$ and $10^{-4} < b \ll 1$ based on the following considerations. The number of oscillations per e-folding is roughly $10^{-2} M_{pl}/f$. Hence for $f \gtrsim 0.1 M_{pl}$ there is less than one oscillation in the whole range of scales probed by the CMB, and the signal from modulations becomes degenerate with the overall amplitude. Furthermore, in section 5.3, we systematically used the expansion $b \ll 1$, where $b = 1$ divides monotonic from non-monotonic potentials. Finally, the lower boundaries $10^{-4} < f$ and $10^{-4} < b$ exclude regions that are relatively uninteresting in the present context: smaller values of b lead to an unobservably small signal, while smaller values of f are rather difficult to obtain in the class of string theory constructions we considered. In the $\{\log_{10} f, \log_{10} b\}$ plane, the theoretically allowed region looks like an interval in f , whose size is determined by \mathcal{V}_E , with an upper cut effectively determined by g_s as in (5.7.1).

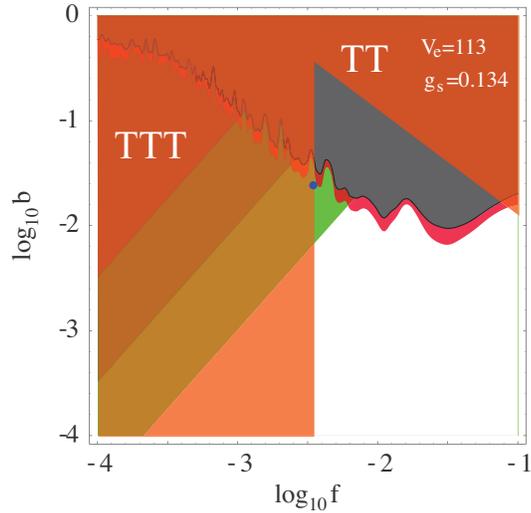


Figure 5.8: The blue dot represents the explicit numerical example presented in appendix E. It represents a case in which upcoming experiments could detect the signatures of modulations in both the two-point function and the three-point function.

Finally, in figure 5.8 we show where a particular numerical toy example, with specific choices of the intersection numbers, lies in the $\{\log_{10} f, \log_{10} b\}$ plane.

5.8 Summary

The goal of this investigation was to characterize the predictions of axion monodromy inflation for the CMB temperature anisotropies. Nonperturbative effects in these models generically introduce sinusoidal modulations of the inflaton potential, which in turn lead to resonantly-enhanced modulations of the scalar spectrum and bispectrum.

We have provided a simple analytic result for the modulated scalar

power spectrum in this class of models. We also presented an alternative derivation focussing on a resonance between a mode inside the horizon and the driving force of the oscillatory background evolution. We then determined in detail the constraints that the five-year WMAP data places on models with modulations of this sort.

Next, after reviewing the realization of axion monodromy inflation in string theory, we performed a comprehensive study of the parameter constraints implied by the requirements of microphysical consistency and computability. The resulting allowed parameter regions are very plausibly realizable in sensible string theory constructions.

We also identified a new contribution to the inflaton potential in axion monodromy inflation: the backreaction of the inflationary energy on the compact space can source an important correction to the potential by correcting the volumes of four-cycles and hence affecting the scale of nonperturbative moduli-stabilizing effects. We then presented a model-building solution to this problem, in which the NS5-brane and anti-NS5-brane driving inflation are in the same warped region, or more generally are distant from the four-cycles of interest.

Finally, we combined the observational and theoretical constraints, in order to ascertain whether detectable modulations of the scalar spectrum and/or bispectrum are possible, consistent with current observational bounds and known theoretical restrictions. Our conclusion is that both sorts of modulations are possible, and in fact in many cases the strongest bound on the amplitude of the modulations comes from data, not from microphysics. Moreover, even though observational limits on the amplitude and frequency of modulations in the scalar power spectrum provide a strong constraint on the

parameter space of axion monodromy models, and even though microphysical constraints sharply restrict the allowed frequencies, detectably-large non-Gaussianity can indeed be produced in a class of controllable models. Such models enjoy three nontrivial signatures: detectable tensors with $r \approx 0.07$, a modulated scalar power spectrum, and resonant non-Gaussianity.

Let us remark that even in the absence of non-Gaussianity, this class of models is eminently testable: axion monodromy inflation unambiguously predicts a large tensor signal, and the parameters of the models are already strongly constrained by limits on modulations in the scalar power spectrum.

There are several interesting directions for future work. First, we have not analyzed the constraints on the model from the three-point function; more generally, understanding the prospects for constraining or detecting resonant non-Gaussianity is an important task. Moreover, it would be instructive to construct an explicit model in which the many theoretical constraints we have checked can be combined in a coordinated way. In addition, it would be interesting to determine whether chain inflation can be realized in this context.

It is intriguing that the modulated power spectrum we have found is very similar in form to that proposed in the context of modifications of the initial state, as in *e.g.* [166],[167],[153],[154],[155]. In light of our calculation of the power spectrum in the saddle point approximation in subsection 5.3.2, this coincidence is not entirely surprising, and could be made more transparent by repeating the derivation in the interaction picture. In that case, the driving force of the oscillating background eventually generates an excited state, even if one begins in the Bunch-Davies vacuum. We leave a more systematic exploration of this connection for the future.

Finally, it would be most valuable to develop a broader understanding

of the connection, if any, between symmetries and signatures in models of large-field inflation.

Chapter 6

On Slow-Roll Moduli Inflation in Massive IIA Supergravity with Metric Fluxes

In this chapter, we derive several no-go theorems in the context of massive type IIA string theory compactified to four dimensions in a way that, in the absence of fluxes, preserves $\mathcal{N} = 1$ supersymmetry. The derivation is based on the dilaton, Kähler and complex structure moduli dependence of the potential of the four-dimensional effective field theory, that is generated by the presence of D6-branes, O6-planes, RR fluxes, NSNS 3-form flux, and geometric fluxes. To demonstrate the usefulness of our theorems, we apply them to the most commonly studied class of toroidal orientifolds. We show that for all but two of the models in this class the slow-roll parameter ϵ is bounded from below by numbers of order unity as long as the fluxes satisfy the Bianchi identities, ruling out slow-roll inflation and even the existence of de Sitter extrema in these models. For the two cases that avoid the no-go theorems, we provide some details of our numerical studies, demonstrating that small ϵ can indeed be achieved. We stress that there seems to be an η -problem, suggesting that none of the models in this class are viable from a cosmological point view at least at large volume, small string coupling, and at leading order in the α' -expansion.

The increasing precision of cosmological data collected over the past two decades has strengthened the case for inflation as the paradigm for early

universe cosmology. Not only is the universe observed to be flat to within one per cent, the density perturbations are also found to be nearly scale-invariant, Gaussian, and adiabatic, just as predicted by inflation. Conclusive evidence for inflation has not yet been found, but future experiments may well strengthen the case for inflation further.

The UV sensitivity of inflation has sparked work by many people aiming toward an understanding of what string theory predicts in this regard, as well as work toward an understanding of the reliability of the approximations made to derive these predictions. Given that we do not understand why space-time has four large dimensions, the best we can do is to take this as an input and construct a reliable four dimensional effective theory, with stabilized moduli, that produces inflation consistent with current observations in a way that allows us to control the corrections. Most of the modern approaches are based on flux compactifications. As explained in [188], one constructs a string inflation model by specifying the compactification manifold, its dimensionality and topology, the location of orientifold planes and D-branes, the type and amount of fluxes that are turned on and through which cycle. The size of the corrections will depend, among other things, on the value of g_s , which controls the loop expansion, α' , which sets the scale of the string modes, and the volume of the compactification manifold, which determines the mass scale for the Kaluza-Klein modes.

Since the ground breaking work of KKLT [170], which provided a mechanism to stabilize all moduli in the context of type IIB string theory, a lot of work has been done toward understanding inflation in string theory setups that are under control [189], [190], [191], [192], [193], [194], [188]. This effort has uncovered a variety of potential mechanisms for inflation which are usu-

ally broadly classified according to the origin of the inflaton field into moduli inflation or D-brane inflation. A lot of progress has been made in the context of type IIB string theory; type IIA on the other hand still remains much less explored.

In type IIA, work by DeWolfe et al. [195] gave an explicit construction for moduli stabilization relying on perturbative effects alone, thus making it more reliable as far as calculability is concerned and allowing for more explicit constructions than in type IIB. All is not good, however, since the presence of orientifolds in these compactifications may invalidate the usual effective field theory treatment as pointed out by Banks and van den Broek [196]. Although, we lack a good argument to appease those concerned by the latter point, in this work we will continue to ignore the backreaction of the orientifold plane in our search for inflation in massive type IIA, following Hertzberg et al. [197], [198]. The concerned reader may view our work as an attempt to derive some expertise and insight into the analysis of inflation in models with a large number of fields and fluxes that may also be useful in other contexts.

Generically, the various moduli fields are coupled to each other, and one has to search for slow-rolling regions in the typically rather high-dimensional moduli space. Because of the large number of fields and the complexity that comes with it, the existence of a slow-roll path will typically either be excluded by analytical methods (no-go theorems) or will have to be confirmed by numerical analysis.

The original work of HTKSÖ [197] studied three simple orientifolds of T^6 [195], [199], [200] in the hope to find inflation, only to be disappointed, and HKTT [198] extended this work and proved a no-go theorem that applies to compactifications of IIA string theory on general Calabi-Yau manifolds with

standard NSNS fluxes, RR fluxes, D6-branes, and O6-planes at large volume and small string coupling. This no-go theorem shows that the slow-roll parameter ϵ is bounded from below by some positive number independent of the choice of fluxes, implying that ϵ cannot be made small enough to allow for slow-roll inflation. This illustrates that even an infinite number of vacua does not guarantee that one of them will inflate.

As with any no-go theorem, the most important information contained in the theorem are the assumptions that went into its derivation. Beyond the usual assumptions of large volume, small string coupling, and leading order in the α' -expansion, an important assumption in the case of HKTT was that the models did not include geometric or non-geometric fluxes. In this work we allow for geometric fluxes, but otherwise use the same assumptions as [198]. In addition, we assume that there is a hierarchy between the mass scales corresponding to the twisted and blow-up modes and the modes of the untwisted sector that we keep. For those D6-branes whose backreaction cannot be ignored, we limit ourselves to rigid embeddings, making a hierarchy between the open string modes and the modes we keep plausible so that they can be integrated out.¹ For branes that can consistently be treated as probes, this limitation does not apply. Under these assumptions, we derive several new no-go theorems, and, to demonstrate their usefulness, apply them to toroidal orientifolds with abelian orbifold groups generated by rotations and reflections, that, in the absence of fluxes and after orientifolding, preserve $\mathcal{N} = 1$ supersymmetry. We show that in all these models except for the two $\mathbb{Z}_2 \times \mathbb{Z}_2$ cases, even in the presence of geometric fluxes, the slow-roll parameter ϵ is bounded

¹Phrased differently, the hierarchy guarantees that the branes will relax to their static configurations on time scales much shorter than the time scales we are interested in so that their dynamics can be ignored.

from below by numbers of order unity as long as the fluxes satisfy the Bianchi identities. For these two special cases, we numerically succeed in identifying regions in moduli space with arbitrarily small ϵ , but all these regions seem to have too short a period of inflation. In other words, there is an η -problem similar to the one that has already appeared in many of the type IIB compactifications, with the additional feature that at least one of these directions always turns out to be not only steep but also tachyonic.

While the original work of HKTT [198] was based exclusively on the dependence of the potential on the dilaton and volume moduli field our work is based on the dependence of the potential on the dilaton, volume as well as the Kähler and complex structure moduli. No-go theorems of this kind are useful because they sharpen our understanding of the ingredients necessary for successful slow-roll inflation and allow us to exclude entire regions in the large landscape of solutions of string theory. The analysis of HKTT shows that localized sources such as NS5-branes or geometric or non-geometric fluxes are a necessary condition for slow-roll in type IIA string theory, we show that geometric fluxes alone are not generally sufficient.

There have by now been other studies of type IIA compactifications that have successfully identified regions of slow-roll inflation or found de Sitter vacua.

In [201], Silverstein constructed de Sitter solutions based on compactifications of type IIA on a product of two Nil-manifolds with orientifold planes, fivebranes, fractional Chern-Simons forms, and fluxes. As was shown by Silverstein and Westphal [150], in these setups it is not only possible to realize successful slow-roll inflation, but even to realize large-field inflation in a controlled way thus generating an observable tensor signal [150], making these the

most exciting solutions in type IIA from a cosmological point of view so far.

While these vacua are presumably much closer to what one might expect a generic vacuum to look like, and one might argue that these are a natural place to start looking for inflation, they clearly are also somewhat less explicit and hence less controlled. So it seems worthwhile asking what the minimal set of ingredients for successful inflation or de Sitter solutions in type IIA is. Our studies are very much in this spirit, and it is also in this spirit that Haque et al. [202] studied compactifications on a product of two maximally symmetric hyperbolic spaces. Another possible approach to evading the no-go theorem of [198], complementary to the work here, is to appeal to corrections to the tree-level effective action. This is the direction pursued by the authors of [203].

Very recently, compactifications of type IIA on orientifolds of SU(3)-structure manifolds with non-vanishing geometric fluxes were studied in [204]. Some of the models were ruled out in [204] based on our results, for others numerical studies indicate that there are regions in moduli space with small ϵ . However, the same η -problem seems to make an appearance, ruling these out from a cosmological perspective as well. This seems natural because the models studied there are in fact close cousins of the models that lead to small ϵ and an η -problem for us.

The organization of the chapter is as follows. In section 2.1 and 2.2, we review in some detail the low energy effective theory in the presence of both geometric and non-geometric fluxes. In section 2.3, we write explicit expressions for the slow-roll parameters ϵ and η as functions of the moduli and the fluxes. In section 3, we derive several no-go theorems based on the dependence of the effective potential on the dilaton, as well as Kähler and

complex structure moduli. In section 4, we present a classification of the possible orientifolds of T^6 as well as the possible constraints on the fluxes, Bianchi identities and tadpoles conditions. In section 5, we apply the no-go theorems to the toroidal orientifolds discussed in section 4 and show that most of them cannot have small ϵ , ruling out both slow-roll inflation and the existence of de Sitter extrema. In section 6, we present some details of our numerical studies for the models for which we found small ϵ . We note that there seems to be an η problem, implying that none of the models in this class can accommodate extended periods of inflation as seems to be required by observations. We conclude in section 7. Appendix A contains a summary of our conventions.

This chapter is based on [205]

6.1 Low Energy Theory

6.1.1 Metric and non-geometric fluxes

As mentioned above, recent work [198] has shown that type IIA string theory with only ordinary fluxes (H -flux and RR fluxes) is not sufficient to allow for slow-roll inflation. For this reason we would like to include some extra ingredients in an effort to overcome this obstacle. There are many objects that one could add, and arguments that generically their presence will allow for de Sitter extrema [201], but we would like to look instead for a minimal set of additional ingredients, and we will focus on one particular class of ingredients which are known as generalized NSNS fluxes. Our motivation is that T-duality guarantees that these fluxes can appear (they should be on the same footing as ordinary H -flux), and T-duality also shows us how they must appear for consistency.

It is well-known that by T-dualizing a circle that is threaded by H -flux (that is if the circle isometry contracted with H is non-zero), one obtains a new solution in which some components of H -flux have been exchanged for some non-constant metric components, whose effect can be thought of as a twist of the circle over the rest of the geometry. These twists can be encoded in components f_{jk}^i , analogous to the individual components H_{ijk} of the H -flux. By performing an explicit Kaluza-Klein reduction from ten dimensions, one can learn how such objects appear in the low-energy theory in four dimensions [206], [207], [208]. It turns out that these objects, which are usually called metric fluxes (or sometimes geometric fluxes) because of the analogy with H , appear in the low-energy theory in much the same way that H does. If we started with an underlying space preserving $\mathcal{N} = 1$, such as a Calabi-Yau orientifold of IIA, then the objects appear as parameters in the four-dimensional superpotential and the tadpole constraints. It turns out that metric fluxes can also have the effect of giving a charge to some of the fields (fields that were RR axions before fluxes were turned on) under the four-dimensional vector multiplets, with the result that they can also appear in D-terms in four dimensions [209], [210].

Sometimes there are further T-dualities one can perform, converting the metric flux components f_{jk}^i into new objects Q_k^{ij} known as non-geometric fluxes. In the presence of these, the six-dimensional compact space is not a geometric manifold anymore, but rather it has the structure of a torus fiber glued over a base with transition functions that sit inside the full T-duality group [211], [212]. In fact, from a four-dimensional perspective, it is quite reasonable to expect a full set of these objects, H_{ijk} , f_{jk}^i , Q_k^{ij} , as well as objects labeled R^{ijk} , but not all of these have been explicitly constructed

from a ten-dimensional string theory. However, that doesn't necessarily stop us from discussing how they would appear in the low-energy theory, since that is determined by symmetry considerations [213], [214], [215], [209], [210]. For a review of these constructions, the reader is encouraged to refer to the review [216], and references therein.

Here, we will be satisfied to include a few more comments about metric fluxes. Consider the case of T^6 in particular (this will be the starting point for all of our explicit constructions later). In the presence of f_{jk}^i , the underlying geometry changes from a torus to a *twisted torus*, and the globally defined one-forms are no longer the closed forms dx^i , but rather forms η^i that satisfy

$$d\eta^i = -\frac{1}{2}f_{jk}^i\eta^j \wedge \eta^k. \quad (6.1.1)$$

Simply demanding that $d^2 = 0$ gives us some constraints,

$$f_{j[k}^i f_{\ell m]}^j = 0. \quad (6.1.2)$$

Similarly, if we now expand H in this basis, $H = \frac{1}{6}H_{ijk}\eta^i \wedge \eta^j \wedge \eta^k$, then the usual condition that H be closed gives

$$f_{[jk}^i H_{\ell m]i} = 0. \quad (6.1.3)$$

These two sets of equations will be referred to as Bianchi identities, and the flux components we turn on must satisfy them for reasons of consistency. Finally, it turns out to be much more convenient to express our fluxes instead using a basis of forms that, in the absence of metric fluxes, would be the harmonic forms of the underlying space (see Appendix A.2 for our conventions),

$$H = p_K b^K, \quad d\omega_a = -r_{aK} b^K, \quad d\mu_\alpha = -\hat{r}_\alpha^K a_K. \quad (6.1.4)$$

Here p_K , r_{aK} , and \hat{r}_α^K are linear combinations of H_{ijk} and f_{jk}^i . More details can be found in section 6.3.2.

Throughout this work we will only turn on H -flux and metric fluxes, and not any of the non-geometric fluxes, since we generically expect that in the presence of non-geometric fluxes, some volume moduli will be stuck near the string scale (since the transition functions involve T-dualities that include volume inversions). However, it will be useful to keep the non-geometric fluxes in our minds when thinking about using T-dualities to convert one configuration of fluxes into a more useful configuration, as discussed in section 6.3.3.

6.1.2 Effective potential

Our starting point is a Calabi-Yau orientifold of type IIA string theory. We will add RR fluxes, as well as H -flux and metric flux from the NSNS sector. Our conventions are listed in Appendix A.2.

These ingredients then lead to an effective $\mathcal{N} = 1$ supergravity theory in four dimensions. To describe this effective theory, and particularly the effective potential for the complex scalar fields t^a and N^K , we must provide the Kähler potential K , the holomorphic superpotential W , the holomorphic gauge kinetic couplings $f_{\alpha\beta}$, and the gauge transformations of the scalar fields under the different $\mathcal{U}_{mod}(1)$ gauge groups arising from the four-dimensional vectors (i.e. we must give the electric and magnetic charges of the scalar fields). Then the effective action for the scalars is

$$S = - \int \left\{ K_{a\bar{b}} dt^a \wedge *d\bar{t}^{\bar{b}} + K_{I\bar{J}} dN^I \wedge *d\bar{N}^{\bar{J}} + V * 1 \right\}, \quad (6.1.5)$$

where the scalar potential is

$$V = e^K \left(K^{a\bar{b}} D_a W \overline{D_b W} + K^{I\bar{J}} D_I W \overline{D_J W} - 3|W|^2 \right) + \frac{1}{2} (\text{Re } f)^{-1\alpha\beta} D_\alpha D_\beta. \quad (6.1.6)$$

Here, $*$ is the four-dimensional Hodge star, $K_{a\bar{b}} = \frac{\partial}{\partial t^a} \frac{\partial}{\partial t^{\bar{b}}} K$, $K^{a\bar{b}}$ is its (transpose) inverse, $D_a W = \frac{\partial}{\partial t^a} W + (\frac{\partial}{\partial t^a} K)W$, and similarly for the N^K , and the D-terms are

$$\begin{aligned} D_\alpha &= \frac{i}{W} (\partial_\alpha t^a D_a W + \partial_\alpha N^K D_K W) \\ &= i (\partial_\alpha t^a \partial_a K + \partial_\alpha N^K \partial_K K) + i \frac{\partial_\alpha W}{W}, \end{aligned} \quad (6.1.7)$$

where $\lambda^\alpha \partial_\alpha \phi$ is the variation of the field ϕ under an infinitesimal gauge transformation $A^\alpha \rightarrow A^\alpha + d\lambda^\alpha$. One can also discuss D-terms arising from the magnetic gauge groups, but the details are similar.

For the IIA orientifolds at hand, we can provide this information [217]. The Kähler potential is given by²

$$K = 4D - \ln \left(\frac{4}{3} (\kappa v^3) \right). \quad (6.1.8)$$

In the sector of Kähler moduli, this leads to

$$\partial_a K = \frac{3i}{2} \frac{(\kappa v^2)_a}{(\kappa v^3)}, \quad (6.1.9)$$

and

$$K_{a\bar{b}} = \frac{9}{4} (\kappa v^3)^{-2} (\kappa v^2)_a (\kappa v^2)_b - \frac{3}{2} (\kappa v^3)^{-1} (\kappa v)_{ab}, \quad (6.1.10)$$

²Here and in some formulae below we will sometimes drop certain indices in cases where the contractions are obvious. For instance, (κv^3) will be short hand for $\kappa_{abc} v^a v^b v^c$. We hope that this does not cause the reader too much difficulty.

with inverse

$$K^{a\bar{b}} = -\frac{2}{3} (\kappa v^3) (\kappa v)^{-1ab} + 2v^a v^b. \quad (6.1.11)$$

Note that there is a no-scale condition

$$K^{a\bar{b}} \partial_a K \overline{\partial_b K} = 3. \quad (6.1.12)$$

In the complex structure sector, our moduli are defined by $N^K = \frac{1}{2}\xi^K + ie^{-D}\mathcal{Z}^K$, where D is the four-dimensional dilaton and the \mathcal{Z}^K come from the expansion of the holomorphic three-form Ω . These \mathcal{Z}^K are not all independent (there are $h^{2,1} + 1$ of them which are functions of the $h^{2,1}$ complex structure moduli), and in fact they satisfy a relation which can always be written as

$$p_n(\mathcal{Z}) = 1, \quad (6.1.13)$$

where $p_n(\mathcal{Z})$ is a homogeneous polynomial of degree $n = h^{2,1} + 1$ in the \mathcal{Z}^K . In terms of this polynomial (which of course plays the role of a prepotential) we can then write

$$\mathcal{F}_K = -\frac{i}{2n} \frac{\partial}{\partial \mathcal{Z}^K} p_n(\mathcal{Z}), \quad (6.1.14)$$

and

$$D = -\frac{1}{n} \ln [p_n(I)], \quad (6.1.15)$$

where $I^K = \text{Im } N^K = e^{-D}\mathcal{Z}^K$. It then follows that

$$\frac{\partial}{\partial N^J} K = \frac{2i}{n} \frac{\partial_J p_n(I)}{p_n(I)} = \frac{2i}{n} e^D \partial_J p_n(\mathcal{Z}) = -4e^D \mathcal{F}_J, \quad (6.1.16)$$

$$K_{J\bar{K}} = \frac{1}{n} e^{2D} [\partial_J p_n(\mathcal{Z}) \partial_{\bar{K}} p_n(\mathcal{Z}) - \partial_J \partial_{\bar{K}} p_n(\mathcal{Z})]. \quad (6.1.17)$$

It can be useful to pull out the dilaton dependence here and define

$$\widehat{K}_{JK} = e^{-2D} K_{J\bar{K}} = \frac{1}{n} [\partial_J p_n(\mathcal{Z}) \partial_{\bar{K}} p_n(\mathcal{Z}) - \partial_J \partial_{\bar{K}} p_n(\mathcal{Z})]. \quad (6.1.18)$$

The inverse of \widehat{K}_{JK} will simply be denoted by \widehat{K}^{KL} , so that $K^{KL} = e^{-2D} \widehat{K}^{KL}$.

With these results it is easy to verify the identities

$$\widehat{K}_{JK} \mathcal{Z}^K = 2i \mathcal{F}_J, \quad \widehat{K}^{JK} \mathcal{F}_K = -\frac{i}{2} \mathcal{Z}^J, \quad (6.1.19)$$

$$\mathcal{Z}^K \mathcal{F}_K = -\frac{i}{2}, \quad \widehat{K}^{JK} \mathcal{F}_J \mathcal{F}_K = -\frac{1}{4}, \quad \widehat{K}_{JK} \mathcal{Z}^J \mathcal{Z}^K = 1, \quad (6.1.20)$$

and the no-scale-type condition

$$K^{J\bar{K}} \partial_J K \overline{\partial_{\bar{K}} K} = 4. \quad (6.1.21)$$

The gauge kinetic couplings are

$$f_{\alpha\beta} = i (\widehat{\kappa} t)_{\alpha\beta}, \quad (6.1.22)$$

with

$$(\text{Re } f)^{-1} = -(\widehat{\kappa} v)^{-1}. \quad (6.1.23)$$

The corresponding D-terms are

$$D_\alpha = 2ie^D (\widehat{r} \mathcal{F})_\alpha, \quad (6.1.24)$$

so that the D-term contribution to the potential is

$$V_D = \frac{1}{2} (\text{Re } f)^{-1\alpha\beta} D_\alpha D_\beta = 2e^{2D} (\widehat{\kappa} v)^{-1} (\widehat{r} \mathcal{F})^2. \quad (6.1.25)$$

Unlike the Kähler potential, the superpotential depends on the fluxes,

$$W = e_0 + te + \frac{1}{2} \kappa t^2 m + \frac{1}{6} \widetilde{m} (\kappa t^3) + 2Np + 2Nrt. \quad (6.1.26)$$

The covariant derivatives of W are

$$\begin{aligned} D_a W &= e_a + (\kappa m t)_a + \frac{1}{2} \tilde{m} (\kappa t^2)_a + 2 (N r)_a + \frac{3i (\kappa v^2)_a}{2 (\kappa v^3)} W, \\ D_K W &= 2p_K + 2(rt)_K - 4e^D \mathcal{F}_K W. \end{aligned} \quad (6.1.27)$$

We will also mention that there are tadpole conditions which the generalized fluxes should satisfy,

$$-\sqrt{2} (p_K \tilde{m} - r_{aK} m^a) = 2N_K^{(\text{O6})} - N_K^{(\text{D6})}. \quad (6.1.28)$$

where the right hand side represents the contribution of localized sources, both O6-planes and D6-branes.

6.1.3 Slow-roll parameters

As discussed in [197], we can express the slow-roll parameters ϵ and η in terms of the scalar potential and Kähler metric.

The expression for ϵ is

$$\begin{aligned} \epsilon &= V^{-2} \left\{ K^{a\bar{b}} \frac{\partial}{\partial t^a} V \frac{\partial}{\partial \bar{t}^{\bar{b}}} V + K^{I\bar{J}} \frac{\partial}{\partial N^I} V \frac{\partial}{\partial \bar{N}^{\bar{J}}} V \right\} \\ &= \frac{1}{4} V^{-2} \left\{ K^{a\bar{b}} \left(\frac{\partial}{\partial v^a} V \frac{\partial}{\partial v^{\bar{b}}} V + \frac{\partial}{\partial u^a} V \frac{\partial}{\partial u^{\bar{b}}} V \right) \right. \\ &\quad \left. + K^{I\bar{J}} \left(\frac{\partial}{\partial \text{Re } N^I} V \frac{\partial}{\partial \text{Re } N^{\bar{J}}} V + \frac{\partial}{\partial \text{Im } N^I} V \frac{\partial}{\partial \text{Im } N^{\bar{J}}} V \right) \right\}. \end{aligned} \quad (6.1.29)$$

If we further define $v^a = \rho \gamma^a$, where

$$\kappa_{abc} \gamma^a \gamma^b \gamma^c = 6, \quad (6.1.30)$$

so that the over-all volume is $\mathcal{V}_6 = \rho^3$, and use the expressions for the Kähler

metric above, then we can further simplify this to

$$\begin{aligned}
\epsilon = V^{-2} & \left\{ \frac{1}{3} \rho^2 \left(\frac{\partial V}{\partial \rho} \right)^2 + \frac{1}{4} \left(\frac{\partial V}{\partial D} \right)^2 \right. \\
& + \left[-(\kappa\gamma)^{-1ab} + \frac{1}{6} \gamma^a \gamma^b \right] \frac{\partial V}{\partial \gamma^a} \frac{\partial V}{\partial \gamma^b} + \frac{1}{4} \left[\widehat{K}^{JK} - \mathcal{Z}^J \mathcal{Z}^K \right] \frac{\partial V}{\partial \mathcal{Z}^J} \frac{\partial V}{\partial \mathcal{Z}^K} \\
& \left. + \rho^2 \left[-(\kappa\gamma)^{-1ab} + \frac{1}{2} \gamma^a \gamma^b \right] \frac{\partial V}{\partial u^a} \frac{\partial V}{\partial u^b} + e^{-2D} \widehat{K}^{JK} \frac{\partial V}{\partial \xi^J} \frac{\partial V}{\partial \xi^K} \right\}. \tag{6.1.31}
\end{aligned}$$

This expression splits ϵ into six non-negative pieces. The first line involves the overall volume modulus ρ and the four-dimensional dilaton D . In this class of models, the potential V can always be written as a polynomial in ρ and e^D , so this line is often very easy to compute. The no-go theorems of [198] have been derived by focussing only on this line.

The second line involves the angular Kähler moduli, γ^a and the complex structure moduli \mathcal{Z}^J . Both these sets of variables are constrained ($\kappa\gamma^3 = 6$, $\widehat{K}\mathcal{Z}^2 = 1$), so there is no unique way of writing the potential in terms of them. However, the metrics which appear in the expression above are such that ϵ doesn't depend on these choices. For example, if a $\kappa\gamma^3$ appears anywhere in V , then when the derivative with respect to γ^a hits it we get a contribution proportional to $\kappa_{abc}\gamma^b\gamma^c$, but this is annihilated by the term in square brackets above. Finally the third line contains the axions u^a and ξ^J .

The expression for η is slightly more complicated. First we must define a canonical metric g_{ij} on the moduli space of real fields, given by

$$\frac{1}{2} g_{ij} d\phi^i d\phi^j = K_{A\bar{B}} d\Phi^A d\bar{\Phi}^{\bar{B}}. \tag{6.1.32}$$

Here we are using the indices i and j to run over all real valued fields, while A and B run over all complex-valued fields, from both the complex structure and

Kähler sectors of the theory. From this metric we can then compute Christoffel symbols Γ_{jk}^i , and then we have

$$\eta = \text{minimum eigenvalue of } \left\{ \frac{g^{ik} (\partial_k \partial_j V - \Gamma_{kj}^\ell \partial_\ell V)}{V} \right\}. \quad (6.1.33)$$

6.1.4 Scalar potential with metric fluxes

The full expression for the scalar potential in this case is given by

$$\begin{aligned} V = & \frac{3}{(\kappa v^3)} e^{2D} \left\{ \widehat{K}^{IJ} (p_I + r_{aI} u^a) (p_J + r_{bJ} u^b) \right. \\ & \left. + \widehat{K}^{IJ} r_{aI} r_{bJ} v^a v^b - 2 (\mathcal{Z}^K r_{aK} v^a)^2 - \frac{2}{3} (\kappa v^3) (\kappa v)^{-1ab} \mathcal{Z}^I \mathcal{Z}^J r_{aI} r_{bJ} \right\} \\ & + 2e^{2D} (\widehat{\kappa} v)^{-1\alpha\beta} (\widehat{r}_\alpha^I \mathcal{F}_I) (\widehat{r}_\beta^J \mathcal{F}_J) + 2e^{3D} \mathcal{Z}^K (\widetilde{m} p_K - r_{aK} m^a) \\ & + \frac{3}{4(\kappa v^3)} e^{4D} \left\{ \left[-\frac{2}{3} (\kappa v^3) (\kappa v)^{-1ab} + 2v^a v^b \right] \right. \\ & \quad \times \left(\xi^I r_{aI} + e_a + (\kappa m u)_a + \frac{1}{2} \widetilde{m} (\kappa u^2)_a \right) \\ & \quad \times \left(\xi^J r_{bJ} + e_b + (\kappa m u)_b + \frac{1}{2} \widetilde{m} (\kappa u^2)_b \right) \\ & \quad + 4 \left[\xi^K (p_K + r_{aK} u^a) + \widetilde{e} + (eu) + \frac{1}{2} (\kappa m u^2) + \frac{1}{6} \widetilde{m} (\kappa u^3) \right]^2 \\ & \quad \left. + \left[-\frac{2}{3} (\kappa v^3) (\kappa v)_{ab} + (\kappa v^2)_a (\kappa v^2)_b \right] (m^a + \widetilde{m} u^a) (m^b + \widetilde{m} u^b) \right. \\ & \quad \left. + \frac{1}{9} \widetilde{m}^2 (\kappa v^3)^2 \right\}. \quad (6.1.34) \end{aligned}$$

6.2 No-Go Theorems

In this section we will prove a series of no-go theorems. Each one will show that, given some restrictions on the model, the slow-roll parameter is bounded below by some positive number of order unity, thus ruling out both

slow-roll inflation and the existence of de Sitter extrema. There will be two types of restrictions that we will consider. We might impose conditions on the intersection numbers of the model, as encoded by the polynomials κ_{abc} , $\widehat{\kappa}_{a\alpha\beta}$ and the polynomial p_n , and we will further restrict which fluxes can be turned on.

6.2.1 General manifolds

Consider first the case where no restrictions are assumed on the intersection numbers of the model, that is κ , $\widehat{\kappa}$ and p_n are unconstrained.

Here we will demonstrate two no-go theorems. The first was shown by [198] and pertains to the case of no metric fluxes, that is $r_{aK} = \widehat{r}_\alpha^K = 0$. In this case the scalar potential simplifies to

$$\begin{aligned}
V = & \frac{1}{2}\rho^{-3}e^{2D}\widehat{K}^{IJ}p_Ip_J + 2\widetilde{m}e^{3D}\mathcal{Z}^Kp_K \tag{6.2.1} \\
& + \frac{1}{8}\rho^{-3}e^{4D}\left\{4\rho^2\left[-(\kappa\gamma)^{-1ab} + \frac{1}{2}\gamma^a\gamma^b\right]\right. \\
& \quad \times \left(e_a + (\kappa mu)_a + \frac{1}{2}\widetilde{m}(\kappa u^2)_a\right)\left(e_b + (\kappa mu)_b + \frac{1}{2}\widetilde{m}(\kappa u^2)_b\right) \\
& \quad + \left[\xi^K p_K + \widetilde{e} + (eu) + \frac{1}{2}(\kappa mu^2) + \frac{1}{6}\widetilde{m}(\kappa u^3)\right]^2 \\
& \quad \left. + \rho^4[-4(\kappa\gamma)_{ab} + (\kappa\gamma^2)_a(\kappa\gamma^2)_b](m^a + \widetilde{m}u^a)(m^b + \widetilde{m}u^b) + 4\widetilde{m}^2\rho^6\right\}.
\end{aligned}$$

Note that the metrics $[-(\kappa\gamma)^{-1ab} + \frac{1}{2}\gamma^a\gamma^b]$ and $[-4(\kappa\gamma)_{ab} + (\kappa\gamma^2)_a(\kappa\gamma^2)_b]$ are both positive definite since they are equal to $\frac{1}{4}\rho^{-2}K^{a\bar{b}}$ and $16\rho^2K_{a\bar{b}}$ respectively, so that the only term in the potential which can be negative is the second term on the first line above.

From this we can easily check that

$$3\partial_D V - \rho\partial_\rho V \geq 9V. \tag{6.2.2}$$

Finally, if we also have $V > 0$, then we can write

$$\begin{aligned}\epsilon &\geq V^{-2} \left[\frac{1}{3} \rho^2 (\partial_\rho V)^2 + \frac{1}{4} (\partial_D V)^2 \right] \\ &= V^{-2} \left[\frac{1}{39} (3\partial_D V - \rho\partial_\rho V)^2 + \frac{1}{52} (\partial_D + 4\rho\partial_\rho V)^2 \right] \geq \frac{27}{13}. \quad (6.2.3)\end{aligned}$$

Let us consider another example of possible interest, where we allow metric fluxes, but don't allow a Romans mass parameter, that is we take $\tilde{m} = 0$. This would be the models one would look at if one wished to have a straightforward lift to M-theory, for example. In this case one can easily check that there is another no-go theorem. Indeed, we have

$$\partial_D V - \rho\partial_\rho V \geq 3V, \quad (6.2.4)$$

and so

$$\epsilon \geq V^{-2} \left[\frac{1}{7} (\partial_D V - \rho\partial_\rho V)^2 + \frac{1}{84} (3\partial_D V + 4\rho\partial_\rho V)^2 \right] \geq \frac{9}{7}. \quad (6.2.5)$$

Thus in both these cases there is no possibility of slow-roll inflation, and no possibility of finding a de Sitter extremum of the potential anywhere in field space (such a point would have $\epsilon = 0$, of course). For this reason we will assume that $\tilde{m} \neq 0$ in subsequent sections, and we will focus on cases in which some metric fluxes are non-zero. The observation that a non-zero \tilde{m} is necessary for the existence of de Sitter vacua was made independently in [202].

6.2.2 Factorization in the Kähler sector

Now consider a more restricted class of models, in which there is one distinguished Kähler modulus v^0 , such that the only non-zero intersections are

$$\kappa_{0ij} = X_{ij}, \quad \hat{\kappa}_{0\alpha\beta} = \hat{X}_{\alpha\beta}, \quad (6.2.6)$$

and their permutations, where i and j run over the remaining Kähler moduli. When considering general Calabi-Yau orientifolds, this is a very unnatural condition. However, frequently one is interested in orientifolds of T^6 and there is a hierarchy between the moduli of the untwisted sector and those of the twisted sectors. In such a case one typically truncates to the untwisted moduli, and in this sector the constraints above on the intersection numbers are not uncommon; they correspond to the presence of a T^2 factor in the T^6 which is preserved by the orientifold group.

In the general case, we found it profitable to split the Kähler moduli into an overall volume variable ρ and a set of angular variables γ^a . In the factorized case it is more useful to take v^0 and then split the remaining moduli by defining $v^i = \sigma\chi^i$, where the angular variables χ^i are constrained by

$$X_{ij}\chi^i\chi^j = 2. \quad (6.2.7)$$

Then, for instance, the volume of the space is $\mathcal{V}_6 = v^0\sigma^2$. With these conventions we find that the Kähler metric and its inverse have the form

$$K_{a\bar{b}} = \begin{pmatrix} \frac{1}{4(v^0)^2} & 0 \\ 0 & \frac{1}{4\sigma^2} [(X\chi)_i (X\chi)_j - X_{ij}] \end{pmatrix}, \quad (6.2.8)$$

$$K^{a\bar{b}} = \begin{pmatrix} 4(v^0)^2 & 0 \\ 0 & 4\sigma^2 [-X^{-1ij} + \chi^i\chi^j] \end{pmatrix}. \quad (6.2.9)$$

The combinations that appear in the expression for ϵ are

$$\begin{aligned} \frac{1}{4}K^{a\bar{b}}\frac{\partial V}{\partial v^a}\frac{\partial V}{\partial v^b} &= (v^0)^2 \left(\frac{\partial V}{\partial v^0}\right)^2 \\ &+ \frac{1}{2}\sigma^2 \left(\frac{\partial V}{\partial \sigma}\right)^2 + \left[-X^{-1ij} + \frac{1}{2}\chi^i\chi^j\right] \frac{\partial V}{\partial \chi^i}\frac{\partial V}{\partial \chi^j}. \end{aligned} \quad (6.2.10)$$

We can now look for no-go theorems involving the three variables D , v^0 , and σ . For example, suppose $r_{0K} = 0$, but we allow non-zero r_{iK} and \widehat{r}_α^K , then

$$\begin{aligned}
V = & \frac{1}{2v^0\sigma^2} e^{2D} \left\{ \widehat{K}^{IJ} (p_I + r_{iI}u^i) (p_J + r_{jJ}u^j) \right. \\
& \left. + \sigma^2 \widehat{K}^{IJ} r_{iI} r_{jJ} \chi^i \chi^j - 4\sigma^2 \mathcal{Z}^I \mathcal{Z}^J X^{-1ij} r_{iI} r_{jJ} \right\} \\
& + 2(v^0)^{-1} e^{2D} \widehat{X}^{-1\alpha\beta} (\widehat{r}_\alpha^I \mathcal{F}_I) (\widehat{r}_\beta^J \mathcal{F}_J) + 2e^{3D} \mathcal{Z}^K (\widetilde{m} p_K - r_{iK} m^i) \\
& + \frac{1}{2v^0\sigma^2} e^{4D} \left\{ (v^0)^2 \left(e_0 + X_{ij} m^i u^j + \frac{1}{2} \widetilde{m} X_{ij} u^i u^j \right)^2 \right. \\
& + \sigma^2 [-X^{-1ij} + \chi^i \chi^j] (\xi^I r_{iI} + e_i + m^0 X_{ik} u^k + u^0 X_{ik} m^k + \widetilde{m} u^0 X_{ik} u^k) \\
& \quad \times (\xi^J r_{jJ} + e_j + m^0 X_{jl} u^l + u^0 X_{jl} m^l + \widetilde{m} u^0 X_{jl} u^l) \\
& + \left[\xi^K (p_K + r_{iK} u^i) + \widetilde{e} + e_0 u^0 + e_i u^i \right. \\
& \quad \left. + \frac{1}{2} (m^0 + \widetilde{m} u^0) X_{ij} u^i u^j + u^0 X_{ij} m^i u^j \right]^2 \\
& + \sigma^4 (m^0 + \widetilde{m} u^0)^2 \\
& + (v^0)^2 \sigma^2 [-X_{ij} + X_{ik} X_{jl} \chi^k \chi^l] (m^i + \widetilde{m} u^i) (m^j + \widetilde{m} u^j) \\
& \left. + \widetilde{m}^2 (v^0)^2 \sigma^4 \right\}. \tag{6.2.11}
\end{aligned}$$

In this case it is easy to check that

$$\partial_D V - v^0 \partial_{v^0} V \geq 3V, \tag{6.2.12}$$

so that

$$\epsilon \geq V^{-2} \left[\frac{1}{5} (\partial_D V - v^0 \partial_{v^0} V)^2 + \frac{1}{20} (\partial_D V + 4v^0 \partial_{v^0} V)^2 \right] \geq \frac{9}{5}. \tag{6.2.13}$$

Similarly, if $r_{iK} = \widehat{r}_\alpha^K = 0$, but we have arbitrary r_{0K} , the potential has the

form

$$\begin{aligned}
V = & \frac{1}{2v^0\sigma^2}e^{2D} \left\{ \widehat{K}^{IJ} (p_I + r_{0I}u^0) (p_J + r_{0J}u^0) + (v^0)^2 \widehat{K}^{IJ} r_{0I}r_{0J} \right\} \\
& + 2e^{3D} \mathcal{Z}^K (\tilde{m}p_K - r_{0K}m^0) \\
& + \frac{1}{2v^0\sigma^2}e^{4D} \left\{ (v^0)^2 \left(\xi^K r_{0K} + e_0 + X_{ij}m^i u^j + \frac{1}{2}\tilde{m}X_{ij}u^i u^j \right)^2 \right. \\
& + \sigma^2 [-X^{-1ij} + \chi^i \chi^j] (e_i + m^0 X_{ik}u^k + u^0 X_{ik}m^k + \tilde{m}u^0 X_{ik}u^k) \\
& \quad \times (e_j + m^0 X_{jl}u^l + u^0 X_{jl}m^l + \tilde{m}u^0 X_{jl}u^l) \\
& + \left[\xi^K (p_K + r_{0K}u^0) + \tilde{e} + e_0 u^0 + e_i u^i \right. \\
& \quad \left. + \frac{1}{2} (m^0 + \tilde{m}u^0) X_{ij}u^i u^j + u^0 X_{ij}m^i u^j \right]^2 \\
& + \sigma^4 (m^0 + \tilde{m}u^0)^2 \\
& + (v^0)^2 \sigma^2 [-X_{ij} + X_{ik}X_{jl}\chi^k \chi^l] (m^i + \tilde{m}u^i) (m^j + \tilde{m}u^j) \\
& \left. + \tilde{m}^2 (v^0)^2 \sigma^4 \right\}, \tag{6.2.14}
\end{aligned}$$

and one can check that

$$2\partial_D V - \sigma\partial_\sigma V \geq 6V, \tag{6.2.15}$$

giving

$$\epsilon \geq V^{-2} \left[\frac{1}{18} (2\partial_D V - \sigma\partial_\sigma V)^2 + \frac{1}{36} (\partial_D V + 4\sigma\partial_\sigma V)^2 \right] \geq 2. \tag{6.2.16}$$

Thus in order to get slow-roll inflation, we must have non-zero metric fluxes both with a 0 index and without, where by fluxes without a 0 index we mean either r_{iK} or \hat{r}_α^K .

6.2.3 Factorization in the complex structure sector

It is possible to find a similar sort of factorization in the complex structure sector. Recall that the computations of the Kähler potential in this sector was determined by a polynomial p_n which is homogeneous of degree $n = h^{2,1} + 1$ in n variables. We defined the usual dilaton D by writing $I^K = \text{Im}(N^K) = e^{-D} \mathcal{Z}^K$, where the \mathcal{Z}^K were constrained by $p_n(\mathcal{Z}) = 1$, or alternatively, $p_n(I) = e^{-nD}$.

Now suppose that we can divide the I^K into two sets, $I_{(1)}^A$ and $I_{(2)}^P$ (we will use letters from different parts of the alphabet for the different sets), and that the polynomial p_n factorizes as

$$p_n(I) = p_{n_1}^{(1)}(I_{(1)}) \cdot p_{n_2}^{(2)}(I_{(2)}), \quad (6.2.17)$$

where n_1 and n_2 are the degrees of the polynomials and satisfy $n_1 + n_2 = n$.³ In this case we can define two dilatons, D_1 and D_2 by

$$e^{-n_1 D_1} = p_{n_1}^{(1)}(I_{(1)}), \quad e^{-n_2 D_2} = p_{n_2}^{(2)}(I_{(2)}), \quad (6.2.18)$$

and we can define two sets of \mathcal{Z} by $\mathcal{Z}_{(1)}^A = e^{D_1} I_{(1)}^A$ and $\mathcal{Z}_{(2)}^P = e^{D_2} I_{(2)}^P$. Each of these sets will be constrained, since $1 = p_{n_1}^{(1)}(\mathcal{Z}_{(1)}) = p_{n_2}^{(2)}(\mathcal{Z}_{(2)})$. The Kähler potential in this sector is now given by

$$K = -\frac{4}{n} \ln [p_n(I)] = 4 \left(\frac{n_1}{n} D_1 + \frac{n_2}{n} D_2 \right). \quad (6.2.19)$$

³Note that the degrees n_1 and n_2 do not have to correspond to the cardinality of the sets $I_{(1)}^A$ and $I_{(2)}^P$.

The Kähler metric will then be block diagonal, with non-zero entries

$$\begin{aligned} K_{A\bar{B}} &= e^{2D_1} \widehat{K}_{(1)AB} \\ &= e^{2D_1} \frac{1}{n} \left[\partial_{AP} p_{n_1}^{(1)}(\mathcal{Z}_{(1)}) \partial_{BP} p_{n_1}^{(1)}(\mathcal{Z}_{(1)}) - \partial_A \partial_B p_{n_1}^{(1)}(\mathcal{Z}_{(1)}) \right], \end{aligned} \quad (6.2.20)$$

$$\begin{aligned} K_{P\bar{Q}} &= e^{2D_2} \widehat{K}_{(2)PQ} \\ &= e^{2D_2} \frac{1}{n} \left[\partial_{PP} p_{n_2}^{(2)}(\mathcal{Z}_{(2)}) \partial_{QQ} p_{n_2}^{(2)}(\mathcal{Z}_{(2)}) - \partial_P \partial_Q p_{n_2}^{(2)}(\mathcal{Z}_{(2)}) \right]. \end{aligned} \quad (6.2.21)$$

Furthermore, in ϵ we will find the combination

$$\begin{aligned} \frac{1}{4} K^{J\bar{K}} \frac{\partial V}{\partial I^J} \frac{\partial V}{\partial I^{\bar{K}}} &= \frac{n}{4n_1} \left(\frac{\partial V}{\partial D_1} \right)^2 + \frac{n}{4n_2} \left(\frac{\partial V}{\partial D_2} \right)^2 \\ &\quad + \frac{1}{4} \left[\widehat{K}_{(1)}^{AB} - \frac{n}{n_1} \mathcal{Z}_{(1)}^A \mathcal{Z}_{(1)}^B \right] \frac{\partial V}{\partial \mathcal{Z}_{(1)}^A} \frac{\partial V}{\partial \mathcal{Z}_{(1)}^B} \\ &\quad + \frac{1}{4} \left[\widehat{K}_{(2)}^{PQ} - \frac{n}{n_2} \mathcal{Z}_{(2)}^P \mathcal{Z}_{(2)}^Q \right] \frac{\partial V}{\partial \mathcal{Z}_{(2)}^P} \frac{\partial V}{\partial \mathcal{Z}_{(2)}^Q}. \end{aligned} \quad (6.2.22)$$

We can now try to concoct more no-go theorems working with the variables D_1 , D_2 , and ρ . However, it turns out that we only gain an advantage over the general case if the non-zero flux contributions to the six-brane tadpoles come from only one of the subsets above, say only the subset labeled (1). In other words, we need to demand that $\tilde{m}_{pP} = r_{aP} m^a$ (so that the tadpole contributions with a P -index vanish), while we allow $\tilde{m}_{pA} - r_{aA} m^a \neq 0$. In this case, and under certain extra conditions on the fluxes, we can find no-go theorems. We present three examples, but the list is not exhaustive.

If $r_{aP} = \widehat{r}_\alpha^A = 0$ and $n_1 \geq n_2$, then we can show that $\partial_{D_2} V \geq \frac{4n_2}{n} V$, which gives $\epsilon \geq \frac{4n_2}{n}$. Note in this case that combining $r_{aP} = 0$ with our assumptions about the tadpoles forces $p_P = 0$.

If $r_{aA} = \widehat{r}_\alpha^P = 0$ then we have a family of inequalities of the form $3\partial_{D_1} V + x\partial_{D_2} V - \rho\partial_\rho V \geq \frac{9n_1 + (4x-3)n_2}{n} V$, where x is any real number satisfying

inequalities

$$x \leq 2, \quad x \leq \frac{5}{2} - \frac{n_1}{2n_2}, \quad x > \frac{3}{4} - \frac{9n_1}{4n_2}. \quad (6.2.23)$$

There are always solutions for x and ϵ turns out to always be maximized by taking x at the top of the allowed interval, which leads to

$$\begin{aligned} \epsilon &\geq \frac{49n_2}{n_1+28n_2}, & n_1 &\geq n_2, \\ \epsilon &\geq \frac{(9n_1+5n_2)^2}{n(39n_1+19n_2)}, & n_1 &\leq n_2. \end{aligned} \quad (6.2.24)$$

Similarly, if $r_{aK} = 0$ (i.e. both r_{aA} and r_{aP} vanish, and hence so also does p_P) then we can show that $3\partial_{D_1}V + x\partial_{D_2}V - \rho\partial_\rho V \geq \frac{9n_1+(4x-3)n_2}{n}V$, where now x is a real number satisfying

$$x > \frac{3}{4} - \frac{9n_1}{4n_2}, \quad x \leq \frac{5}{2} - \frac{n_1}{2n_2}, \quad (n_1 - n_2)x \geq 4n_1 - 2n_2. \quad (6.2.25)$$

In this case there are solutions only when $5n_2 > 9n_1$, in which case the strongest bound is

$$\epsilon \geq \frac{(9n_1 - 5n_2)^2}{39n_1^2 - 50n_1n_2 + 19n_2^2}. \quad (6.2.26)$$

6.2.4 Factorization in both sectors

Now, finally, let us briefly consider the case where there is factorization in both the Kähler and complex structure sectors of the theory, with the notation of the previous two sections. There are many possible no-go theorems which can be derived in various situations. One situation is relevant for our analysis below, so we present the derivation here.

This case occurs when we have $r_{0K} = 0$ (that is both $r_{0A} = r_{0P} = 0$) and $r_{iP} = p_P = \hat{r}^A = 0$, but allow non-zero r_{iA} , p_A , and \hat{r}^P . Here we find a

family of inequalities, $\partial_{D_1} V + x \partial_{D_2} V - v^0 \partial_{v^0} V \geq \frac{3n_1 + (4x-1)n_2}{n} V$, where the real number x must satisfy

$$x > \frac{1}{4} - \frac{3n_1}{4n_2}, \quad (n_1 - n_2)x \geq n_1 - n_2. \quad (6.2.27)$$

These always admit solutions for x , and the corresponding bound on ϵ is given by

$$\begin{aligned} \epsilon &\geq \frac{9n_1 + 5n_2}{5n_1 + n_2}, & n_1 &\geq n_2, \\ \epsilon &\geq \frac{9}{5}, & n_1 &< n_2. \end{aligned} \quad (6.2.28)$$

6.3 Toroidal Orientifolds

For a generic Calabi-Yau three-fold, it is not presently understood exactly how to consistently include metric fluxes or non-geometric fluxes. When the manifold is at a point in its moduli space that admits a description as an orbifold of T^6 , however, we can identify a subset of these generalized fluxes which can be turned on simply by twisting the torus construction. In this case we can derive the full set of consistency conditions which must be satisfied, and it is for this reason that toroidal orbifolds and orientifolds are the most well-studied compactifications with generalized fluxes.

What will follow in the next section is a (partial) classification of orientifolds of T^6 which preserve $\mathcal{N} = 1$ supersymmetry in four dimensions (see also a related classification in [218]). The goal is simply to generate a list of examples in which to look for slow-roll inflation or to test the utility of our no-go theorems.

6.3.1 Classification of orientifolds

There is a well-known classification of abelian orbifold groups which act on T^6 without shifts and which preserve $\mathcal{N} = 2$ supersymmetry [219], [220],

Table 6.1: Cyclic orbifold groups

Group \mathbb{Z}_N	Generator $\frac{1}{N}(n_1, n_2, n_3)$
\mathbb{Z}_3	$\frac{1}{3}(1, 1, 1)$
\mathbb{Z}_4	$\frac{1}{4}(1, 1, 2)$
\mathbb{Z}_{6-I}	$\frac{1}{6}(1, 1, 4)$
\mathbb{Z}_{6-II}	$\frac{1}{6}(1, 2, 3)$
\mathbb{Z}_7	$\frac{1}{7}(1, 2, 4)$
\mathbb{Z}_{8-I}	$\frac{1}{8}(1, 2, 5)$
\mathbb{Z}_{8-II}	$\frac{1}{8}(1, 3, 4)$
\mathbb{Z}_{12-I}	$\frac{1}{12}(1, 4, 7)$
\mathbb{Z}_{12-II}	$\frac{1}{12}(1, 5, 6)$

[221]. We will not be too concerned with the explicit action on the lattice, except in some specific cases in section 6.5. As we will see, the action on the lattice only enters the story for us once we attempt to derive the correct quantization conditions on the generalized fluxes, but there is a great deal of information which can be obtained without these details. With this in mind, then, we have nine different cyclic groups and eight more products of cyclic groups which can occur as lattice-preserving subgroups of $SU(3)$, and hence give rise to $\mathcal{N} = 2$ orbifolds of T^6 . Moreover, this list exhausts the possibilities for abelian orbifold groups, up to isomorphism. These groups are listed in tables 6.1 and 6.2. In each table, the generator of the orbifold action is written as $\frac{1}{N}(n_1, n_2, n_3)$, which is shorthand for

$$(z_1, z_2, z_3) \mapsto (e^{2\pi i n_1/N} z_1, e^{2\pi i n_2/N} z_2, e^{2\pi i n_3/N} z_3). \quad (6.3.1)$$

Let us now classify the supersymmetric orientifolds of these models. An orientifold will be a \mathbb{Z}_2 extension \widehat{G} of the orbifold group G ,

$$1 \longrightarrow G \longrightarrow \widehat{G} \longrightarrow \mathbb{Z}_2 \longrightarrow 1, \quad (6.3.2)$$

Table 6.2: $\mathbb{Z}_N \times \mathbb{Z}_M$ orbifold groups

Group $\mathbb{Z}_N \times \mathbb{Z}_M$	Generator 1 $\frac{1}{N} (n_1, n_2, n_3)$	Generator 2 $\frac{1}{M} (m_1, m_2, m_3)$
$\mathbb{Z}_2 \times \mathbb{Z}_2$	$\frac{1}{2} (1, 0, 1)$	$\frac{1}{2} (0, 1, 1)$
$\mathbb{Z}_2 \times \mathbb{Z}_4$	$\frac{1}{2} (1, 0, 1)$	$\frac{1}{4} (0, 1, 3)$
$\mathbb{Z}_2 \times \mathbb{Z}_6$	$\frac{1}{2} (1, 0, 1)$	$\frac{1}{6} (0, 1, 5)$
$\mathbb{Z}_2 \times \mathbb{Z}'_6$	$\frac{1}{2} (1, 0, 1)$	$\frac{1}{6} (1, 1, 4)$
$\mathbb{Z}_3 \times \mathbb{Z}_3$	$\frac{1}{3} (1, 0, 2)$	$\frac{1}{3} (0, 1, 2)$
$\mathbb{Z}_3 \times \mathbb{Z}_6$	$\frac{1}{3} (1, 0, 2)$	$\frac{1}{6} (0, 1, 5)$
$\mathbb{Z}_4 \times \mathbb{Z}_4$	$\frac{1}{4} (1, 0, 3)$	$\frac{1}{4} (0, 1, 3)$
$\mathbb{Z}_6 \times \mathbb{Z}_6$	$\frac{1}{6} (1, 0, 5)$	$\frac{1}{6} (0, 1, 5)$

where each element of \widehat{G} which is not in the image of G (or equivalently not in the kernel of the map to \mathbb{Z}_2) must be accompanied by orientation reversal Ω and $(-1)^F$. In order to preserve $\mathcal{N} = 1$ supersymmetry in type IIA, we also require not only that the elements of G act as linear holomorphic maps, $G \subset \text{SU}(3)$, but we also demand that the orientation-reversing elements of \widehat{G} act as linear *antiholomorphic* maps. Thus for our classification we would like to find, for each of the orbifold group actions in the list above, all \mathbb{Z}_2 extensions, where the extension acts antiholomorphically. To be more explicit, we want to find an antiholomorphic linear map σ such that for every element g in the orbifold group G , we have $g\sigma g\sigma \in G$. As a set then, $\widehat{G} = G \cup \sigma G$. In given holomorphic coordinates z_1, z_2, z_3 , we can write each element as a three-by-three complex matrix with entries $g^i_j, \sigma^{\bar{i}}_{\bar{j}}$, and then the condition above is

$$g^i_k \sigma^{\bar{k}}_{\bar{\ell}} \bar{g}^{\bar{\ell}}_{\bar{m}} \sigma^{\bar{m}}_j = (g')^i_j, \quad (6.3.3)$$

for some $g' \in G$.

In all of the examples we will consider, the elements of G are all diagonal with entries $g^i_j = \partial_j^i \exp[i\theta_i]$ (for instance see the generators in tables 6.1

and 6.2), so we have

$$\sum_{\bar{k}} e^{i(\theta_i - \theta_{\bar{k}})} \bar{\sigma}_{\bar{k}}^i \sigma_{\bar{j}}^{\bar{k}} = \partial_j^i e^{i\theta'_i}, \quad (6.3.4)$$

with no sum over i .

We will now consider different cases.

First consider the cases of cyclic groups whose generators have $\theta_1, \theta_2, \theta_3$ all distinct (the last six cases in table 6.1). In this case we can show that (6.3.4) requires $\bar{\sigma}_{\bar{k}}^i \sigma_{\bar{j}}^{\bar{k}} = 0$ for all \bar{k} (no sum on \bar{k} here) and all $i \neq j$. These equations can be shown to imply that only three entries of σ are non-zero; either σ is diagonal, or it is block diagonal with a one-by-one block and a two-by-two block with zeros on the diagonal. In the diagonal case, we can apply phase changes to our holomorphic coordinates in order to transform σ into the three-by-three identity matrix, so that it acts by simple conjugation, $z_i \mapsto \bar{z}_i$. It turns out that this choice for σ will be valid for each of our orbifold groups, and so we will denote it as the *standard orientifold* for each case. These are summarized in table 6.3. In the non-diagonal cases, we can again use a phase rotation to set the one-by-one block to one, and we can set one of the non-zero entries of the two-by-two block to one. Then demanding that (6.3.4) be satisfied for each element of the orbifold group restricts the possibilities. We find that there are no allowed non-standard orientifolds for \mathbb{Z}_{6-II} and \mathbb{Z}_7 , one choice for each of the \mathbb{Z}_8 groups, and two choices each for the \mathbb{Z}_{12} groups, where to correctly count the number of independent choices, we should also recall that we can relabel our element σ as $\sigma' = g\sigma$, for any $g \in G$, and then drop the prime.

Next let us consider \mathbb{Z}_4 and \mathbb{Z}_{6-I} . In this case we can easily show that σ must be block diagonal with a two-by-two block for z_1 and z_2 , and a one-by-

Table 6.3: Standard $\mathcal{N} = 1$ Orientifolds ($z_i \mapsto \bar{z}_i$)

Group	$h_{-untw}^{1,1}$	$h_{+untw}^{1,1}$	$h_{untw}^{2,1}$
\mathbb{Z}_3	6	3	0
\mathbb{Z}_4	4	1	1
\mathbb{Z}_{6-I}	4	1	0
\mathbb{Z}_{6-II}	3	0	1
\mathbb{Z}_7	3	0	0
\mathbb{Z}_{8-I}	3	0	0
\mathbb{Z}_{8-II}	3	0	1
\mathbb{Z}_{12-I}	3	0	0
\mathbb{Z}_{12-II}	3	0	1
$\mathbb{Z}_2 \times \mathbb{Z}_2$	3	0	3
$\mathbb{Z}_2 \times \mathbb{Z}_4$	3	0	1
$\mathbb{Z}_2 \times \mathbb{Z}_6$	3	0	1
$\mathbb{Z}_2 \times \mathbb{Z}'_6$	3	0	0
$\mathbb{Z}_3 \times \mathbb{Z}_3$	3	0	0
$\mathbb{Z}_3 \times \mathbb{Z}_6$	3	0	0
$\mathbb{Z}_4 \times \mathbb{Z}_4$	3	0	0
$\mathbb{Z}_6 \times \mathbb{Z}_6$	3	0	0

one block for z_3 . A phase rotation can be used to set the latter entry to one (i.e. $\sigma \cdot z_3 = \bar{z}_3$), but we have quite a bit more symmetries at our disposal in the two-by-two block, since any $\text{GL}(2, \mathbb{C})$ matrix commutes with the orbifold group. It turns out that solving the constraints and then using the symmetries allows us to put σ into one of two canonical forms. Either we can set the two-by-two block to the identity, giving the standard orientifold, or we can set it to be the canonical antisymmetric matrix, so that $\sigma \cdot (z_1, z_2) = (\bar{z}_2, -\bar{z}_1)$.

In the case of \mathbb{Z}_3 , we have $\theta_1 = \theta_2 = \theta_3 = 2\pi i/3$. Here (6.3.4) is the least restrictive, but we also have the most symmetry, since the full $\text{GL}(3, \mathbb{C})$ commutes with the orbifold group. Here we can use this symmetry to always convert to the standard case.

We move on now to the product groups of table 6.2. As in the case of the cyclic groups with distinct angles, we can show that σ must be either diagonal, leading to the standard case, or block diagonal, with the two-by-two block having vanishing diagonal entries. Finally, by carefully examining the remaining constraints and symmetries we are able to find the independent non-standard orientifolds in each case. All of the non-standard orientifolds are summarized in table 6.4.

6.3.2 Turning on NSNS fluxes

Now for each of the orientifolds discussed in the last subsection, we would like to understand which generalized fluxes can be turned on consistently. The discussion will be very brief, and we will refer the interested reader to [209], [210] for a more careful discussion of our approach.

As explained in section 6.1.1, the H -flux and metric fluxes we would like to turn on can be thought of in terms of their components H_{ijk} and f_{jk}^i , where

Table 6.4: Non-Standard $\mathcal{N} = 1$ Orientifolds

Group	$\sigma \cdot (z_1, z_2, z_3)$	$h_{-\text{untw}}^{1,1}$	$h_{+\text{untw}}^{1,1}$	$h_{\text{untw}}^{2,1}$
\mathbb{Z}_4	$(\bar{z}_2, -\bar{z}_1, \bar{z}_3)$	2	3	1
\mathbb{Z}_{6-I}	$(\bar{z}_2, -\bar{z}_1, \bar{z}_3)$	2	3	0
\mathbb{Z}_{8-I}	$(\bar{z}_3, \bar{z}_2, \bar{z}_1)$	2	1	0
\mathbb{Z}_{8-II}	$(\bar{z}_2, \bar{z}_1, \bar{z}_3)$	2	1	1
\mathbb{Z}_{12-I}	$(\bar{z}_3, \bar{z}_2, \bar{z}_1)$	2	1	0
	$(\bar{z}_3, \bar{z}_2, i\bar{z}_1)$	2	1	0
\mathbb{Z}_{12-II}	$(\bar{z}_2, \bar{z}_1, \bar{z}_3)$	2	1	1
	$(\bar{z}_2, e^{\pi i/3} \bar{z}_1, \bar{z}_3)$	2	1	1
$\mathbb{Z}_2 \times \mathbb{Z}_2$	$(\bar{z}_1, \bar{z}_3, \bar{z}_2)$	2	1	3
$\mathbb{Z}_2 \times \mathbb{Z}_4$	$(\bar{z}_1, \bar{z}_3, \bar{z}_2)$	2	1	1
	$(\bar{z}_1, \bar{z}_3, i\bar{z}_2)$	2	1	1
$\mathbb{Z}_2 \times \mathbb{Z}_6$	$(\bar{z}_1, \bar{z}_3, \bar{z}_2)$	2	1	1
$\mathbb{Z}_2 \times \mathbb{Z}'_6$	$(\bar{z}_1, \bar{z}_3, \bar{z}_2)$	2	1	0
	$(\bar{z}_2, \bar{z}_1, \bar{z}_3)$	2	1	0
$\mathbb{Z}_3 \times \mathbb{Z}_3$	$(\bar{z}_1, \bar{z}_3, \bar{z}_2)$	2	1	0
$\mathbb{Z}_3 \times \mathbb{Z}_6$	$(\bar{z}_1, \bar{z}_3, \bar{z}_2)$	2	1	0
	$(\bar{z}_1, \bar{z}_3, -\bar{z}_2)$	2	1	0
$\mathbb{Z}_4 \times \mathbb{Z}_4$	$(\bar{z}_1, \bar{z}_3, \bar{z}_2)$	2	1	0
$\mathbb{Z}_6 \times \mathbb{Z}_6$	$(\bar{z}_1, \bar{z}_3, \bar{z}_2)$	2	1	0

the indices run over the six legs of the torus. For each toroidal orientifold, we need these objects to transform correctly under the quotient group. This means that both H_{ijk} and f_{jk}^i must be invariant under the orbifold group, and H_{ijk} should be odd under the action of σ , while f_{jk}^i should be even.

Next, we must also impose the Bianchi identities, which in this case take the form

$$H_{i[jk}f_{\ell m]}^i = 0, \quad f_{j[k}^i f_{\ell m]}^j = 0. \quad (6.3.5)$$

In terms of the geometry of the underlying twisted torus, the latter equation is simply that the exterior derivative is nilpotent ($d^2 = 0$), and the former condition is simply that H is a closed three-form ($dH = 0$). In section 6.4 we will tabulate the fluxes that can be turned on and all solutions to the Bianchi identities for each of our models. In principle one could violate the Bianchi identities by including localized NSNS sources [222], but we will not include such objects in this work.

Once we have found the set of independent fluxes which survive the orientifold quotient, it turns out to be more convenient, for the purposes of the low-energy theory, to refer not to the components H_{ijk} and f_{jk}^i , but instead to certain coefficients p_K , r_{aK} , and \hat{r}_α^K , given by

$$H = p_K b^K, \quad d\omega_a = -r_{aK} b^K, \quad d\mu_\alpha = -\hat{r}_\alpha^K a_K, \quad (6.3.6)$$

where a_K , b^K , ω_a , and μ_α are forms which descend from the untwisted cohomology of the torus *without* fluxes. There is an invertible linear map between the p_K and the independent components H_{ijk} . Similarly, the coefficients r_{aK} and \hat{r}_α^K are always given by linear combinations of the independent f_{jk}^i , but in this case the map is not always invertible; there can be more independent

f_{jk}^i than r_{aK} and \hat{r}_α^K . In these cases, the extra metric fluxes do appear in the Bianchi identities, and should be taken into account when classifying the solutions, but the scalar potential and the tadpole constraints (see below) depend only on r_{aK} and \hat{r}_α^K .

The Ramond-Ramond tadpoles are given by⁴

$$-\sqrt{2}(p_K \tilde{m} - r_{aK} m^a) = 2N_K^{(\text{O6})} - N_K^{(\text{D6})}, \quad (6.3.7)$$

where the right hand side of the equation above represents the contribution from localized sources, both O6-planes, which sit at the fixed points of the orientation reversing elements of the orientifold group, and the D6-branes which we allow to be added anywhere. We will not really be viewing these tadpoles as constraints in the present work, taking the attitude that D6-branes can be added as needed. Clearly, for a detailed analysis of any given model, one would have to proceed more carefully, taking into account RR quantization as well as the action of the orientifold on the open string sector.

From the perspective of the low-energy effective theory, these seem to be the only constraints that need to be obeyed. Indeed, we will find that for most models, these constraints are already enough for us to be able to apply our no-go theorems and rule out slow-roll inflation and de Sitter extrema from the corresponding scalar potential. However, we would like to understand exactly which models can be constructed consistently from a ten-dimensional perspective. There are at least two approaches to this problem, via the coset

⁴This expression is actually not exactly correct. Rather, this is a cohomological condition (in the sense of the cohomology of the torus without metric fluxes). There is an exact tadpole constraint of the schematic form $DF = J$, where D is the generalized derivative ($d + H$ on the twisted torus), F is the formal sum of the RR fluxes, and J is a delta function form describing the configuration of O6-planes and D6-branes.

constructions of [223], [224], [225], and the base-fiber twisted torus constructions of [209], which are inspired by [211], [212] and others. In the present work we will focus mainly on the latter approach when we want explicit constructions, so we will now briefly review it.

For a given orientifold of T^6 , we first pick a division of the torus into a base and a fiber, in such a way so that the orientifold group does not mix the two. More precisely, we require the tangent spaces of the base and fiber to be invariant subspaces of the orientifold action on the tangent space of the T^6 . Once this splitting has been chosen, then for each direction in the base, labeled by an index $i = 1, \dots, n$, we will choose a matrix $M_i \in \mathfrak{so}(6-n, 6-n)$. The entries of these matrices will correspond to the components of our fluxes,

$$(M_i) = \begin{pmatrix} -f_{ia}^b & H_{iab} \\ -Q_i^{ab} & f_{ib}^a \end{pmatrix}. \quad (6.3.8)$$

Here a and b are indices in the fiber directions. Q_i^{ab} are non-geometric fluxes which we don't want to turn on in this work, so the matrices in our examples will be upper-block-diagonal. Note that using these constructions we can only turn on sets of fluxes that have one lower index lying along the base, and the other two indices (upper or lower) lying along the fiber.

Using this language, it is straightforward to restrict to matrices that are compatible with the orientifold group. The Bianchi identities are reproduced simply by demanding that M_i and M_j commute for all i and j . Finally, quantization conditions for the generalized fluxes become simply the condition

$$\exp \left[\vec{\lambda} \cdot \vec{M} \right] \in \text{SO}(6, 6; \mathbb{Z}), \quad \vec{\lambda} \in \Lambda \cong \mathbb{Z}^6 \quad (6.3.9)$$

where $\Lambda \cong \mathbb{Z}^6$ is the lattice of torus identifications embedded in the tangent space of T^6 , and where we identify the M_i with a Lie algebra valued cotangent

vector. One of the great values of the base-fiber construction is that it enables one to identify the correct quantization conditions on the generalized fluxes, something that is not apparent from considerations of the low energy theory alone. And, in fact, there can be cases where the quantization conditions forbid us from turning on certain components of metric fluxes, for instance. These constructions are slightly generalized and formulated more precisely in [226].

6.3.3 Residual symmetries

Each of the no-go theorems we derived requires certain assumptions about which fluxes can be non-vanishing. Sometimes a solution of the Bianchi identities will automatically satisfy the assumptions for one of our theorems, but there will also be cases which do not appear to fall into one of these cases, but which we found numerically to still satisfy a bound on ϵ . In almost all of these cases we were able to find a symmetry (that is a field redefinition that preserved the form of the potential while simply changing which fluxes were turned on) that mapped us into a configuration for which the no-go theorems apply. For this reason, it is important to identify the group of symmetries which can act in this way.

This turns out to be fairly straightforward. Recall that type IIA on T^6 has a group of T-duality symmetries $SO(6, 6; \mathbb{Z})$, or, somewhat more precisely, we should use the double cover, $Spin(6, 6; \mathbb{Z})$. This group includes large diffeomorphisms of the torus (living in a $GL(6; \mathbb{Z})$ subgroup), shifts of the B -field, and also non-geometric symmetries such as performing a T-duality on a $T^2 \subset T^6$. In fact, the orientifold group \widehat{G} by which we are to quotient can also be considered as a subgroup of this T-duality group. The orbifold group

G sits inside $SU(3) \subset GL(6; \mathbb{Z})$, while the orientation-reversing elements of \widehat{G} sit inside of a \mathbb{Z}_2 extension of $GL(6; \mathbb{Z})$ in $Spin(6, 6; \mathbb{Z})$ [209]. The resulting space will still have a group of duality symmetries given by the elements h of the full T-duality group which satisfy

$$h\widehat{G}h^{-1} = \widehat{G}. \tag{6.3.10}$$

All of the fields and fluxes (here we should include the full set of non-geometric fluxes), as well as the Bianchi identities and tadpole constraints, will transform as representations of these residual symmetries. We will find situations where we can use these symmetries to map one set of fluxes which solves the Bianchi identities, to a new set of fluxes which still solves the Bianchi identities, but also satisfy the assumptions of one of our no-go theorems. The resulting bound on ϵ will apply to both configurations of fluxes (since we have simply performed a field redefinition).

6.4 Application of the No-Go Theorems to Toroidal Orientifolds

In this section we will apply our no-go theorems to the toroidal orientifold models discussed above. Since we are restricting to the untwisted sector we have eleven different models that are uniquely determined by their triple intersection numbers (A.2.2) and the prepotential in the complex structure sector (6.1.13). These models are summarized in table 6.5. Note that for all but the \mathbb{Z}_3 quotient we have a factorization in the Kähler sector and for all models with $h^{2,1} > 0$ we have a factorization in the complex structure sector.

For the benefit of the reader primarily interested in the results rather than how they arise, we summarize the weakest bounds on the slow-roll pa-

#	$\kappa, \hat{\kappa}$	$p_n(\mathcal{Z}) (= 1)$	Group
I	$\kappa_{123} = 1$	$2\mathcal{Z}^1$	$\mathbb{Z}_7, \mathbb{Z}_{8-I}, \mathbb{Z}_{12-I},$ $\mathbb{Z}_2 \times \mathbb{Z}_{6'}, \mathbb{Z}_3 \times \mathbb{Z}_3,$ $\mathbb{Z}_3 \times \mathbb{Z}_6, \mathbb{Z}_4 \times \mathbb{Z}_4,$ $\mathbb{Z}_6 \times \mathbb{Z}_6$
II	$\kappa_{122} = 1, \hat{\kappa}_{111} = -1$	$2\mathcal{Z}^1$	$\mathbb{Z}_{8-I}, \mathbb{Z}_{12-I},$ $\mathbb{Z}_2 \times \mathbb{Z}_{6'}, \mathbb{Z}_3 \times \mathbb{Z}_3,$ $\mathbb{Z}_3 \times \mathbb{Z}_6, \mathbb{Z}_4 \times \mathbb{Z}_4,$ $\mathbb{Z}_6 \times \mathbb{Z}_6$
III	$\kappa_{123} = 1, \kappa_{144} = -1,$ $\hat{\kappa}_{111} = -1$	$2\mathcal{Z}^1$	\mathbb{Z}_{6-I}
IV	$\kappa_{122} = 1, \hat{\kappa}_{1\alpha\alpha} = -1,$ $\alpha \in \{1, 2, 3\}$	$2\mathcal{Z}^1$	\mathbb{Z}_{6-I}
V	$\kappa_{123} = 1, \kappa_{456} = -2,$ $\kappa_{144} = \kappa_{255} = \kappa_{366} = -2,$ $\hat{\kappa}_{111} = \hat{\kappa}_{222} = \hat{\kappa}_{333} = -1,$ $\hat{\kappa}_{423} = \hat{\kappa}_{513} = \hat{\kappa}_{612} = 1,$	$2\mathcal{Z}^1$	\mathbb{Z}_3
VI	$\kappa_{123} = 1$	$2^2 \mathcal{Z}^1 \mathcal{Z}^2$	$\mathbb{Z}_{6-II}, \mathbb{Z}_{8-II}, \mathbb{Z}_{12-II},$ $\mathbb{Z}_2 \times \mathbb{Z}_4, \mathbb{Z}_2 \times \mathbb{Z}_6$
VII	$\kappa_{122} = 1, \hat{\kappa}_{111} = -1$	$2^2 \mathcal{Z}^1 \mathcal{Z}^2$	$\mathbb{Z}_{8-II}, \mathbb{Z}_{12-II},$ $\mathbb{Z}_2 \times \mathbb{Z}_4, \mathbb{Z}_2 \times \mathbb{Z}_6$
VIII	$\kappa_{123} = 1, \kappa_{144} = -1,$ $\hat{\kappa}_{111} = -1$	$2^2 \mathcal{Z}^1 \mathcal{Z}^2$	\mathbb{Z}_4
IX	$\kappa_{122} = 1, \hat{\kappa}_{1\alpha\alpha} = -1,$ $\alpha \in \{1, 2, 3\}$	$2^2 \mathcal{Z}^1 \mathcal{Z}^2$	\mathbb{Z}_4
X	$\kappa_{123} = 1$	$2^4 \mathcal{Z}^1 \mathcal{Z}^2 \mathcal{Z}^3 \mathcal{Z}^4$	$\mathbb{Z}_2 \times \mathbb{Z}_2$
XI	$\kappa_{122} = 1, \hat{\kappa}_{111} = -1$	$2^4 (\mathcal{Z}^1)^2 (\mathcal{Z}^2 \mathcal{Z}^3 - (\mathcal{Z}^4)^2)$	$\mathbb{Z}_2 \times \mathbb{Z}_2$

Table 6.5: This table summarizes the models we are considering. It contains the number of invariant forms and the non-vanishing triple intersection numbers (A.2.2) together with the prepotential for the complex structure sector (6.1.13).

parameter ϵ for the various cases in table 6.6. There are several special cases that can be shown to satisfy stronger bounds. These are omitted from table

Case	I,II,III,IV,V	VII	I',II', VI, VIII, IX	X, XI
$\epsilon \geq$	$\frac{27}{13}$	2	$\frac{9}{5}$	0

Table 6.6: This table presents a summary of the weakest bounds on the slow-roll parameter ϵ for the various cases. The Z_{8-I} -quotient turns out to be special for both case I and II, and we denote it by I' and II'.

6.6, but they are discussed in some detail below.

We will now discuss the solutions to the Bianchi identities for all of these cases and check which of our no-go theorems can be applied. We have also minimized ϵ numerically by allowing the moduli and fluxes to vary. We found generically that the bound given by the no-go theorem can be attained which proves that it is impossible to derive a stronger no-go theorem. For the two models X and XI we will find solutions to the Bianchi identities that escape all of our no-go theorems. In these cases it is possible to get vanishing ϵ and we will discuss this in detail in the next section.

6.4.1 Case I

There are two special quotients Z_7 and Z_{8-I} in this first case. Both of these have extra metric fluxes that are not contained in the matrix r_{aK} . The generic solution to the Bianchi identities for all models in case I has only one of the three entries in the r vector non-zero. (For Z_7 and Z_{8-I} the extra metric fluxes are zero.) For these solutions there is an $SL(2, \mathbb{R})$ subgroup of the residual T-duality symmetries (see section 6.3.3) under which p_1 and the nonvanishing r -flux transform as a doublet. This symmetry can be used to set the r -flux to zero (note that if other components of r were nonzero, then these T-dualities would mix them with nongeometric fluxes). A more pedestrian way to see this is that a shift in one of the B -axions allows us to set $p_1 = 0$.

A T-duality then takes us to a configuration that has no metric flux, and we find the bound $\epsilon \geq \frac{27}{13}$.

For \mathbb{Z}_{8-I} there is one more solution to the Bianchi identities due to the extra metric fluxes, which we call f_1 and f_2 . It reads $r_{21} = 0$, $r_{11}r_{31} = -f_1^2 = -f_2^2$. The factorization in the Kähler sector allows us to apply our no-go theorem since $r_{21} = "r_{01}" = 0$, and we obtain $\epsilon \geq \frac{9}{5}$.

6.4.2 Case II

As before, \mathbb{Z}_{8-I} is special because it has two extra metric fluxes f_1, f_2 that are not contained in the r matrix.

The solution to the Bianchi identities common to all quotients forces $r_{21} = \hat{r}_1^1 = 0$, ($f_1 = f_2 = 0$) so that we are again left with only one single metric flux that can be mapped to H -flux just as discussed in case I, and we again find $\epsilon \geq \frac{27}{13}$.

For \mathbb{Z}_{8-I} there is one more solution to the Bianchi identities $r_{11} = r_{21} = p_1 = 0$, $\hat{r}_1^1 = f_1^2 + f_2^2$.

From the factorization in the Kähler sector and using $r_{11} = "r_{0K}" = 0$ we find $\epsilon \geq \frac{9}{5}$.

6.4.3 Case III

This case has two solutions to the Bianchi identities $r_{21} = r_{31} = r_{41} = \hat{r}_1^1 = 0$ and $r_{11} = 2r_{21}r_{31} - r_{41}^2 = \hat{r}_1^1 = 0$. The second case can be brought into a form where only one of the r_{21}, r_{31}, r_{41} is non-zero using field redefinitions as described in subsection 6.3.3. To see this, note that the potential has an $SO(2, 1)$ symmetry under which r_{21}, r_{31} , and r_{41} transform as the components of a null vector. This means we can boost to a configuration that has only

r_{21} or r_{31} non-zero. After shifting one of the B -axions to absorb p_1 , both configurations can be T-dualized to a case with only H -flux resulting in $\epsilon \geq \frac{27}{13}$.

6.4.4 Case IV

This case again has only one non-vanishing entry in the r matrix since the Bianchi identities are $r_{21} = \hat{r}_\alpha^1 = 0, \forall \alpha \in \{1, 2, 3\}$. Using a shift in one of the axions together with a T-duality transformation, we see that the configuration is equivalent to one with H -flux and no metric flux. Therefore, we find $\epsilon \geq \frac{27}{13}$.

6.4.5 Case V

The Bianchi identities are $\hat{r}_\alpha^1 = 0, 2r_{11}r_{41} + r_{51}r_{61} = 2r_{21}r_{51} + r_{41}r_{61} = 2r_{31}r_{61} + r_{41}r_{51} = 0, 4r_{11}r_{21} - r_{61}^2 = 4r_{11}r_{31} - r_{51}^2 = 4r_{21}r_{31} - r_{41}^2 = 0$. These have two classes of solutions.

The first class is characterized by the vanishing of two of the fluxes $r_{41}, r_{51},$ and r_{61} . With the others being related to this by symmetries, let us consider $r_{51} = r_{61} = 0$, for definiteness. The remaining Bianchi identities then are $r_{11}r_{21} = 0, r_{11}r_{31} = 0,$ and $4r_{21}r_{31} - r_{41}^2 = 0$. One obvious solution is $r_{21} = r_{31} = r_{41} = 0$, but this is equivalent by symmetries to a special case of the solution $r_{11} = 0,$ and $4r_{21}r_{31} - r_{41}^2 = 0$, so let us focus on the latter solution. The potential has a manifest $SO(2, 1)$ symmetry under which r_{21}, r_{31}, r_{41} transform as the components of a vector. The Bianchi identity enforces this vector to be null. We can perform a boost such that only r_{21} or r_{31} is non-zero. At least locally, we can also redefine one of the B -axions to set $p_1 = 0$. It is then easy to see that the remaining configuration is T-dual to a configuration with H-flux but no metric fluxes, which implies $\epsilon \geq \frac{27}{13}$.

A naively inequivalent class of solutions has r_{41} , r_{51} , and r_{61} non-zero. The Bianchi identities then determine r_{11} , r_{21} , r_{31} in terms of these as $r_{11} = -r_{51}r_{61}/2r_{41}$, $r_{21} = -r_{41}r_{61}/2r_{51}$, $r_{31} = -r_{51}r_{61}/2r_{41}$. This class turns out to be related by symmetries to the first class, and consequently also obeys $\epsilon \geq \frac{27}{13}$. To see this, we can pick our favorite among r_{11} , r_{21} , and r_{31} , which, without loss of generality, we will take to be r_{21} . Let us make use of another $SO(2, 1)$ symmetry of the potential under which r_{21} transforms as a scalar, r_{11} , r_{31} , and r_{51} as the components of a vector, and r_{41} and r_{61} as the components of a spinor. We can use the $SO(2)$ subgroup of $SO(2, 1)$ to set one component of the spinor to zero, say, r_{61} . The Bianchi identities together with the invariance of r_{21} guarantee that as we take r_{61} to zero, because r_{41} remains finite, r_{51} must vanish such that $r_{41}r_{61}/r_{51}$ remains constant. With r_{51} and r_{61} zero, we are back to the first class of solutions, and a boost of our earlier $SO(2, 1)$ group followed by a field redefinition of one of the axions and a T-duality again takes us to a configuration without metric flux implying $\epsilon \geq \frac{27}{13}$.

6.4.6 Case VI

The \mathbb{Z}_{6-II} quotient is special because it allows for one extra metric flux we denote f_1 that is not contained in the r matrix.

We can exclude all solutions by using our two no-go theorems that rely on the factorization in the Kähler sector. There are two solutions up to permutation of the $a \in \{1, 2, 3\}$ index, that are common to all models (and for which the extra flux f_1 vanishes). The first has $r_{2K} = r_{3K} = 0, \forall K \in \{1, 2\}$ and our no-go theorem gives $\epsilon \geq 2$. The other solution reads $r_{1K} = r_{21}r_{32} + r_{22}r_{31} = 0, \forall K \in \{1, 2\}$ and we find $\epsilon \geq \frac{9}{5}$.

Finally, for \mathbb{Z}_{6-II} we have one more solution for which the extra metric flux is

non-vanishing: $r_{1K} = 0, \forall K \in \{1, 2\}, r_{2L} = r_{3M} = 0, r_{2M}r_{3L} = f_1^2, L \neq M, L, M \in \{1, 2\}$. We again find $\epsilon \geq \frac{9}{5}$.

6.4.7 Case VII

Here we have four different solutions to the Bianchi identities that can be dealt with using four different no-go theorems.

The first solution is $\hat{r}^K = r_{1K} = 0, \forall K \in \{1, 2\}$ and r_{21} or $r_{22} = 0$. We are left with one single entry in the r matrix. A field redefinition relates this to a configuration with only H -flux and we again have $\epsilon \geq \frac{27}{13}$.

Another solution is $\hat{r}^K = r_{2K} = 0, \forall K \in \{1, 2\}$ and our no-go theorem for factorization in the Kähler sector gives $\epsilon \geq 2$.

The third solution is $r_{2K} = 0, \forall K \in \{1, 2\}, \hat{r}^L = p_M = r_{1M} = 0, L \neq M, L, M \in \{1, 2\}$. Here we can use the factorization in $p_n(\mathcal{Z})$ and apply one of our no-go theorem for $n_1 = n_2 = 1$ and find $\epsilon \geq \frac{4n_2}{n_1+n_2} = 2$.

Using the factorization in both complex and Kähler sector we can show that the last solution $r_{1K} = 0, \forall K \in \{1, 2\}, \hat{r}^L = p_M = r_{2M} = 0, L \neq M, L, M \in \{1, 2\}$ has $\epsilon \geq \frac{9n_1+5n_2}{5n_1+n_2} = \frac{7}{3}$, where we used $n_1 = n_2 = 1$.

6.4.8 Case VIII

This case has six different solutions to the Bianchi identities. Three of those are

- $r_{1K} = \hat{r}_1^K = 0, \forall K \in \{1, 2\}, r_{21}r_{32} + r_{22}r_{31} - r_{41}r_{42} = 0$
- $r_{1K} = r_{4K} = p_K \hat{r}_1^K = r_{2K} \hat{r}_1^K = r_{3K} \hat{r}_1^K = 0, \forall K \in \{1, 2\}, 2r_{2L}r_{3L} + (\hat{r}_1^M)^2 = \epsilon^{LM} p_L r_{2M} = 0, 2r_{2L}r_{3M} - \hat{r}_1^1 \hat{r}_1^2 = 0, \epsilon^{LM} p_L r_{3M} = 0, L \neq M, L, M \in \{1, 2\}$

- $r_{1K} = p_K \hat{r}_1^K = r_{aK} \hat{r}_1^K = 0, \forall K \in \{1, 2\}, 2r_{2L}r_{3L} - (r_{4L})^2 + (\hat{r}_1^M)^2 = 0, \epsilon^{LM} p_L r_{aM} = 0, 2r_{2L}r_{3M} - r_{41}r_{42} - \hat{r}_1^1 \hat{r}_1^2 = 0, \epsilon^{LM} r_{4L} r_{aM} = 0, L \neq M, L, M \in \{1, 2\}$

These all have $r_{1K} = 0$ and we can use the factorization in the Kähler sector to show that $\epsilon \geq \frac{9}{5}$.

The next solution $\hat{r}_1^K = r_{2K} = r_{3K} = r_{4K} = 0, \forall K \in \{1, 2\}$ has $\epsilon \geq 2$ again due to the factorization in the Kähler sector.

The fifth solution $r_{2K} = r_{3K} = r_{4K} = 0, \forall K \in \{1, 2\}, \hat{r}_1^L = p_M = r_{1M} = 0, L \neq M, L, M \in \{1, 2\}$ has $\epsilon \geq \frac{4n_2}{n_1+n_2} = 2$. This follows from the no-go theorem that relies on the factorization of the complex structure sector and we have used $n_1 = n_2 = 1$.

The last case $r_{1K} = 0, \forall K \in \{1, 2\}, \hat{r}_1^L = p_M = r_{2M} = r_{3M} = r_{4M} = 0, L \neq M, L, M \in \{1, 2\}$, can be dealt with using the no-go theorem that relies on the factorization in both complex and Kähler sector and has $\epsilon \geq \frac{9n_1+5n_2}{5n_1+n_2} = \frac{7}{3}$ since $n_1 = n_2 = 1$.

6.4.9 Case IX

In this case we again need several different no-go theorems.

The first class of solutions $r_{1K} = \hat{r}_\alpha^K = 0, \forall K \in \{1, 2\}, \forall \alpha \in \{1, 2, 3\}$ and $r_{21} = 0$ or $r_{22} = 0$ has only one non-vanishing metric flux, and the configuration is dual to one with only H -flux, so we find $\epsilon \geq \frac{27}{13}$.

The next solution $r_{2K} = \hat{r}_\alpha^K = 0, \forall K \in \{1, 2\}, \forall \alpha \in \{1, 2, 3\}$ gives $\epsilon \geq 2$ due to the factorization in the Kähler sector.

From the factorization in the complex structure sector we find $\epsilon \geq \frac{4n_2}{n_1+n_2} = 2$ for $n_1 = n_2 = 1$ for the third solution $r_{2K} = 0, \forall K \in \{1, 2\}$ and $\hat{r}_1^L = p_M = r_{2M} = 0, L \neq M, L, M \in \{1, 2\}$.

The next solution $r_{aK} = p_K = \sum_{\alpha} \hat{r}_{\alpha}^1 \hat{r}_{\alpha}^2 = 0, \forall a, K \in \{1, 2\}$ satisfies the condition for our no-go based on the factorization in the Kähler sector and we find $\epsilon \geq \frac{9}{5}$.

Using the factorization in both complex and Kähler sector we can show that the last solution $r_{1K} = 0, \forall K \in \{1, 2\}, \hat{r}_{\alpha}^L = p_M = r_{2M} = 0, \forall \alpha \in \{1, 2, 3\}, L \neq M, L, M \in \{1, 2\}$ has $\epsilon \geq \frac{9n_1+5n_2}{5n_1+n_2} = \frac{7}{3}$, where we used $n_1 = n_2 = 1$.

6.4.10 Case X

Here the solutions to the Bianchi identities can be grouped into five different classes.

The first class of solutions has two of the three rows in the r matrix equal to zero. The third one is arbitrary and can be identified with " r_{0K} " and our no-go theorem that relies on the factorization of the Kähler sector can be used to obtain $\epsilon \geq 2$.

The next case has partially non-vanishing entries in two rows and at least two columns. The entire third row is zero so that we have this time " $r_{0K} = 0$ " and $\epsilon \geq \frac{9}{5}$.

The third class encompasses four solutions that each have only three non-vanishing entries with one in each row. The non-vanishing r components for the four different cases are 1) $r_{11}, r_{24}, r_{33} \neq 0$, 2) $r_{12}, r_{23}, r_{34} \neq 0$, 3) $r_{13}, r_{22}, r_{31} \neq 0$, 4) $r_{14}, r_{21}, r_{32} \neq 0$. Each of these cases leads numerically to $\epsilon \approx 1.57721$. We leave it up to the interested reader to try to find a no-go theorem that gives this value using the residual symmetries and factorization in both Kähler and complex structure sector.

The fourth class of solutions has one of the four columns of the r matrix non-vanishing and arbitrary. Numerically one obtains $\epsilon \approx \frac{4}{3}$ and we leave it again

to the interested reader to find the corresponding no-go theorem.

The last class has two, three, or all four columns non-zero. The non-zero metric fluxes are not all independent but have to satisfy constraints that result from the Bianchi identities. The most generic case has all twelve entries in the r matrix non-zero. There are six constraints so that we are left with six independent metric fluxes. For this class there cannot be a bound on ϵ from a no-go theorem since one can find numerically solutions that have $\epsilon \approx 0$. We will discuss an explicit example in more detail in subsection 6.5.1.

6.4.11 Case XI

The solutions to the Bianchi identities can be again grouped into five classes.

The first class has $r_{11} = r_{21} = p_1 = \hat{r}_1^2 = \hat{r}_1^3 = \hat{r}_1^4 = 0$ and $2r_{13}r_{22} + 2r_{12}r_{23} - r_{14}r_{24} = 0$. From the factorization in the complex sector we find $\epsilon \geq \frac{4n_2}{n_1+n_2} = 2$, where now $n_1 = n_2 = 2$.

The next class has $r_{1K} = 0, \forall K \in \{1, 2, 3, 4\}$. The remaining NSNS fluxes r_{2K}, \hat{r}_1^K, p_K are constrained by the remaining Bianchi identities. We can use the factorization in the Kähler moduli sector since $r_{1K} = "r_{0K}" = 0$ to obtain $\epsilon \geq \frac{9}{5}$.

The third class has $r_{21} = \hat{r}_1^1 = 0$ and $r_{11} \neq 0$. The Bianchi identities then reduce to $p_K \hat{r}_1^K = r_{1K} \hat{r}_1^K = 0$ and $r_{22} = \frac{-r_{14} \hat{r}_1^3 - 2r_{12} \hat{r}_1^4}{2r_{11}}$, $r_{23} = \frac{r_{14} \hat{r}_1^2 + 2r_{13} \hat{r}_1^4}{2r_{11}}$, $r_{24} = \frac{r_{12} \hat{r}_1^2 - r_{13} \hat{r}_1^3}{r_{11}}$. None of the no-go theorems apply and numerically one finds $\epsilon \approx 0$. We will discuss this class and the next two that all allow for extremal points with very small ϵ in subsection 6.5.2 in greater detail.

The next class has $r_{11} = r_{21} = \hat{r}_1^1 = 0$. The remaining fluxes are again constrained by the Bianchi identities. Numerically we find vanishing ϵ in the limit

where we have $\hat{r}_1^K \approx 0, \forall K \in \{1, 2, 3, 4\}$. In this limit the only Bianchi identity that constrains the non-vanishing NSNS fluxes is $2r_{13}r_{22} + 2r_{12}r_{23} - r_{14}r_{24} = 0$. The last class of solutions has $r_{11} = 0$ but $r_{21} \neq 0$. Again there are several Bianchi identities that constrain the remaining fluxes. Nevertheless, it is generically possible to obtain small ϵ so that there cannot be a no-go theorem. We will present some of the details of our numerical studies of this case in the next section.

6.5 Examples with Small ϵ

In the previous section, we have shown that our no-go theorems rule out slow-roll inflation and de Sitter vacua in large classes of models. However, there were solutions to the Bianchi identities for the two $\mathbb{Z}_2 \times \mathbb{Z}_2$ orbifold models that escaped all no-go theorems. In this section we give explicit examples for the models that have regions in moduli space with (very) small ϵ that likely correspond to de Sitter extrema.⁵ Since we are now presenting explicit solutions rather than no-go theorems some words of caution are in order. “Adding metric fluxes” to an existing geometry only leads to a well-defined compact space if the fluxes are properly quantized. One way to explicitly construct a space with metric fluxes is the base-fiber splitting framework [211], [212], [209]. This framework also allows to derive the quantization conditions for all the NSNS fluxes. We will therefore check whether we can explicitly construct the compact spaces that lead to de Sitter extrema or whether the base-fiber approach is not compatible with the solutions to the Bianchi identities that

⁵Since the analysis is purely numerical, it is impossible to tell whether these are extrema with all derivatives of the potential vanishing. All we know is that the values we find are compatible with zero to within our working precision, but they might just be very shallow.

gave small ϵ . Further constraints on these models arise from the quantization of the RR fluxes and the tadpole cancellation condition. To ensure the validity of the supergravity approximation, one also has to check that the volume of the internal space is large in string units and that the string coupling is small. Since we always find at least one tachyonic direction for the extremal points with vanishing ϵ we will only consider the restrictions from the base-fiber constructions in the subsections below.

6.5.1 Standard orientifold of $\mathbb{Z}_2 \times \mathbb{Z}_2$

The solutions to the Bianchi identities that allow for vanishing ϵ have at least non-vanishing entries in two columns and all three rows. For simplicity we take the case where $r_{a3} = r_{a4} = 0, \forall a \in \{1, 2, 3\}$. Assuming that $r_{22} \neq 0$ we can solve the remaining Bianchi identities and find $r_{11} = -\frac{r_{12}r_{21}}{r_{22}}, r_{31} = \frac{r_{21}r_{32}}{r_{22}}$. Minimizing ϵ by letting all the moduli and remaining fluxes vary, we find $\epsilon \approx 10^{-21}$ for the following values⁶ of the fluxes and moduli⁷

$$\begin{aligned} \tilde{m} &\approx -.2026, m_1 = m_2 = m_3 \approx .6990, e_1 = e_2 = e_3 \approx -1.076, e_0 = 0, \\ p_1 = p_2 &= 0, p_3 = p_4 \approx -1.310, r_{12} \approx .6215, r_{21} \approx .5004, r_{22} \approx -.02231, \\ r_{32} &\approx -.1930, \xi^1 \approx -.1504, \xi^2 \approx 2.682, \xi^3 + \xi^4 \approx -2.573, \\ u^1 = u^2 = u^3 &\approx 1.336, e^D \approx .3481, \mathcal{Z}^1 \approx .1845, \mathcal{Z}^2 \approx 2.333, \mathcal{Z}^3 = \mathcal{Z}^4 \approx .3810, \\ v^1 &\approx 2.202, v^2 \approx 18.73, v^3 \approx 4.023. \end{aligned}$$

⁶For cosmetic reasons, the following values are rounded to four digits and give $\epsilon \approx 10^{-4}$.

⁷The C_3 axions ξ^K appear in the potential only through the linear combinations $p_K \xi^K$ and $r_{aK} \xi^K$. Since in this simple case r has rank two we can stabilize only three linear combinations of them. In particular $\xi^3 - \xi^4$ is a flat direction. By allowing for at least three non-vanishing columns in the r matrix one finds examples without flat directions.

Since we have $\epsilon \approx 0$ this corresponds to a de Sitter extremum. Calculating the η parameter for this solution we find $\eta \approx -3.7$. So this solution is not a minimum but rather a saddle point. From the mass matrix for the moduli one sees that there is exactly one tachyonic direction that is a mixture of several moduli including the axions. We have looked at several extremal points for this model but always found at least one tachyonic direction with $\eta \lesssim -2.4$. We did not pursue this model further since it is not compatible with the base-fiber construction mentioned above. Splitting the compact space into a base and a fiber always results in an r matrix that has one row equal to zero. Due to the factorization in the Kähler sector we therefore find $\epsilon \geq \frac{9}{5}$ for all models that can be obtained from the base-fiber construction. In a related work [204] the authors searched for slow-roll inflation and de Sitter vacua in coset spaces [223], [224], [225]. They found that for an orientifold of $SU(2) \times SU(2)$ one can obtain de Sitter extrema with one tachyonic direction. This orientifold of $SU(2) \times SU(2)$ can be thought of as a $\mathbb{Z}_2 \times \mathbb{Z}_2$ quotient of T^6 with metric fluxes as was discussed in [215]. This means that at least a subset of the compact spaces exists although it is not possible to obtain them from the base-fiber construction. The authors of [204] also checked whether the no-go theorems related to the η parameter [227], [228] can be applied to their $SU(2) \times SU(2)$ orientifold but found that this is not the case. It would be interesting to study this model further to verify whether all solutions to the Bianchi identities that give small ϵ have a corresponding compact space and whether it is possible to find de Sitter vacua that have no tachyonic directions.

6.5.2 Non-standard orientifold of $\mathbb{Z}_2 \times \mathbb{Z}_2$

For the non-standard orientifold projection we can explicitly construct solutions to the Bianchi identities that lead to vanishing ϵ . The two interesting cases have the 2 dimensional submanifolds spanned by the 3 and 5 or 4 and 6 directions as base and the other four transverse directions as fiber. The first case leads to $p_1, p_4, r_{11}, r_{14}, r_{23}, \hat{r}_1^2$ fluxes with all other NSNS fluxes equal to zero. The non-zero fluxes have to satisfy the Bianchi identity $2r_{11}r_{23} - r_{14}\hat{r}_1^2 = 0$. The second case with the 4 and 6 direction as base allows for $p_1, p_4, r_{11}, r_{14}, r_{22}, \hat{r}_1^3$ fluxes with all other NSNS fluxes equal to zero. The non-zero fluxes have to satisfy the Bianchi identity $2r_{11}r_{22} - r_{14}\hat{r}_1^3 = 0$. This case is related to the first one by symmetry so we will only focus on the first case with the 3 and 5 direction as base. In the first case we can solve the remaining Bianchi identity $2r_{11}r_{23} = r_{14}\hat{r}_1^2$ by setting one of the metric fluxes appearing on either side to zero. If we have $r_{23} = \hat{r}_1^2 = 0$ or $r_{11} = r_{14} = 0$ our no-go theorems based on the factorization in the Kähler sector give $\epsilon \geq 2$ and $\epsilon \geq \frac{9}{5}$, respectively. The other two possibilities $r_{11} = \hat{r}_1^2 = 0$ and $r_{23} = r_{14} = 0$ give numerically $\epsilon \approx \frac{4}{3}$ and $\epsilon \approx .2$, so that we focus on solutions that have $2r_{11}r_{23} = r_{14}\hat{r}_1^2 \neq 0$. For $2r_{11}r_{23} = r_{14}\hat{r}_1^2 < 0$ we find numerically that $\epsilon \geq .2$ where the lower bound is attained in the limit where $r_{23} = r_{14} \approx 0$. So the only solution to the Bianchi identities that leads to vanishing ϵ is $2r_{11}r_{23} = r_{14}\hat{r}_1^2 > 0$. The quantization condition in this case forces $2r_{11}r_{23} = r_{14}\hat{r}_1^2 = n^2\pi^2$, $n \in \mathbb{Z}$. There are two different solutions. For $n = 2k$ even we find

$$r_{14} = \frac{n_1}{n_2}r_{11}, \hat{r}_1^2 = \frac{4k^2\pi^2n_2}{n_1r_{11}}, r_{23} = \frac{2k^2\pi^2}{r_{11}},$$

$$p_1 = n_2 \left(12\sqrt{2} + \frac{p_4}{n_1} \right), \quad k, n_1, n_2 \in \mathbb{Z}. \quad (6.5.1)$$

For $n = 2k + 1$ odd we find

$$\begin{aligned} r_{14} &= \pm r_{11}, \hat{r}_1^2 = \pm \frac{(2k+1)^2 \pi^2}{r_{11}}, r_{23} = \frac{(2k+1)^2 \pi^2}{2r_{11}}, \\ p_1 &= 6 \left(\sqrt{2} n_1 + n_2 r_{11} \right), p_4 = \mp 6 \left(\sqrt{2} n_1 - n_2 r_{11} \right), \quad k, n_1, n_2 \in \mathbb{Z}. \end{aligned} \quad (6.5.2)$$

Note that not all metric fluxes are quantized. r_{11} can take arbitrary values. Both solutions respect the symmetry arising from shifting the B field and H flux. Under a shift of $u^1 \rightarrow u^1 + a^1$ we have $p_1 \rightarrow r_{11} a^1$ and $p_4 \rightarrow r_{14} a^1$ so that we can set one of p_1 and p_4 equal to zero. We will set $p_4 = 0$ and minimize ϵ numerically for integers $k, n_1, n_2 \in \mathbb{Z}$. One particular solution for $n = 2k$ even with $\epsilon \approx 10^{-19}$ is⁸

$$\begin{aligned} \tilde{m} &\approx -3.74854, m_1 \approx -32.5482, m_2 \approx -22.5086, \\ e_1 &\approx 2.76717, e_2 \approx -2.92192, e_0 \approx -.251057, \\ k &= 1, n_1 = -3, n_2 = -1, r_{11} \approx -1.62809, \\ \xi^1 &\approx -6.39013, \xi^2 \approx \text{unstabilized}, \xi^3 \approx -1.66584, \xi^4 \approx -15.7204, \\ u^1 &\approx -2.49321, u^2 \approx 3.16322, v^1 \approx 3.32339, v^2 \approx 11.6507, \\ e^D &\approx .0745145, \mathcal{Z}^1 \approx .413947, \mathcal{Z}^2 \approx 38.0222, \mathcal{Z}^3 \approx .360619, \mathcal{Z}^4 \approx 3.65332. \end{aligned}$$

This particular solution has one tachyonic direction and $\eta \approx -2.5$. The tachyon is a mixture of several moduli including the axions. We have scanned over ranges where the flux quanta n_1, n_2, k are of order 1 for both n even and odd and found dozens of solutions. All of these solutions had at least one

⁸The following values are rounded to six digits and give $\epsilon \approx 10^{-4}$. Note also, that similar to the previous case, the C_3 axions ξ^K appear in the potential only through the linear combinations $p_K \xi^K$ and $r_{aK} \xi^K$. Since for this model r has two rows we can stabilize only three linear combinations of them. In particular for $p_4 = 0$ we can stabilize $p_1 \xi^1$, $r_{11} \xi^1 + r_{14} \xi^4$ and $r_{23} \xi^3$. This means that ξ^2 remains unstabilized.

tachyonic direction that generically is a mixture of all moduli. We generically found that $\eta \lesssim -2.4$ and solutions close to that bound have only one single tachyonic directions. The no-go theorems of [227], [228] cannot be applied to this particular model since we have D-terms. It would be very interesting to understand this tachyonic direction that appears in both of the models in this section better. We, of course, cannot rule out that there are solutions corresponding to metastable de Sitter vacua since we only did a numerical study but due to the large number of solutions that all have this tachyonic direction with roughly the same value for η that furthermore is independent of fluxes, we suspect that this model has no metastable de Sitter vacua. We have examined the vicinity of our extrema in which ϵ is still small to see whether this enables us to find small $|\eta|$ to satisfy the conditions for slow-roll inflation, but we found that η changes at most by a factor of two in this region. We have also minimized ϵ with constraints ensuring small $|\eta|$, but have not been able to find small ϵ in this case. We take this as a strong indication that these models are incompatible with slow-roll inflation. However, we do not have an analytic proof of this, and it would be very interesting to investigate this further. We will leave this for future research.

6.6 Summary

We have explored the possibility of slow-roll inflation and de Sitter vacua in type IIA compactifications that include standard NSNS 3-form fluxes, RR fluxes, D6-branes and O6-planes as well as metric fluxes. We have derived a set of no-go theorems based on the dependence of the potential on the dilaton, volume, Kähler and complex structure moduli, extending previous work by HKTT [198]. Theorems of this kind are valuable because they specify the

minimal set of ingredients required to have slow-roll inflation or de Sitter vacua in this type of compactifications, or put differently they rule out entire regions in the vast landscape of solutions of string theory. To demonstrate their usefulness, we applied these no-go theorems to toroidal orientifolds with abelian orbifold groups generated by rotations and reflections, that, in the absence of fluxes and after orientifolding, preserve $\mathcal{N} = 1$ supersymmetry. As we showed, the application of the no-go theorems is straightforward in some cases while in others T-dualities and field redefinitions play a crucial role. We find that under the assumptions made in deriving the no-go theorems, the slow-roll parameter ϵ is bounded from below in all models of this class except the two $\mathbb{Z}_2 \times \mathbb{Z}_2$ cases. In those cases, we have succeeded in finding regions of parameter space where the slow-roll parameter ϵ is very small numerically, but unfortunately η turns out to be such that inflation is much too short, making these compactifications uninteresting from a cosmological perspective.

While it would be more satisfying and insightful to have no-go theorems for ϵ and η simultaneously, the ones obtained in this work are exclusively for ϵ . Our exploration of the range of η has always been numerical. We either computed η where ϵ had already been found to be small or have failed to find a small value for ϵ when we restricted the minimization procedure to keeping η small. Thus, although we are confident of our results we lack the insight as to why the necessary conditions for small η are not compatible with those for small ϵ .

There are several effects we have not considered. We have ignored twisted sector modes and blow-up modes. We have also ignored more general brane configurations such as backreacting D6-branes that do not wrap rigid cycles and are far from their static configuration, coisotropic branes, or NSNS

sources. Even though this is by no means guaranteed, all of these ingredients might render the no-go theorems invalid and may be interesting to investigate further. We leave this for future work.

There is an orthogonal line of research pursued in [229] that comes to similar conclusions. While we have not had the chance to do so, we think it would be interesting to understand if there is a relation between the two.

Chapter 7

Conclusions

Since its discovery almost half a century ago, the cosmic microwave background radiation has greatly improved our understanding of the very early universe. With the help of current and future experiments this trend is likely to continue. The detection of a B-mode signal would provide valuable information that may allow us to verify and maybe even explore the details of inflation. As explained in Section 2.4, it would teach us about the distance the field must have traveled, about the energy scale of inflation, and as we saw in Chapter 3, it would imply that the graviton has a mass of less than 10^{-30} eV. Though slightly less exciting, even in the case that no primordial B-mode signal is observed, one can still draw a valuable conclusion: the simplest models of inflation, a minimally coupled scalar with monomial potential, such as the linear potential studied in Chapter 5, the popular $m^2\phi^2$ potential, as well as all other models of large field inflation are ruled out. Searching for departures from Gaussianity will provide another useful tool to learn about the early universe. A detection of non-Gaussianities of the so-called local shape would for instance allow us to rule out all single field *slow-roll* inflationary models. Once non-Gaussianities are seen, studying their shape would contain a large amount of information and strongly constrain possible models. Another possibility is that isocurvature modes are discovered, ruling out single field models of inflation altogether. It may also be, however, that an explanation of the

experimental results necessitates nothing but a small field model of a single slowly rolling scalar field. In this case it seems very hard to make further progress.

The situation is rather similar to that in particle physics, where the Large Hadron Collider will start taking data soon. We are guaranteed to learn about electroweak symmetry breaking, we may discover supersymmetry, large extra dimensions, new strongly coupled sectors, and we may learn about the nature of dark matter by observing it in the form of missing energy and confirming its properties in direct detection experiments. While the experiment could provide us with a huge amount of information, it is equally possible that we will discover a Higgs and nothing else, in which case further progress will also become difficult.

Both in cosmology and in particle physics, we are guaranteed to make some progress in learning about fundamental physics over the course of the next decade or two, but to see how much progress we can make, we have to be patient and await the experimental results. Whatever the final conclusion may be, the next years promise to be interesting both in cosmology and particle physics.

Appendices

Appendix A

Conventions and Notations

A.1 Conventions for IIB

In this appendix we review our conventions, emphasizing differences with the existing literature.

A good starting point is the ten-dimensional string-frame action¹ [169]

$$S_{10} = \frac{1}{(2\pi)^7 \alpha^4} \int d^{10}x \sqrt{g_{string}} \left(e^{-2\Phi} R_{string} - \frac{1}{2} |dC_2|^2 \right) \quad (\text{A.1.1})$$

which after the rescaling to the ten-dimensional Einstein-frame metric $e^{-\Phi/2} g_{string, MN} = g_{E, MN}$ becomes

$$S_{10} = \frac{1}{(2\pi)^7 \alpha^4} \int d^{10}x \sqrt{g_E} \left(R_E - \frac{1}{2} g_s |dC_2|^2 \right), \quad (\text{A.1.2})$$

where we assumed that the axio-dilaton is $\tau = i/g_s$. Upon compactifying on a six-dimensional manifold Y , the resulting four-dimensional reduced Planck mass is

$$M_{pl}^2 = \frac{\int_Y \sqrt{g_E}}{(2\pi)^6 \alpha^3} \frac{1}{\alpha \pi} \equiv \frac{\mathcal{V}_E}{\alpha \pi}, \quad (\text{A.1.3})$$

where \mathcal{V}_E is the (dimensionless) Einstein volume of Y measured in units of $2\pi\sqrt{\alpha}$. When Y is (conformally equal to) a Calabi-Yau space, we have

$$\mathcal{V}_E = \frac{1}{6} \frac{\int J \wedge J \wedge J}{(2\pi)^6 \alpha^3} = \frac{1}{6} v^I v^J v^K \frac{\int \omega_I \wedge \omega_J \wedge \omega_K}{(2\pi)^6 \alpha^3} \equiv \frac{1}{6} v^I v^J v^K c_{IJK}, \quad (\text{A.1.4})$$

¹Remember that $2\kappa_{10}^2 = (2\pi)^7 \alpha^4$.

where ω_I for $I = 1, \dots, h^{1,1}$ are a basis of the cohomology $H^2(Y, \mathbb{Z})$ normalized such that

$$\int_{\Sigma_I} \omega_J = (2\pi)^2 \alpha \delta_J^I \quad (\text{A.1.5})$$

for a basis Σ_I of the dual homology $H_2(Y, \mathbb{Z})$. With the ansatz for the ten-dimensional RR two-form

$$C_2 = \frac{1}{2\pi} c(x) \omega, \quad (\text{A.1.6})$$

for some base two-cycle ω , we get a four-dimensional axion $c(x)$ with periodicity² 2π , as can be seen *e.g.* via S-duality starting from the world-sheet coupling $\int B/(2\pi\alpha)$. The axion decay constant of c is

$$\frac{f^2}{M_{pl}^2} = \frac{g_s}{12\mathcal{V}_E(2\pi)^2} \left[\frac{\int \omega \wedge * \omega}{(2\pi)^6 \alpha^3} \right]. \quad (\text{A.1.7})$$

The four-dimensional $\mathcal{N} = 1$ Kähler potential for the Kähler moduli is

$$K = -2 \log \mathcal{V}_E. \quad (\text{A.1.8})$$

A.2 Conventions for IIA

Our conventions largely follow [217].

Consider type IIA string theory on a Calabi-Yau three-fold X , equipped with a \mathbb{Z}_2 orientifold action which includes an anti-holomorphic involution σ . The cohomology of X then splits into even and odd parts, depending upon the behavior of each class under σ . We will take the following basis of representative forms:

²Note that this choice differs from that in [151], where the axion periodicity was $(2\pi)^2$.

- The zero-form 1,
- a set of odd two-forms ω_a , $a = 1, \dots, h_-^{1,1}$,
- a set of even two-forms μ_α , $\alpha = 1, \dots, h_+^{1,1}$,
- a set of even four-forms $\tilde{\omega}^a$, $a = 1, \dots, h_-^{1,1}$,
- a set of odd four-forms $\tilde{\mu}^\alpha$, $\alpha = 1, \dots, h_+^{1,1}$,
- a six form φ , odd under σ ,
- a set of even three-forms a_K , $K = 1, \dots, h^{2,1} + 1$,
- and a set of odd three-forms b^K , $K = 1, \dots, h^{2,1} + 1$.

Additionally, it turns out that we can always choose the a_K and b^K to form a symplectic basis such that the only non-vanishing intersections are

$$\int_X a_K \wedge b^J = \partial_K^J. \quad (\text{A.2.1})$$

Similarly, we can take the even-degree forms to obey

$$\int_X \varphi = 1, \quad \int_X \omega_a \wedge \omega_b \wedge \omega_c = \kappa_{abc}, \quad \int_X \omega_a \wedge \mu_\alpha \wedge \mu_\beta = \hat{\kappa}_{a\alpha\beta}, \quad (\text{A.2.2})$$

$$\int_X \omega_a \wedge \tilde{\omega}^b = \partial_a^b, \quad \int_X \mu_\alpha \wedge \tilde{\mu}^\beta = \partial_\alpha^\beta. \quad (\text{A.2.3})$$

Now let us describe the four-dimensional fields of this class of compactifications, restricting ourselves, for simplicity, to the bosonic sector. First we

have the Kähler moduli, parametrized by complex scalar fields $t^a = u^a + iv^a$ coming from the expansion

$$B + iJ = J_c = t^a \omega_a, \quad (\text{A.2.4})$$

where the complexified Kähler form J_c must be odd under σ . Note that the Kähler form $J = v^a \omega_a$ determines the compactification volume (in string frame) via

$$\mathcal{V}_6 = \frac{1}{3!} \int_X J \wedge J \wedge J = \frac{1}{6} \kappa_{abc} v^a v^b v^c. \quad (\text{A.2.5})$$

To describe the complex moduli, let us write the holomorphic three-form as

$$\Omega = \mathcal{Z}^K a_K - \mathcal{F}_K b^K. \quad (\text{A.2.6})$$

We will use conventions in which

$$i \int_X \Omega \wedge \bar{\Omega} = 1, \quad \sigma^* \Omega = \bar{\Omega}, \quad (\text{A.2.7})$$

so that the \mathcal{Z}^K are real functions of the complex moduli and \mathcal{F}_K are pure imaginary, and together they satisfy the constraint $\mathcal{Z}^K \mathcal{F}_K = -i/2$. We can now define a complexified version [217]

$$\Omega_c = C_3 + 2ie^{-D} \text{Re} \Omega = (\xi^K + 2ie^{-D} \mathcal{Z}^K) a_K, \quad (\text{A.2.8})$$

where $e^{-D} = \mathcal{V}_6^{1/2} e^{-\phi}$ contains the dilaton and we expand the periods of C_3 (which must be even under σ in order to survive the orientifold projection) as $C_3 = \xi^K a_K$. Note that we abuse notation somewhat here as we ignore other pieces which contribute to the ten-dimensional RR three-form potential C_3 , namely pieces that give rise to four-dimensional vectors and (local) pieces

that give the four-form RR flux, both of which will be discussed below. The complex moduli $N^K = \frac{1}{2}\xi^K + ie^{-D}\mathcal{Z}^K$ are then simply given by the expansion

$$\Omega_c = 2N^K a_K, \quad (\text{A.2.9})$$

and include the complex structure moduli of the metric, the dilaton, and the RR three-form periods.

There are also $h_+^{1,1}$ four-dimensional vectors from the decomposition of the RR three-form potential, which includes a contribution

$$C_3 = A^\alpha \wedge \mu_\alpha. \quad (\text{A.2.10})$$

We can now consider turning on fluxes. In the RR sector, this leads us to include

$$F_0 = \tilde{m}, \quad F_2 = m^a \omega_a, \quad F_4 = e_a \tilde{\omega}^a, \quad F_6 = \tilde{e} \varphi. \quad (\text{A.2.11})$$

From the NSNS sector, we can include the usual H -flux,

$$H = p_K b^K, \quad (\text{A.2.12})$$

but we can also consider generalized metric fluxes. For more details, please refer to sections 6.1.1 and 6.3.2, but for completeness we list the definitions of our parameters r_{aK} and \hat{r}_α^K ,

$$d\omega_a = -r_{aK} b^K, \quad d\mu_\alpha = -\hat{r}_\alpha^K a_K. \quad (\text{A.2.13})$$

Appendix B

Phase Conventions for Spherical Harmonics

Consider an irreducible representation of the group of rotations in three dimensions on a $2\ell+1$ -dimensional vector space with a basis $|\ell, m\rangle$ chosen such that

$$L^2 |\ell, m\rangle = \ell(\ell+1) |\ell, m\rangle, \quad (\text{B.0.1})$$

$$L_3 |\ell, m\rangle = m |\ell, m\rangle, \quad (\text{B.0.2})$$

where $L^2 = L_1^2 + L_2^2 + L_3^2$ is the quadratic casimir operator, and the L_i are the generators of rotations satisfying

$$[L_i, L_j] = i\epsilon_{ijk} L_k. \quad (\text{B.0.3})$$

The action of the raising and lowering operators, $L_{\pm} \equiv L_1 \pm iL_2$, is a priori only defined up to a phase:

$$L_{\pm} |\ell, m\rangle = e^{i\phi_{\pm}(\ell, m)} \sqrt{\ell(\ell+1) - m(m \pm 1)} |\ell, m \pm 1\rangle. \quad (\text{B.0.4})$$

In quantum mechanics this phase is usually fixed such that:

$$L_{\pm} |\ell, m\rangle = \sqrt{\ell(\ell+1) - m(m \pm 1)} |\ell, m \pm 1\rangle. \quad (\text{B.0.5})$$

This convention leads to the standard representation of rotation matrices and to spherical harmonics that behave under complex conjugation as:

$$Y_{\ell}^m(\hat{n})^* = (-1)^m Y_{\ell}^{-m}(\hat{n}). \quad (\text{B.0.6})$$

The spherical harmonics that behave under complex conjugation as:

$$Y_\ell^m(\hat{n})^* = Y_\ell^{-m}(\hat{n}) \quad (\text{B.0.7})$$

do not transform according to this standard representation of rotations but according to a representation that is equivalent to this standard one.

One choice of phase in equation (B.0.4) that leads to spherical harmonics with the desired reality property is the following:

$$L_\pm |\ell, m\rangle = i(-1)^m \sqrt{\ell(\ell+1) - m(m \pm 1)} |\ell, m \pm 1\rangle. \quad (\text{B.0.8})$$

Together with the usual action of L_3 on the state vectors:

$$L_3 |\ell, m\rangle = m |\ell, m\rangle \quad (\text{B.0.9})$$

this completely determines the representation and the only non-trivial rotation we need, namely the one about the 2-axis by an angle θ , in this representation takes the form:

$$D_{mm'}^{(\ell)}(e^{-i\theta L_2}) = i^{m^2 - m'^2} \sqrt{\frac{(\ell+m)!(\ell-m)!}{(\ell+m')!(\ell-m')!}} \sin^{2\ell} \left(\frac{\theta}{2} \right) \sum_r (-1)^{\ell+m-r} \binom{\ell+m'}{r} \binom{\ell-m'}{r-m-m'} \cot^{2r-m-m'} \left(\frac{\theta}{2} \right) \quad (\text{B.0.10})$$

As usual we define the spherical harmonics as $Y_\ell^m(\hat{n}) = \langle \hat{n} | \ell, m \rangle$. Using:

$$|\hat{n}\rangle = \mathcal{D}(S(\hat{n})) |\hat{z}\rangle \quad (\text{B.0.11})$$

it then follows that:

$$Y_\ell^m(\hat{n})^* = D_{mm'}^{(\ell)}(S(\hat{n})) Y_\ell^{m'}(\hat{z})^* = \sqrt{\frac{2\ell+1}{4\pi}} D_{m0}^{(\ell)}(S(\hat{n})) \quad (\text{B.0.12})$$

or equivalently that:

$$Y_\ell^m(\hat{n}) = \sqrt{\frac{2\ell+1}{4\pi}} D_{m0}^{(\ell)}(S(\hat{n}))^*. \quad (\text{B.0.13})$$

Together with the explicit form of the rotation matrix given in equation (B.0.10) equation (B.0.13) gives the following formula for the spherical harmonics:

$$Y_\ell^m(\hat{n}) = (-i)^{m^2} e^{im\phi} \sqrt{\frac{2\ell+1}{4\pi}} \sqrt{\frac{(\ell+m)!(\ell-m)!}{\ell!^2}} \sin^{2\ell} \left(\frac{\theta}{2} \right) \sum_r (-1)^{\ell+m-r} \binom{\ell}{r} \binom{\ell}{r-m} \cot^{2r-m} \left(\frac{\theta}{2} \right). \quad (\text{B.0.14})$$

This expression can easily be checked to have the desired reality property:

$$Y_\ell^m(\hat{n})^* = Y_\ell^{-m}(\hat{n}), \quad (\text{B.0.15})$$

and clearly they transform under rotations according to:

$$Y_\ell^m(R\hat{n}) = \sum_m D_{mm'}^{(\ell)}(R)^* Y_\ell^{m'}(\hat{n}). \quad (\text{B.0.16})$$

For some applications it is convenient to rewrite this in terms of associated Legendre polynomials:

$$Y_\ell^m(\hat{n}) = (-i)^{m^2} \sqrt{\frac{2\ell+1}{4\pi}} \sqrt{\frac{(\ell-m)!}{(\ell+m)!}} e^{im\phi} P_\ell^m(\cos \theta), \quad (\text{B.0.17})$$

where:

$$P_\ell^m(\cos \theta) = \frac{(\ell+m)!}{\ell!} \sin^{2\ell} \left(\frac{\theta}{2} \right) \sum_r (-1)^{\ell+m-r} \binom{\ell}{r} \binom{\ell}{r-m} \cot^{2r-m} \left(\frac{\theta}{2} \right) \quad (\text{B.0.18})$$

are the associated Legendre polynomials and satisfy

$$P_\ell^{-m}(\mu) = (-1)^m \frac{(\ell - m)!}{(\ell + m)!} P_\ell^m(\mu). \quad (\text{B.0.19})$$

A useful property is the orthonormality condition

$$\int d^2\hat{n} Y_\ell^m(\hat{n}) Y_{\ell'}^{m'*}(\hat{n}) = \delta_{\ell\ell'} \delta_{mm'}. \quad (\text{B.0.20})$$

In a way similar to how we defined the ordinary spherical harmonics we can define the spin-weighted spherical harmonics as:

$${}_s Y_\ell^m(\hat{n})^* = \sqrt{\frac{2\ell + 1}{4\pi}} D_{m-s}^{(\ell)}(S(\hat{n})). \quad (\text{B.0.21})$$

This leads to an explicit expression for ${}_2 Y_\ell^m(\hat{n})$ of the form:

$$\begin{aligned} {}_2 Y_\ell^m(\hat{n}) &= (-i)^{m^2} e^{im\phi} \sqrt{\frac{2\ell + 1}{4\pi}} \sqrt{\frac{(\ell + m)!(\ell - m)!}{(\ell + 2)!(\ell - 2)!}} \sin^{2\ell} \left(\frac{\theta}{2} \right) \\ &\quad \sum_r (-1)^{\ell+m-r} \binom{\ell - 2}{r} \binom{\ell + 2}{r + 2 - m} \cot^{2r+2-m} \left(\frac{\theta}{2} \right). \end{aligned} \quad (\text{B.0.22})$$

It is again useful to have expressions for the spherical harmonics of spin 2 in terms of associated Legendre polynomials rather than sums. One has

$$\begin{aligned} {}_2 Y_\ell^m(\hat{n}) &= (-i)^{m^2} e^{im\phi} \sqrt{\frac{2\ell + 1}{4\pi}} \sqrt{\frac{(\ell - m)!(\ell - 2)!}{(\ell + m)!(\ell + 2)!}} \\ &\quad \left(-\ell(\ell + 1) P_\ell^m(\mu) + \frac{2(m + \mu)(m - \ell\mu)}{1 - \mu^2} P_\ell^m(\mu) + \frac{2(m + \mu)(\ell + m)}{1 - \mu^2} P_{\ell-1}^m(\mu) \right), \end{aligned} \quad (\text{B.0.23})$$

where as usual $\mu = \cos \theta$. They again satisfy an orthonormality condition

$$\int d^2\hat{n} {}_2 Y_\ell^m(\hat{n}) {}_2 Y_{\ell'}^{m'*}(\hat{n}) = \delta_{\ell\ell'} \delta_{mm'}. \quad (\text{B.0.24})$$

Appendix C

Induced Shift of the Four-Cycle Volume

In this appendix we address the issue raised in subsection 5.6.2: the inflationary energy can correct the warped volumes of four-cycles in the compact space, leading to corrections to the moduli potential, and hence inducing new terms in the inflaton potential itself.

More specifically, if an NS5-brane wraps some cycle Σ , then a nonvanishing integral $\int_{\Sigma} C_2 \neq 0$ leads to the presence of energy that is localized near Σ in the compact space; this energy corresponds to the increased tension of the NS5-brane. Moreover, there is a corresponding induced D3-brane charge via the coupling $\int C_2 \wedge C_4$. The increased tension creates a backreaction on the metric (and in particular, on the warp factor) of the compact space, while the induced charge sources five-form flux. We must determine whether these effects substantially correct the nonperturbative effects that are responsible, in our KKLT-like scenario, for stabilization of the Kähler moduli.

Whether the nonperturbative superpotential arises from gaugino condensation on D7-branes or from Euclidean D3-branes, it is exponentially sensitive to the warped volume of the four-cycle wrapped by these D-branes. Therefore, we will carefully consider the possibility of an inflaton-dependent shift of the warped volume of various four-cycles.

Concretely, we will consider a fivebrane/anti-fivebrane pair wrapping two homologous cycles, and will compute the leading correction to the volume

of a particular four-cycle in the same throat region as the fivebranes. This will serve as a conservative upper bound on the effect of the worldvolume flux, as more distant four-cycles would be more weakly affected.

It would be very interesting to perform a systematic study of this backreaction in the four-dimensional effective theory and in ten-dimensional supergravity/string theory. We leave this task for future investigation. In what follows, we simply show that the effect described above can be ameliorated by choosing an appropriate configuration.

There are two mechanisms to suppress the backreaction on a given four-cycle volume. A first improvement comes from choosing a setup in which the leading backreaction is due to a dipole as opposed to a monopole potential. This allows for a parametric suppression. The second improvement can be achieved by a carefully chosen geometry of the four-cycle under consideration. In general, this latter mechanism requires fine tuning.

The problem of estimating the backreaction from two-form flux on an NS5-brane pair may be simplified by a series of approximations. First, the inflaton-dependent backreaction is generated by the increased tension and the induced D3-brane charge of the NS5-branes, which may be understood as corresponding to some number of D3-branes (or anti-D3-branes) dissolved in the NS5-branes. In practice, it is much simpler to study the effect of the D3-branes themselves; this captures the leading inflaton-dependent contributions.

The configuration of interest involves an NS5-brane wrapping Σ , with

$$\frac{1}{(2\pi)^2\alpha'} \int_{\Sigma} C_2 \equiv N_w, \quad (\text{C.0.1})$$

as well as a distant NS5-brane wrapping a homologous cycle Σ' , but with opposite orientation. (We will refer to the latter object as the anti-NS5-brane.)

Next, we recall that the COBE normalization requires each fivebrane to be in a warped region. Let us denote by $N/2$ the amount of D3-brane charge that creates the background warping for each of the fivebranes.¹

In light of the above discussion, we may approximate the fivebrane by a stack of N_w D3-branes and the anti-fivebrane by a stack of N_w anti-D3-branes. Combining this with the background D3-brane charge, we conclude that a convenient proxy for our system consists of two stacks of D3-branes, which we call A and B respectively. The first consists of $N/2 + N_w$ D3-branes and the second of $N/2$ D3-branes and N_w anti-D3-branes, which we may more conveniently represent as $N/2 - N_w$ D3-branes and N_w brane-antibrane pairs.

Next, using the results of [180], we recognize that the leading backreaction effect comes from the total D3-brane charge on each stack, while the brane-antibrane pairs lead to subleading effects that are suppressed by powers of the warp factor. Thus, we can simplify even further, so that at last we are considering a supersymmetric system involving two stacks that contain $N/2 + N_w$ and $N/2 - N_w$ D3-branes, respectively.

Equipped with this much simpler system, we may now estimate the inflaton-dependent backreaction, by computing how the presence of the stacks A and B leads to a N_w -dependent change in the warped volume of some four-cycle.

We choose the usual D3-brane ansatz

$$ds_6^2 = \sqrt{H^{-1}(y)} ds_4^2 + \sqrt{H(y)} ds_6^2 \quad (\text{C.0.2})$$

$$\tilde{F}_5 = (1 + *)dH^{-1} \wedge \text{Vol}_4, \quad \Phi = \text{const}. \quad (\text{C.0.3})$$

¹More generally, one could consider different degrees of warping for each fivebrane; extending our considerations to this case is straightforward.

The resulting equation of motion is linear in $H(y)$. Therefore we may simply add the solutions obtained in the presence of either of the two individual stacks. Once the resulting warp factor is used to compute the volume of a four-cycle, the N_w dependence of the result gives us an estimate of the inflaton-dependence of the nonperturbative superpotential.

We tackle the problem in two steps of increasing complexity. First, in C.0.1 we give a very simple, (conformally) flat toy example in which the calculations are easy. This already shows the relevant features of the more complicated solution. The inflaton-dependent shift of the volume can be suppressed by having the distance between A and B much smaller than the distance between the four-cycle and either of A and B. This corresponds to a configuration in which the leading interaction is via a dipole. In addition, one can fine-tune the four-cycle embedding so that the N_w -dependent correction to its volume actually cancels.²

Then, in C.0.2, C.0.2.1 and C.0.3, we describe the case of a resolved conifold using the solution of [230],[231]. We consider a particular holomorphic embedding of a four-cycle and compute numerically the inflaton-dependence of its warped volume.

C.0.1 A simple illustration of the suppression mechanism

Consider two stacks of $N/2 \pm N_w$ D3-branes, called A and B, respectively, in conformally flat space $M_4 \times \mathbb{R}^6$. The A stack is located at the origin

²Although suppression from symmetry of the embedding appears unappealing because of the fine-tuning required, one should keep in mind that it could conceivably be enforced by a discrete symmetry of the compactification.

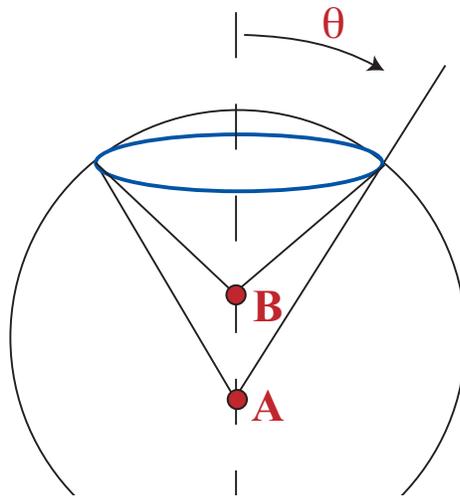


Figure C.1: This diagram illustrates the positions of the A and B stacks of D3-branes in \mathbb{R}^6 , the choice of the angular coordinate θ , and, in blue, a (topologically trivial) four-cycle.

of \mathbb{R}^6 and the B stack is located at some position $(u, 0, 0, 0, 0, 0)$ for³ $u \in \mathbb{R}^+$, where we have chosen spherical coordinates (see figure C.0.1) with the metric

$$ds_6^2 = dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\Omega_4^2) , \quad (\text{C.0.4})$$

where $d\Omega_4$ is the volume form of S^4 . With the usual D3-brane ansatz (C.0.2) one finds the solution

$$H = 1 + \frac{R_A^4}{r^4} + \frac{R_B^4}{(r^2 + u^2 - 2ru \cos \theta)^2} , \quad (\text{C.0.5})$$

$$R_{A,B} = 4\pi g_s \alpha'^2 \left(\frac{N}{2} \pm N_w \right) . \quad (\text{C.0.6})$$

Let us consider a (topologically trivial) four-cycle Σ_4 defined by⁴ $r = \mu$ and $\theta = \bar{\theta}$, whose unwarped volume is $V_4 = \frac{8\pi^2}{3} \bar{r}^4 \sin^4 \bar{\theta}$. The warped volume is

$$\begin{aligned} \int_{\Sigma_4} H(r, \theta) \sin^4 \theta d\Omega_4 &= V_4 H(\mu, \bar{\theta}) \\ &= V_4 \left[1 + \frac{R^4}{\mu^4} \left(1 + \mathcal{O} \left(\frac{u}{\mu} \right) \right) \right. \\ &\quad \left. - N_w \frac{u}{\mu} \left(4 \cos \bar{\theta} - 2 \frac{u}{\mu} + \mathcal{O} \left(\frac{u^2}{\mu^2} \right) \right) \right] . \end{aligned} \quad (\text{C.0.7})$$

From this result, one can see that a fine-tuning of the embedding can suppress the backreaction, *i.e.* if $\cos \bar{\theta} \simeq u/(2\mu)$. On the other hand, a parametric suppression is also clearly visible. Both the factors N_w/N and u/μ can be made small by construction. The former must be small in order for the background geometry to be at all trustworthy. The latter can be made small by arranging for all the four-cycles bearing nonperturbative effects to be far away in units

³We choose the letter u in analogy with the setup of the next subsections, where the distance between the A and the B stack is given by the resolution parameter of the resolved conifold.

⁴We choose the letter μ in analogy with the (usually complex) parameter appearing in other known embeddings of four-cycles in the conifold [232, 233].

of the separation of the fivebranes. Physically, this means that the four-cycle is sensitive only to the dipole field generated by the A and B stacks.

C.0.2 The conifold and its resolution

In the following we review some relevant definitions and conventions regarding the resolved conifold. The treatment is based on [234],[231]. The (singular) conifold is a cone over $T^{1,1}$ (the coset space $SU(2) \times SU(2)/U(1)$, which is topologically $S^2 \times S^3$). It is defined as the hyperspace in \mathbb{C}^4 that is a solution of the complex constraint

$$\det W \equiv \det \begin{pmatrix} X & U \\ V & Y \end{pmatrix} = XY - VU = 0 \quad (\text{C.0.8})$$

where (X, U, V, Y) are coordinates on \mathbb{C}^4 . The resolved conifold can be defined as the zero locus in $\mathbb{C}^4 \times \mathbb{CP}_1$ of the two linear complex equations

$$\begin{pmatrix} X & U \\ V & Y \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = 0 \quad (\text{C.0.9})$$

where (λ_1, λ_2) are complex coordinates on \mathbb{CP}_1 , *i.e.* they are identified by $(\lambda_1, \lambda_2) \simeq (\alpha\lambda_1, \alpha\lambda_2)$, for every $\alpha \in \mathbb{C}_*$. For every $W \neq 0$ (C.0.8) and (C.0.9) are equivalent, but when $W = 0$, *i.e.* at the tip, (λ_1, λ_2) are arbitrary and (C.0.9) defines a $\mathbb{CP}_1 \simeq S^2$. The radial direction is defined by

$$\text{Tr} W^\dagger W = r^2. \quad (\text{C.0.10})$$

One can check that the resolved conifold is an $\mathcal{O}(-1) \oplus \mathcal{O}(-1)$ bundle over \mathbb{CP}_1 with fiber \mathbb{C}_2 . If we define $\lambda \equiv \lambda_2/\lambda_1$, then we can choose coordinates on a patch $H_+ \equiv \{\lambda \neq 0\}$ of the resolved conifold using the following solution of (C.0.9):

$$W = \begin{pmatrix} -\lambda U & U \\ -\lambda Y & Y \end{pmatrix}. \quad (\text{C.0.11})$$

Defining $\tilde{\lambda} \equiv \lambda_1/\lambda_2$, one can find coordinates on a complementary patch $H_- \equiv \{\tilde{\lambda} \neq 0\}$ using the following solution:

$$W = \begin{pmatrix} X & -\tilde{\lambda}X \\ V & -\tilde{\lambda}V \end{pmatrix}. \quad (\text{C.0.12})$$

The complex structure is given by

$$\Omega = dU \wedge dY \wedge d\lambda = dV \wedge dX \wedge d\tilde{\lambda}. \quad (\text{C.0.13})$$

For later use, we introduce a parametrization of the resolved conifold in terms of real coordinates and give the explicit Kähler metric. We start by noting that a particular solution of (C.0.9) is given by

$$W_0 = \begin{pmatrix} 0 & r \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} \lambda_0 \\ 0 \end{pmatrix} \Rightarrow \lambda = 0. \quad (\text{C.0.14})$$

The base of the resolved conifold with respect to r can be obtained by acting on this solution with two $SU(2)$ transformations, L_1 and L_2 ,

$$L_i = \begin{pmatrix} \cos \frac{\theta_i}{2} e^{\frac{i}{2}(\psi_i + \phi_i)} & -\sin \frac{\theta_i}{2} e^{-\frac{i}{2}(\psi_i - \phi_i)} \\ \sin \frac{\theta_i}{2} e^{\frac{i}{2}(\psi_i - \phi_i)} & \cos \frac{\theta_i}{2} e^{-\frac{i}{2}(\psi_i + \phi_i)} \end{pmatrix}, \quad i = 1, 2 \quad (\text{C.0.15})$$

written in terms of Euler angles. This gives

$$W = L_1 W_0 L_2^\dagger, \quad \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = L_2 \begin{pmatrix} \lambda_0 \\ 0 \end{pmatrix} \quad (\text{C.0.16})$$

which depends only on the combination $\psi \equiv \psi_1 + \psi_2$. A Kähler metric on the resolved conifold with resolution parameter u is given by [231]

$$ds_6^2 = \kappa^{-1}(\rho) d\rho^2 + \frac{1}{9} \kappa(\rho) \rho^2 e_\psi^2 + \frac{1}{6} \rho^2 (e_{\theta_1}^2 + e_{\phi_1}^2) + \frac{1}{6} (\rho^2 + 6u^2) (e_{\theta_2}^2 + e_{\phi_2}^2), \quad (\text{C.0.17})$$

where

$$\kappa(\rho) = \frac{\rho^2 + 9u^2}{\rho^2 + 6u^2}. \quad (\text{C.0.18})$$

Here, following [231], we have defined a new radial coordinate ρ by

$$r^4 = \frac{4}{9}\rho^4 \left(\frac{2}{3}\rho^2 + 6u^2 \right), \quad (\text{C.0.19})$$

The explicit expression for the e 's is

$$e_\psi = d\psi + \sum_{i=1}^2 \cos \theta_i d\phi_i, \quad e_{\theta_i} = d\theta_i, \quad e_{\phi_i} = \sin \theta_i d\phi_i. \quad (\text{C.0.20})$$

C.0.2.1 The λUY embedding

In this subsection, we consider a particular holomorphic embedding of a four-cycle. A simple embedding would be $\lambda = \mu$ because this is trivial to solve for in real coordinates, $\tan \theta_2 = \mu$ and $\phi_2 = 0$ for $\mu \in \mathbb{R}$. The trouble is that this embedding reaches the tip, and in fact r is unconstrained. This can also be seen from

$$r^2 = (1 + |\lambda|^2) (|U|^2 + |Y|^2). \quad (\text{C.0.21})$$

As a result, this embedding does not give us the dipole suppression factor analogous to the (u/μ) of appendix C.0.1. The next-simplest embedding (whose defining equation depends on r) is

$$\lambda UY = \mu^3, \quad \mu \in \mathcal{R}, \quad (\text{C.0.22})$$

which in real coordinates gives

$$\psi = 0, \quad \sin(\theta_2) \sin(\theta_1) = 4 \frac{\mu^3}{r^2} \sim \frac{\mu^3}{\rho^3} \quad \text{for large } r. \quad (\text{C.0.23})$$

After some algebra (in particular, expressing $d\theta_2$ as a function of $d\theta_1$ and dr) one finds the metric in terms of $d\rho$, $d\theta_2$, $d\phi_1$ and $d\phi_2$. Its determinant

g_4^{ind} is independent of $\phi_{1,2}$ and reads

$$g_4^{ind} = \frac{\rho^2 \csc^4(\theta_2)}{20736 (9u^2 + \rho^2)^3 (\rho^4(9u^2 + \rho^2) \sin^2(\theta_2) - 54\mu^6)} \times \quad (\text{C.0.24})$$

$$\left((6u^2 + \rho^2) (9u^2\rho + \rho^3)^2 \cos(4\theta_2) \right.$$

$$- 4 \cos(2\theta_2) (486u^6\rho^2 + 189u^4\rho^4 + u^2(24\rho^6 - 324\mu^6) - 27\mu^6\rho^2 + \rho^8)$$

$$\left. + 3(54\rho^2(9u^6 + 2\mu^6) + 189u^4\rho^4 + 864u^2\mu^6 + 24u^2\rho^6 + \rho^8) \right)^2 .$$

We see that there is a boundary beyond which the sign of the determinant becomes negative, which thus defines the integration boundary in ρ, θ_2 -space:

$$\rho_{min}(\theta_2) = \sqrt{3} \mu \sqrt{A - \frac{u^2}{\mu^2} \cdot \left(1 - \frac{u^2}{\mu^2} \frac{1}{A}\right)} \quad (\text{C.0.25})$$

with

$$A = \sqrt[3]{\sqrt{-\csc^2(\theta_2) \left(2\frac{u^6}{\mu^6} - \csc^2(\theta_2)\right)} - \frac{u^6}{\mu^6} + \csc^2(\theta_2)} . \quad (\text{C.0.26})$$

In the limit $\frac{u}{\mu} \ll 1$ one thus has

$$\rho_{min}(\theta_2) \rightarrow \sqrt{3 \cdot 2^{1/3}} \mu \csc^{1/3}(\theta_2) \quad (\text{C.0.27})$$

C.0.3 The shift of the four-cycle volume

The solution with the branes smeared over the S^2 was obtained in [231]. Later, the solutions with pointlike sources were given in [230]. If the D3-brane stacks are at the north and south pole of the resolution S^2 , respectively, *i.e.*

$\theta_2^A = \pi - \theta_2^B = 0$, then one finds

$$H = \sum_l (2l+1) H_l(\rho) [L_A^4 P_l(\cos(\theta_2)) \quad (C.0.28)$$

$$+ L_B^4 P_l(\cos(\theta_2)) (-1)^l] , \quad (C.0.29)$$

$$H_l = \frac{2}{9u^2} \frac{C_\beta}{\rho^{2+2\beta}} {}_2F_1 \left(\beta, 1+\beta, 1+2\beta, -\frac{9u^2}{\rho^2} \right) , \quad (C.0.30)$$

$$C_\beta = \frac{(3u)^{2\beta} \Gamma(1+\beta)^2}{\Gamma(1+2\beta)} \quad \beta = \sqrt{1 + (3/2)l(l+1)} , \quad (C.0.31)$$

$$L_{A,B} = \frac{27}{16} 4\pi g_s (\alpha')^2 (N \mp N_w) , \quad (C.0.32)$$

where ${}_2F_1$ is a hypergeometric function. We want to integrate this warp factor on some supersymmetric four-cycle Σ_4 . This gives us an estimate of the inflation dependence of the gauge kinetic function of a stack of D7-branes wrapping Σ_4 .

Using this information and (C.0.29) and (C.0.30) we can now calculate the integral

$$\begin{aligned} \mathcal{V}_{warped} &= \int_{\Sigma_4} d\rho d\theta_2 d\phi_1 d\phi_2 \sqrt{-g_4^{ind}} H(\rho, \theta_2) \\ &= 4\pi^2 \int_{\Sigma_4} d\rho d\theta_2 \sqrt{-g_4^{ind}} H(\rho, \theta_2) \end{aligned} \quad (C.0.33)$$

numerically, as a function of μ . To facilitate this we will expand (C.0.29) up to $\ell = 1$, the dipole term, and take the large ρ limit

$$\begin{aligned} H(\rho, \theta_2) &= \frac{L^4}{2\rho^4} \left[1 + 3(2\ell+1) \frac{N_w}{N} \frac{u^2}{\rho^2} P_\ell(\cos \theta_2) \Big|_{\ell=1} \right] \\ &= H_{\ell=0}(\rho) + \delta H_{\ell=1}(\rho, \theta_2) \quad . \end{aligned} \quad (C.0.34)$$

We can now calculate

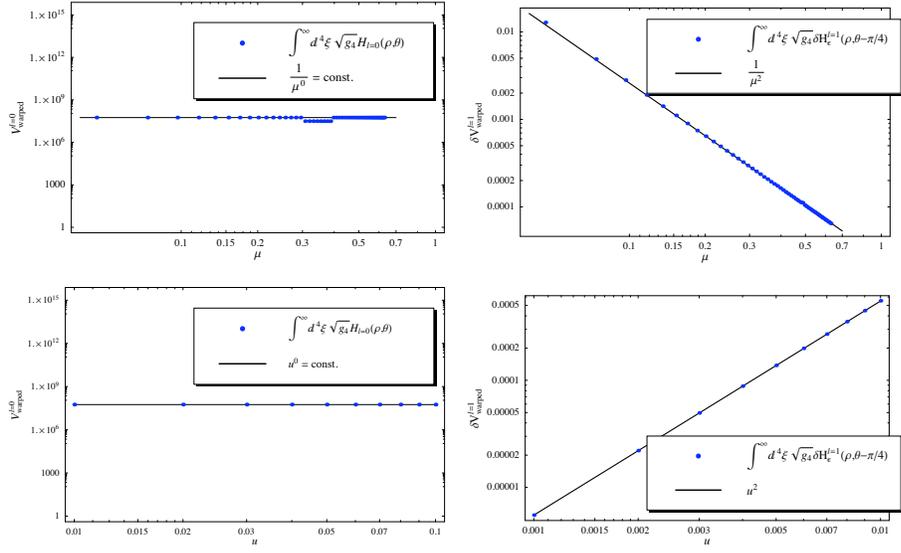


Figure C.2: 1st row: Plot of $\mathcal{V}_{warped}^{(0)}$ and $\delta\mathcal{V}_{warped}(\delta\theta_2)$ as functions of μ at constant $u = 0.01$. 2nd row: Plot of $\mathcal{V}_{warped}^{(0)}$ and $\delta\mathcal{V}_{warped}(\delta\theta_2)$ as functions of u at constant $\mu = 0.1$. The leading $\ell = 0$ term scales as $u^0\mu^0 = \text{const.}$ while the $\ell = 1$ dipole term scales as $(u/\mu)^2$. Note that the $\ell = 0$ scaling ensues only in the strictly noncompact limit (*i.e.* when the integration goes all the way $\rho \rightarrow \infty$), while for a finite cutoff, resembling a crude approximation to a compact setting, there remains a weak dependence of the $\ell = 0$ term on μ , of the form $(u/\mu)^\delta$, where $\delta \rightarrow 0$ for $\rho_{bulk} \rightarrow \infty$. For the example we have chosen $\epsilon \equiv N_w/N = 0.1$.

$$\mathcal{V}_{warped}^{(0)} = 4\pi^2 \int_0^\pi d\theta_2 \int_{\rho_{min}(\theta_2)}^{\rho_R} d\rho \sqrt{-g_4^{ind}} H_{\ell=0}(\rho) \quad (\text{C.0.35})$$

$$\delta\mathcal{V}_{warped}^{\ell=1}(\delta\theta_2) = 4\pi^2 \int_0^\pi d\theta_2 \int_{\rho_{min}(\theta_2)}^{\rho_R} d\rho \quad (\text{C.0.36})$$

$$\sqrt{-g_4^{ind}} \delta H_{\ell=1}(\rho, \theta_2 + \delta\theta_2) \quad (\text{C.0.37})$$

where $\rho_R \gg 1$ denotes a UV cutoff to compactify the resolved conifold geometry for the purpose of integration, and $\delta\theta_2$ denotes the angular misalignment of the D3-brane dipole configuration with respect to the four-cycle symmetry axis at $\theta_2 = \pi/2$.

As $\sqrt{g_4^{ind}}$ is a symmetric function with respect to $\theta_2 = \pi/2$ and $\delta H_{\ell=1}(\rho, \theta_2)$ is anti-symmetric with respect to $\theta_2 = \pi/2$, we immediately find $\delta \mathcal{V}_{warped}(\delta\theta_2 = 0) = 0$. So by fine-tuning a \mathbb{Z}_2 -symmetric configuration we can forbid the $\ell = 1$ term in the warped volume, whose corrections in this case start with the $\ell = 2$ quadrupole terms.

We will now display the numerical results for the case $\delta\theta_2 = -\pi/4$ in which the $\ell = 1$ term will not vanish under the integral, and compare the scaling with μ between $\delta \mathcal{V}_{warped}(\delta\theta_2)$ and $\mathcal{V}_{warped}^{(0)}$. This is displayed in Fig. C.2. We see clearly that the leading $\ell = 0$ term scales as $u^0 \mu^0 = const.$ while the $\ell = 1$ dipole term scales as $(u/\mu)^2$. Therefore, the $\ell = 1$ dipole term has a parametric suppression $(u/\mu)^2$ relative to the leading $\ell = 0$ term, and can therefore be made parametrically small (even in the non- \mathbb{Z}_2 -symmetric general situation) in the limit where the four-cycle recedes far from the resolution S^2 (*i.e.* in the limit of large μ/u).

Let us finally note that this relative suppression of the $\ell = 1$ term with $(u/\mu)^2$ might have been guessed without any integration, as the integration boundary tells us that $\rho_{min}(\theta_2) \geq \rho_{min}(\pi/2)$, which corresponds to $r > 2\mu^{3/2}$ or $\rho \gtrsim \mu$, and thus the relative scaling u^2/ρ^2 should be replaced by the scaling u^2/μ^2 .

Appendix D

The Kaluza-Klein Spectrum

In this appendix we obtain the (5+1)-dimensional effective action for a D5-brane wrapped on a two-cycle with $\int B \neq 0$. We show how a Kaluza-Klein reduction to four dimensions leads to masses that are suppressed with respect to the fluxless case. We then comment on the consequences of these light KK modes for axion monodromy inflation.

D.1 The effective theory

The DBI action for a D5-brane is

$$S = T_5 \int d^4x dy dz \sqrt{-\det (G_{ab}^{ind} + \mathcal{F}_{ab})}, \quad (\text{D.1.1})$$

where y, z are two coordinates in the internal space, which we take to be toroidal for the purpose of this derivation. The indices are defined as follows: worldvolume indices are $a, b = 0, \dots, 5$; spacetime indices are $\mu, \nu = 0, \dots, 3$ as usual; ten-dimensional indices are $M, N = 0, \dots, 9$; six-dimensional compact indices are $m, n = 4, \dots, 9$; and indices transverse to the D5-brane are $i, j = 6, \dots, 9$. We first expand the square root using

$$\begin{aligned} \sqrt{\det(M_0 + \delta M)} &= \sqrt{\det M_0} \left\{ 1 + \frac{1}{2} \text{Tr}(M_0^{-1} \delta M) \right. \\ &\quad \left. + \frac{1}{8} [\text{Tr}(M_0^{-1} \delta M)]^2 - \frac{1}{4} \text{Tr}(M_0^{-1} \delta M M_0^{-1} \delta M) + \dots \right\}. \end{aligned} \quad (\text{D.1.2})$$

We will consider a background with two-form flux on the two-cycle

$$\int \mathcal{F} = \int B = \int dy \wedge dz B_{yz}(x, y, z) = b(x) = b, \quad (\text{D.1.3})$$

i.e. the four-dimensional axion field $b(x)$ has a homogeneous expectation value that is approximately constant, up to terms suppressed by the slow-roll parameters. So the background is given by

$$B_{MN} = b\delta_{My}\delta_{Nz} - b\delta_{Mz}\delta_{Ny}, \quad F_{ab} = 0, \quad (\text{D.1.4})$$

$$ds_{10}^2 = g_{\mu\nu}dx^\mu dx^\nu + g_{yy}dy^2 + g_{zz}dz^2 + 2g_{yz}dydz + g_{ij}dy^i dy^j. \quad (\text{D.1.5})$$

Hence

$$(M_0)_{ab} = \begin{pmatrix} g_{\mu\nu} & & \\ & g_{yy} & g_{yz} + b \\ & g_{zy} - b & g_{zz} \end{pmatrix}. \quad (\text{D.1.6})$$

The perturbations are

$$(\delta M)_{ab} = \partial_a X^i \partial_b X^j (g_{ij} + B_{ij}) + F_{ab} + \delta B_{ab}. \quad (\text{D.1.7})$$

The calculation is simplified by the block-diagonal form of the background M_0 . The 2×2 block is the sum of a symmetric and an antisymmetric piece that we call S and A respectively. We have that

$$\det(A + S) = \det(A) + \det(S), \quad (\text{D.1.8})$$

$$(S + A)^{-1} = S^{-1} \frac{\det(S)}{\det(A) + \det(S)} + A^{-1} \frac{\det(A)}{\det(A) + \det(S)}, \quad (\text{D.1.9})$$

which substantially simplifies the calculation. Using (D.1.2) we get at leading order

$$S = T_5 \int d^4x dy dz \sqrt{-g_4} \sqrt{g_2 + b^2} \left[1 + \frac{1}{2} \partial_\mu X^i \partial^\mu X_i \right. \quad (\text{D.1.10})$$

$$\left. + \frac{1}{2} \frac{g_2}{g_2 + b^2} (\partial_y X^i \partial_y X_i + \partial_z X^i \partial_z X_i) \right. \quad (\text{D.1.11})$$

$$\left. + \frac{1}{2} \frac{2b}{g_2 + b^2} (\partial_y X^i \partial_z X^j \delta B_{ij} + F_{yz} + \delta B_{yz}) \right] + \dots, \quad (\text{D.1.12})$$

where $g_2 \equiv g_{yy}g_{zz} - g_{yz}^2$. After a KK reduction one finds the four-dimensional kinetic and potential terms, in the first line, as well as the Kaluza-Klein mass terms, in the second line. The Kaluza-Klein masses in the presence of fluxes are

$$m_{bKK}^2 = \frac{g_2}{g_2 + b^2} m_{KK}^2, \quad (\text{D.1.13})$$

where m_{KK} are the Kaluza-Klein masses in the absence of fluxes. This leads to the central point of this appendix: for $b \gg 1$, the Kaluza-Klein masses are *suppressed* by a factor of $\sqrt{g_2}/b \simeq L^2/b \ll 1$.¹ This phenomenon is intuitively understood in the T-dual picture in which flux becomes the angle of the D-brane. A large flux means that the T-dual brane winds around the torus many times, and thus becomes quite long. The Kaluza-Klein reduction of the fields living on the worldvolume of the T-dual brane therefore produces b-suppressed Kaluza-Klein masses.

D.2 Effects of the light Kaluza-Klein modes

Throughout this paper we have been careful to work in parameter ranges for which the typical Kaluza-Klein mass scale m_{KK} obeys $m_{KK} \gg H$, as required for a consistent four-dimensional analysis of inflation. However, from (D.1.13) we learn that a subclass of Kaluza-Klein modes, namely those associated with transverse excitations of the fivebrane, have considerably smaller masses, $m_{bKK} \ll m_{KK}$. For the numerical examples we have considered, we find that, very roughly, $m_{bKK} \sim (f_c/f)H$, where f_c is a fiducial value of the

¹We have assumed for simplicity that the internal space is isotropic, with typical size $L\sqrt{\alpha}$.

decay constant, $f_c \sim 10^{-2} M_{pl}$. Therefore, for constructions with small values of f , the transverse excitations of the fivebrane can be lighter than H .

We leave a comprehensive study of this constraint for future work, as a proper implementation plausibly requires a more explicit compact model that we have been able to present in this work. In particular, one should carefully compute the Kaluza-Klein mass, incorporating anisotropy in the geometry, warping, and, as we have explained above, the effect of worldvolume two-forms. To accomplish this, one needs a reasonably explicit construction of the warped throat region, of the two-cycle within the throat, and of the gluing of the throat into the compact space, which are beyond the scope of this work.

In this appendix, we will restrict ourselves to some qualitative statements that explain how our inflationary analysis can be consistent even in parameter regimes for which m_{bKK} is slightly smaller than H . Broadly speaking, one might worry about corrections to the inflationary Lagrangian, and about new contributions to the cosmological perturbations. Concerning the first point, we remark that the excitations of the fivebrane depend on the inflaton expectation value only through their masses. Therefore, the primary correction to the background evolution from these light modes would come if large numbers of Kaluza-Klein particles were produced by the time-dependent background. In practice, the particle production is negligible, as can be seen by computing the adiabatic parameter \dot{m}_{bKK}/m_{bKK}^2 and substituting the constraints on the volume, and hence on the Kaluza-Klein mass, from §5.6.

More generally, let us stress that only a small subclass of the Kaluza-Klein modes (a small portion of the tower of excitations of the fivebrane) have masses smaller than H . From the viewpoint of the inflationary analysis, these fields constitute a small number of harmless spectators. These light fields

will fluctuate, absorbing energy, but this yields a very small correction unless the number of fields approaches $(M_{pl}/H)^2$. Moreover, any entropy perturbations produced by these fields can turn into visible isocurvature perturbations only if their decays are distinct from that of the inflaton. Although we have not specified a concrete reheating mechanism, one can argue that the most straightforward scenario involves visible sector degrees of freedom that are well-separated in the compact space from the inflationary fivebranes. Thus, we expect that excitations of the fivebranes will not give visible isocurvature perturbations, because they must first decay [235],[236] to degrees of freedom localized in the inflationary throat, just as the inflaton does, and will plausibly do so with rather similar couplings, as the modes correspond to small excitations of the NS5-brane that drives inflation.

Appendix E

Numerical Examples

In this appendix, we specify two different sets of intersection numbers and show the relevant formulas for the volumes. For these two toy models, we explicitly performed the moduli stabilization outlined in §5.6.5, finding numerical values leading to the dot in figure 5.8.

E.1 Intersection numbers: set I

We consider as a toy-model a manifold with $H_+^{1,1} = \text{span}(\omega^L, \omega^+)$ for the orientifold-even homology two-cycles and $H_-^{1,1} = \text{span}(\omega^-)$ for the orientifold-odd homology two-cycles. We assume the following simple set of intersection numbers¹

$$c_{LLL} = c_{LL+} = c_{+--} = 1, \tag{E.1.1}$$

with all the others vanishing. We believe that, although very simplistic, the above toy model captures the relevant features of more realistic constructions. Notice that the intersection numbers in a basis for the homology of the covering space of the orbifold, *i.e.* without a definite parity with respect to the orbifold projection, are just linear combinations of those given above.

¹As they stand they do not correspond to a Calabi-Yau manifold, but we do not expect this to affect the results.

Using the standard relations

$$\mathcal{V}_E = \frac{1}{6} c_{\alpha\beta\gamma} v^\alpha v^\beta v^\gamma, \quad \tau_\alpha = \partial_{v^\alpha} \mathcal{V}_E = \frac{1}{2} c_{\alpha\beta\gamma} v^\beta v^\gamma, \quad (\text{E.1.2})$$

one finds

$$v_L = \sqrt{2\tau_+}, \quad v_+ = \frac{\tau_L - \tau_+}{\sqrt{2\tau_+}}. \quad (\text{E.1.3})$$

and

$$\mathcal{V}_E = \frac{\sqrt{2\tau_+}}{2} \tau_L - \frac{\sqrt{2}}{6} \tau_+^{3/2}. \quad (\text{E.1.4})$$

E.2 Intersection numbers: set II

Again assuming $H_+^{1,1} = \text{span}(\omega^L, \omega^+)$ and $H_-^{1,1} = \text{span}(\omega^-)$, we consider the intersection numbers

$$c_{LLL} = c_{L++} = c_{+--} = 1, \quad (\text{E.2.1})$$

with all the others vanishing. We find

$$v_L = \frac{1}{\sqrt{2}} \left((\tau_L + \tau_+)^{1/2} + (\tau_L - \tau_+)^{1/2} \right), \quad v_+ = \frac{1}{\sqrt{2}} \left((\tau_L + \tau_+)^{1/2} - (\tau_L - \tau_+)^{1/2} \right) \quad (\text{E.2.2})$$

and

$$\mathcal{V}_E = \frac{1}{3\sqrt{2}} \left((\tau_L + \tau_+)^{3/2} + (\tau_L - \tau_+)^{3/2} \right). \quad (\text{E.2.3})$$

Bibliography

- [1] G. Gamow, “Expanding Universe and the Origin of Elements,” *Physical Review* **70** (Oct., 1946) 572–573.
- [2] R. A. Alpher, H. Bethe, and G. Gamow, “The Origin of Chemical Elements,” *Physical Review* **73** (Apr., 1948) 803–804.
- [3] G. Gamow, “The Origin of Elements and the Separation of Galaxies,” *Physical Review* **74** (Aug., 1948) 505–506.
- [4] G. Gamow, “The Evolution of the Universe,” *Nature* **162** (Oct., 1948) 680–682.
- [5] R. A. Alpher and R. Herman, “Evolution of the Universe,” *Nature* **162** (Nov., 1948) 774–775.
- [6] R. A. Alpher, R. Herman, and G. A. Gamow, “Thermonuclear Reactions in the Expanding Universe,” *Physical Review* **74** (Nov., 1948) 1198–1199.
- [7] R. A. Alpher, “A Neutron-Capture Theory of the Formation and Relative Abundance of the Elements,” *Physical Review* **74** (Dec., 1948) 1577–1589.
- [8] R. A. Alpher and R. C. Herman, “On the Relative Abundance of the Elements,” *Physical Review* **74** (Dec., 1948) 1737–1742.

- [9] R. A. Alpher, R. Herman, and G. A. Gamow, “Erratum: Thermonuclear Reactions in the Expanding Universe,” *Physical Review* **75** (Feb., 1949) 701–701.
- [10] R. A. Alpher and R. C. Herman, “Remarks on the Evolution of the Expanding Universe,” *Physical Review* **75** (Apr., 1949) 1089–1095.
- [11] A. A. Penzias and R. W. Wilson, “A Measurement of Excess Antenna Temperature at 4080 Mc/s.,” *Astrophys. J.* **142** (July, 1965) 419–421.
- [12] R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson, “Cosmic Black-Body Radiation.,” *Astrophys. J.* **142** (July, 1965) 414–419.
- [13] E. K. Conklin, “Velocity of the Earth with Respect to the Cosmic Background Radiation,” *Nature* **222** (June, 1969) 971–972.
- [14] P. S. Henry, “Isotropy of the 3 K Background,” *Nature* **231** (June, 1971) 516–518.
- [15] G. F. Smoot *et al.*, “Structure in the COBE differential microwave radiometer first year maps,” *Astrophys. J.* **396** (1992) L1–L5.
- [16] M. J. Rees, “Polarization and Spectrum of the Primeval Radiation in an Anisotropic Universe,” *Astrophys. J.* **153** (July, 1968) L1.
- [17] J. Kovac *et al.*, “Detection of polarization in the cosmic microwave background using DASI,” *Nature* **420** (2002) 772–787, astro-ph/0209478.
- [18] E. M. Leitch *et al.*, “Measuring Polarization with DASI,” *Nature* **420** (2002) 763–771, astro-ph/0209476.

- [19] E. M. Leitch *et al.*, “DASI Three-Year Cosmic Microwave Background Polarization Results,” *Astrophys. J.* **624** (2005) 10–20, astro-ph/0409357.
- [20] A. C. S. Readhead *et al.*, “Polarization Observations with the Cosmic Background Imager,” astro-ph/0409569.
- [21] D. Barkats *et al.*, “CMB Polarimetry using Correlation Receivers with the PIQUE and CAPMAP Experiments,” astro-ph/0503329.
- [22] T. E. Montroy *et al.*, “A Measurement of the CMB Spectrum from the 2003 Flight of BOOMERANG,” *Astrophys. J.* **647** (2006) 813, astro-ph/0507514.
- [23] S. Masi *et al.*, “Instrument, Method, Brightness and Polarization Maps from the 2003 flight of BOOMERanG,” astro-ph/0507509.
- [24] F. Piacentini *et al.*, “A measurement of the polarization-temperature angular cross power spectrum of the Cosmic Microwave Background from the 2003 flight of BOOMERANG,” *Astrophys. J.* **647** (2006) 833, astro-ph/0507507.
- [25] **Supernova Search Team** Collaboration, A. G. Riess *et al.*, “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant,” *Astron. J.* **116** (1998) 1009–1038, astro-ph/9805201.
- [26] **Supernova Search Team** Collaboration, A. G. Riess *et al.*, “Type Ia Supernova Discoveries at $z < 1$ From the Hubble Space Telescope: Evidence for Past Deceleration and Constraints on Dark Energy Evolution,” *Astrophys. J.* **607** (2004) 665–687, astro-ph/0402512.

- [27] A. G. Riess *et al.*, “New Hubble Space Telescope Discoveries of Type Ia Supernovae at $z > 1$: Narrowing Constraints on the Early Behavior of Dark Energy,” *Astrophys. J.* **659** (2007) 98–121, [astro-ph/0611572](#).
- [28] **Supernova Cosmology Project** Collaboration, M. Kowalski *et al.*, “Improved Cosmological Constraints from New, Old and Combined Supernova Datasets,” *Astrophys. J.* **686** (2008) 749–778, [0804.4142](#).
- [29] M. Colless, “First results from the 2dF galaxy redshift survey,” [astro-ph/9804079](#).
- [30] **The 2dFGRS** Collaboration, W. J. Percival *et al.*, “The 2dF Galaxy Redshift Survey: The power spectrum and the matter content of the universe,” *Mon. Not. Roy. Astron. Soc.* **327** (2001) 1297, [astro-ph/0105252](#).
- [31] **The 2DFGRS** Collaboration, M. Colless *et al.*, “The 2dF Galaxy Redshift Survey: Spectra and redshifts,” *Mon. Not. Roy. Astron. Soc.* **328** (2001) 1039, [astro-ph/0106498](#).
- [32] **The 2dFGRS** Collaboration, S. Cole *et al.*, “The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final dataset and cosmological implications,” *Mon. Not. Roy. Astron. Soc.* **362** (2005) 505–534, [astro-ph/0501174](#).
- [33] **SDSS** Collaboration, C. Stoughton *et al.*, “The Sloan Digital Sky Survey: Early data release,” *Astron. J.* **123** (2002) 485–548.
- [34] **SDSS** Collaboration, M. Tegmark *et al.*, “The 3D power spectrum of galaxies from the SDSS,” *Astrophys. J.* **606** (2004) 702–740, [astro-ph/0310725](#).

- [35] **SDSS** Collaboration, D. J. Eisenstein *et al.*, “Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies,” *Astrophys. J.* **633** (2005) 560–574, astro-ph/0501171.
- [36] **SDSS** Collaboration, K. N. Abazajian *et al.*, “The Seventh Data Release of the Sloan Digital Sky Survey,” *Astrophys. J. Suppl.* **182** (2009) 543–558, 0812.0649.
- [37] A. Challinor and H. Peiris, “Lecture notes on the physics of cosmic microwave background anisotropies,” *AIP Conf. Proc.* **1132** (2009) 86–140, 0903.5158.
- [38] S. Weinberg, “Cosmology,”. Oxford, UK: Oxford Univ. Pr. (2008) 593 p.
- [39] P. J. E. Peebles and J. T. Yu, “Primeval Adiabatic Perturbation in an Expanding Universe,” *Astrophys. J.* **162** (Dec., 1970) 815.
- [40] M. L. Wilson and J. Silk, “On the anisotropy of the cosmological background matter and radiation distribution. I - The radiation anisotropy in a spatially flat universe,” *Astrophys. J.* **243** (Jan., 1981) 14–25.
- [41] J. R. Bond and G. Efstathiou, “Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter,” *Astrophys. J.* **285** (1984) L45–L48.
- [42] N. Vittorio and J. Silk, “Fine-scale anisotropy of the cosmic microwave background in a universe dominated by cold dark matter,” *Astrophys. J.* **285** (1984) L39–L43.

- [43] J. R. Bond and G. Efstathiou, “The statistics of cosmic background radiation fluctuations,” *Mon. Not. Roy. Astron. Soc.* **226** (1987) 655–687.
- [44] M. M. Basko and A. G. Polnarev, “Polarization and Anisotropy of the Primordial Radiation in an Anisotropic Universe,” *Soviet Astronomy* **24** (June, 1980) 268.
- [45] N. Kaiser, “Small-angle anisotropy of the microwave background radiation in the adiabatic theory,” *Mon. Not. Roy. Astron. Soc.* **202** (Mar., 1983) 1169–1180.
- [46] A. G. Polnarev, “Polarization and Anisotropy Induced in the Microwave Background by Cosmological Gravitational Waves,” *Soviet Astronomy* **29** (Dec., 1985) 607.
- [47] R. Crittenden, R. L. Davis, and P. J. Steinhardt, “Polarization of the microwave background due to primordial gravitational waves,” *Astrophys. J.* **417** (1993) L13–L16, [astro-ph/9306027](#).
- [48] D. D. Harari and M. Zaldarriaga, “Polarization of the microwave background in inflationary cosmology,” *Phys. Lett.* **B319** (1993) 96–103, [astro-ph/9311024](#).
- [49] R. A. Frewin, A. G. Polnarev, and P. Coles, “Gravitational Waves and the Polarization of the Cosmic Microwave Background,” *Mon. Not. Roy. Astron. Soc.* **266** (Jan., 1994) L21+, [arXiv:astro-ph/9310045](#).
- [50] D. Coulson, R. G. Crittenden, and N. G. Turok, “Polarization and Anisotropy of the Microwave Sky,” *Phys. Rev. Lett.* **73** (1994) 2390–2393, [astro-ph/9406046](#).

- [51] R. G. Crittenden, D. Coulson, and N. G. Turok, “Temperature - polarization correlations from tensor fluctuations,” *Phys. Rev.* **D52** (1995) 5402–5406, [astro-ph/9411107](#).
- [52] K. L. Ng and K.-W. Ng, “Large scale polarization of the cosmic microwave background radiation,” *Phys. Rev.* **D51** (1995) 364–368, [astro-ph/9305001](#).
- [53] A. Kosowsky, “Cosmic microwave background polarization,” *Ann. Phys.* **246** (1996) 49–85, [astro-ph/9501045](#).
- [54] M. Zaldarriaga and D. D. Harari, “Analytic approach to the polarization of the cosmic microwave background in flat and open universes,” *Phys. Rev.* **D52** (1995) 3276–3287, [astro-ph/9504085](#).
- [55] U. Seljak, “Measuring Polarization in the Cosmic Microwave Background,” *Astrophys. J.* **482** (June, 1997) 6, [arXiv:astro-ph/9608131](#).
- [56] M. Kamionkowski, A. Kosowsky, and A. Stebbins, “Statistics of Cosmic Microwave Background Polarization,” *Phys. Rev.* **D55** (1997) 7368–7388, [astro-ph/9611125](#).
- [57] M. Zaldarriaga and U. Seljak, “An All-Sky Analysis of Polarization in the Microwave Background,” *Phys. Rev.* **D55** (1997) 1830–1840, [astro-ph/9609170](#).
- [58] **WMAP** Collaboration, G. Hinshaw *et al.*, “Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Data Processing, Sky Maps, & Basic Results,” *Astrophys. J. Suppl.* **180** (2009) 225–245, 0803.0732.

- [59] L. Senatore, K. M. Smith, and M. Zaldarriaga, “Non-Gaussianities in Single Field Inflation and their Optimal Limits from the WMAP 5-year Data,” 0905.3746.
- [60] **WMAP** Collaboration, G. Hinshaw *et al.*, “Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Temperature analysis,” *Astrophys. J. Suppl.* **170** (2007) 288, astro-ph/0603451.
- [61] **WMAP** Collaboration, L. Page *et al.*, “Three year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Polarization analysis,” *Astrophys. J. Suppl.* **170** (2007) 335, astro-ph/0603450.
- [62] J. M. Bardeen, “Gauge-invariant cosmological perturbations,” *Phys. Rev.* (Oct., 1980) 1882–1905.
- [63] D. H. Lyth, “Large-scale energy-density perturbations and inflation,” *Phys. Rev.* (Apr., 1985) 1792–1798.
- [64] S. Weinberg, “Adiabatic modes in cosmology,” *Phys. Rev.* **D67** (2003) 123504, astro-ph/0302326.
- [65] S. Weinberg, “Must cosmological perturbations remain non-adiabatic after multi-field inflation?,” *Phys. Rev.* **D70** (2004) 083522, astro-ph/0405397.
- [66] S. Weinberg, “Can non-adiabatic perturbations arise after single-field inflation?,” *Phys. Rev.* **D70** (2004) 043541, astro-ph/0401313.
- [67] J. M. Maldacena, “Non-Gaussian features of primordial fluctuations in single field inflationary models,” *JHEP* **05** (2003) 013, astro-ph/0210603.

- [68] D. H. Lyth, K. A. Malik, and M. Sasaki, “A general proof of the conservation of the curvature perturbation,” *JCAP* **0505** (2005) 004, astro-ph/0411220.
- [69] S. Weinberg, “Non-Gaussian Correlations Outside the Horizon,” *Phys. Rev.* **D78** (2008) 123521, 0808.2909.
- [70] S. Weinberg, “Non-Gaussian Correlations Outside the Horizon II: The General Case,” *Phys. Rev.* **D79** (2009) 043504, 0810.2831.
- [71] S. Seager, D. D. Sasselov, and D. Scott, “A New Calculation of the Recombination Epoch,” astro-ph/9909275.
- [72] S. Seager, D. D. Sasselov, and D. Scott, “How exactly did the Universe become neutral?,” *Astrophys. J. Suppl.* **128** (2000) 407–430, astro-ph/9912182.
- [73] S. Weinberg, “The Quantum theory of fields. Vol. 1: Foundations,” Cambridge, UK: Univ. Pr. (1995) 609 p.
- [74] U. Seljak and M. Zaldarriaga, “A Line of Sight Approach to Cosmic Microwave Background Anisotropies,” *Astrophys. J.* **469** (1996) 437–444, astro-ph/9603033.
- [75] S. Weinberg, “A no-truncation approach to cosmic microwave background anisotropies,” *Phys. Rev.* **D74** (2006) 063517, astro-ph/0607076.
- [76] D. Baskaran, L. P. Grishchuk, and A. G. Polnarev, “Imprints of relic gravitational waves in cosmic microwave background radiation,” *Phys. Rev.* **D74** (2006) 083008, gr-qc/0605100.

- [77] **Particle Data Group** Collaboration, W. M. Yao *et al.*, “Review of particle physics,” *J. Phys.* **G33** (2006) 1–1232.
- [78] **WMAP** Collaboration, E. Komatsu *et al.*, “Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation,” *Astrophys. J. Suppl.* **180** (2009) 330–376, 0803.0547.
- [79] J. R. Bond and A. S. Szalay, “The Collisionless Damping of Density Fluctuations in an Expanding Universe,” *Astrophys. J.* **274** (1983) 443–468.
- [80] C.-P. Ma and E. Bertschinger, “Cosmological perturbation theory in the synchronous and conformal Newtonian gauges,” *Astrophys. J.* **455** (1995) 7–25, astro-ph/9506072.
- [81] S. Dodelson, E. Gates, and A. Stebbins, “Cold + Hot Dark Matter and the Cosmic Microwave Background,” *Astrophys. J.* **467** (1996) 10–18, astro-ph/9509147.
- [82] U. Seljak and M. Zaldarriaga, “Signature of gravity waves in polarization of the microwave background,” *Phys. Rev. Lett.* **78** (1997) 2054–2057, astro-ph/9609169.
- [83] A. Lewis, A. Challinor, and A. Lasenby, “Efficient Computation of CMB anisotropies in closed FRW models,” *Astrophys. J.* **538** (2000) 473–476, astro-ph/9911177.
- [84] S. Weinberg, “Damping of tensor modes in cosmology,” *Phys. Rev.* **D69** (2004) 023503, astro-ph/0306304.

- [85] V. F. Mukhanov, “CMB-slow, or How to Estimate Cosmological Parameters by Hand,” *Int. J. Theor. Phys.* **43** (2004) 623–668, astro-ph/0303072.
- [86] **WMAP** Collaboration, M. R.olta *et al.*, “Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Angular Power Spectra,” *Astrophys. J. Suppl.* **180** (2009) 296–305, 0803.0593.
- [87] C. L. Reichardt *et al.*, “High resolution CMB power spectrum from the complete ACBAR data set,” *Astrophys. J.* **694** (2009) 1200–1219, 0801.1491.
- [88] W. C. Jones *et al.*, “A Measurement of the Angular Power Spectrum of the CMB Temperature Anisotropy from the 2003 Flight of Boomerang,” *Astrophys. J.* **647** (2006) 823, astro-ph/0507494.
- [89] A. C. S. Readhead *et al.*, “Extended Mosaic Observations with the Cosmic Background Imager,” *Astrophys. J.* **609** (2004) 498–512, astro-ph/0402359.
- [90] V. F. Mukhanov and G. V. Chibisov, “Quantum fluctuations and a nonsingular universe,” *Soviet Journal of Experimental and Theoretical Physics Letters* **33** (May, 1981) 532.
- [91] V. F. Mukhanov and G. V. Chibisov, “Energy of vacuum and the large-scale structure of the universe,” *Zhurnal Eksperimental noi i Teoreticheskoi Fiziki* **83** (Aug., 1982) 475–487.
- [92] S. W. Hawking, “The Development of Irregularities in a Single Bubble Inflationary Universe,” *Phys. Lett.* **B115** (1982) 295.

- [93] A. H. Guth and S.-Y. Pi, “Fluctuations in the new inflationary universe,” *Physical Review Letters* **49** (Oct., 1982) 1110–1113.
- [94] A. A. Starobinsky, “Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations,” *Physics Letters B* **117** (Nov., 1982) 175–178.
- [95] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, “Spontaneous creation of almost scale-free density perturbations in an inflationary universe,” *Phys. Rev.* (Aug., 1983) 679–693.
- [96] W. Fischler, B. Ratra, and L. Susskind, “Quantum mechanics of inflation,” *Nuclear Physics B* **259** (Sept., 1985) 730–744.
- [97] A. H. Guth, “The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems,” *Phys. Rev.* **D23** (1981) 347–356.
- [98] A. D. Linde, “A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems,” *Phys. Lett.* **B108** (1982) 389–393.
- [99] A. J. Albrecht and P. J. Steinhardt, “Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking,” *Phys. Rev. Lett.* **48** (1982) 1220–1223.
- [100] A. A. Starobinsky, “Spectrum of relict gravitational radiation and the early state of the universe,” *JETP Lett.* **30** (1979) 682–685.
- [101] D. S. Goldwirth and T. Piran, “Spherical Inhomogeneous Cosmologies and Inflation. 1. Numerical Methods,” *Phys. Rev.* **D40** (1989) 3263.

- [102] D. S. Goldwirth and T. Piran, “Inhomogeneity and the Onset of Inflation,” *Phys. Rev. Lett.* **64** (1990) 2852–2855.
- [103] D. S. Goldwirth, “On inhomogeneous initial conditions for inflation,” *Phys. Rev.* **D43** (1991) 3204–3213.
- [104] D. S. Goldwirth and T. Piran, “Initial conditions for inflation,” *Phys. Rept.* **214** (1992) 223–291.
- [105] E. Calzetta and M. Sakellariadou, “Inflation in inhomogeneous cosmology,” *Phys. Rev.* **D45** (1992) 2802–2805.
- [106] N. Deruelle and D. S. Goldwirth, “Conditions for inflation in an initially inhomogeneous universe,” *Phys. Rev.* **D51** (1995) 1563–1568, [gr-qc/9409056](#).
- [107] H. Kurki-Suonio, R. A. Matzner, J. Centrella, and J. R. Wilson, “Inflation from Inhomogeneous Initial Data in a One- Dimensional Back Reacting Cosmology,” *Phys. Rev.* **D35** (1987) 435–448.
- [108] P. Laguna, H. Kurki-Suonio, and R. A. Matzner, “Inhomogeneous inflation: The Initial value problem,” *Phys. Rev.* **D44** (1991) 3077–3086.
- [109] H. Kurki-Suonio, P. Laguna, and R. A. Matzner, “Inhomogeneous inflation: Numerical evolution,” *Phys. Rev.* **D48** (1993) 3611–3624, [astro-ph/9306009](#).
- [110] A. Linde, “Inflationary Cosmology,” *Lect. Notes Phys.* **738** (2008) 1–54, [0705.0164](#).

- [111] D. H. Lyth and A. Riotto, “Particle physics models of inflation and the cosmological density perturbation,” *Phys. Rept.* **314** (1999) 1–146, hep-ph/9807278.
- [112] B. A. Bassett, S. Tsujikawa, and D. Wands, “Inflation dynamics and reheating,” *Rev. Mod. Phys.* **78** (2006) 537–589, astro-ph/0507632.
- [113] V. F. Mukhanov, “Gravitational Instability of the Universe Filled with a Scalar Field,” *JETP Lett.* **41** (1985) 493–496.
- [114] M. Sasaki, “Large Scale Quantum Fluctuations in the Inflationary Universe,” *Progress of Theoretical Physics* **76** (Nov., 1986) 1036–1046.
- [115] E. D. Stewart and D. H. Lyth, “A more accurate analytic calculation of the spectrum of cosmological perturbations produced during inflation,” *Phys. Lett.* **B302** (1993) 171–175, gr-qc/9302019.
- [116] A. R. Liddle and M. S. Turner, “Second order reconstruction of the inflationary potential,” *Phys. Rev.* **D50** (1994) 758–768, astro-ph/9402021.
- [117] **CMBPol Study Team** Collaboration, D. Baumann *et al.*, “CMBPol Mission Concept Study: Probing Inflation with CMB Polarization,” 0811.3919.
- [118] D. H. Lyth, “What would we learn by detecting a gravitational wave signal in the cosmic microwave background anisotropy?,” *Phys. Rev. Lett.* **78** (1997) 1861–1863, hep-ph/9606387.
- [119] G. Efstathiou and K. J. Mack, “The Lyth Bound Revisited,” *JCAP* **0505** (2005) 008, astro-ph/0503360.

- [120] R. Easther, W. H. Kinney, and B. A. Powell, “The Lyth bound and the end of inflation,” *JCAP* **0608** (2006) 004, [astro-ph/0601276](#).
- [121] H. C. Chiang *et al.*, “Measurement of CMB Polarization Power Spectra from Two Years of BICEP Data,” [0906.1181](#).
- [122] R. Flauger and S. Weinberg, “Tensor Microwave Background Fluctuations for Large Multipole Order,” *Phys. Rev.* **D75** (2007) 123505, [astro-ph/0703179](#).
- [123] M. Kamionkowski, A. Kosowsky, and A. Stebbins, “A probe of primordial gravity waves and vorticity,” *Phys. Rev. Lett.* **78** (1997) 2058–2061, [astro-ph/9609132](#).
- [124] R. Crittenden, J. R. Bond, R. L. Davis, G. Efstathiou, and P. J. Steinhardt, “The Imprint of gravitational waves on the cosmic microwave background,” *Phys. Rev. Lett.* **71** (1993) 324–327, [astro-ph/9303014](#).
- [125] I. S. Gradshteyn and I. M. Ryzhik, “Table of integrals, series, and products,”. New York, Academic Press, (1980).
- [126] J. R. Pritchard and M. Kamionkowski, “Cosmic microwave background fluctuations from gravitational waves: An analytic approach,” *Annals Phys.* **318** (2005) 2–36, [astro-ph/0412581](#).
- [127] A. A. Starobinsky, “Cosmic Background Anisotropy Induced by Isotropic Flat- Spectrum Gravitational-Wave Perturbations,” *Sov. Astron. Lett.* **11** (1985) 133.

- [128] M. Fierz and W. Pauli, “On relativistic wave equations for particles of arbitrary spin in an electromagnetic field,” *Proc. Roy. Soc. Lond.* **A173** (1939) 211–232.
- [129] H. van Dam and M. J. G. Veltman, “Massive and massless Yang-Mills and gravitational fields,” *Nucl. Phys.* **B22** (1970) 397–411.
- [130] V. I. Zakharov, “Linearized gravitation theory and the graviton mass,” *JETP Lett.* **12** (1970) 312.
- [131] A. I. Vainshtein, “To the problem of nonvanishing gravitation mass,” *Phys. Lett.* **B39** (1972) 393–394.
- [132] D. G. Boulware and S. Deser, “Can gravitation have a finite range?,” *Phys. Rev.* **D6** (1972) 3368–3382.
- [133] N. Arkani-Hamed, H. Georgi, and M. D. Schwartz, “Effective field theory for massive gravitons and gravity in theory space,” *Ann. Phys.* **305** (2003) 96–118, [hep-th/0210184](#).
- [134] V. A. Rubakov and P. G. Tinyakov, “Infrared-modified gravities and massive gravitons,” *Phys. Usp.* **51** (2008) 759–792, [0802.4379](#).
- [135] N. Arkani-Hamed, H.-C. Cheng, M. A. Luty, and S. Mukohyama, “Ghost condensation and a consistent infrared modification of gravity,” *JHEP* **05** (2004) 074, [hep-th/0312099](#).
- [136] V. A. Rubakov, “Lorentz-violating graviton masses: Getting around ghosts, low strong coupling scale and vDVZ discontinuity,” [hep-th/0407104](#).

- [137] S. L. Dubovsky, “Phases of massive gravity,” *JHEP* **10** (2004) 076, hep-th/0409124.
- [138] S. L. Dubovsky, P. G. Tinyakov, and I. I. Tkachev, “Massive graviton as a testable cold dark matter candidate,” *Phys. Rev. Lett.* **94** (2005) 181102, hep-th/0411158.
- [139] S. L. Dubovsky, P. G. Tinyakov, and I. I. Tkachev, “Cosmological attractors in massive gravity,” *Phys. Rev.* **D72** (2005) 084011, hep-th/0504067.
- [140] M. V. Bebronne and P. G. Tinyakov, “Massive gravity and structure formation,” *Phys. Rev.* **D76** (2007) 084011, 0705.1301.
- [141] S. Dubovsky, P. Tinyakov, and M. Zaldarriaga, “Bumpy black holes from spontaneous Lorentz violation,” *JHEP* **11** (2007) 083, 0706.0288.
- [142] J. H. Taylor, “Binary pulsars and relativistic gravity,” *Rev. Mod. Phys.* **66** (1994) 711–719.
- [143] S. Dubovsky, R. Flauger, A. Starobinsky, and I. Tkachev, “Signatures of a Graviton Mass in the Cosmic Microwave Background,” 0907.1658.
- [144] A. V. Zakharov, “Effect of collisionless particles on the growth of gravitational perturbations in an isotropic universe,” *Zhurnal Eksperimental noi i Teoreticheskoi Fiziki* **77** (Aug., 1979) 434–450.
- [145] A. A. Starobinskii, “Cosmic Background Anisotropy Induced by Isotropic Flat-Spectrum Gravitational-Wave Perturbations,” *Soviet Astronomy Letters* **11** (May, 1985) 133.

- [146] B. M. Project, “Higher Transcendental Functions, Volume 2,”. New York, McGraw Hill, (1953-55).
- [147] D. Polarski and A. A. Starobinsky, “Semiclassicality and decoherence of cosmological perturbations,” *Class. Quant. Grav.* **13** (1996) 377–392, [gr-qc/9504030](#).
- [148] A. Lewis and S. Bridle, “Cosmological parameters from CMB and other data: a Monte- Carlo approach,” *Phys. Rev.* **D66** (2002) 103511, [astro-ph/0205436](#).
- [149] M. Zaldarriaga and U. Seljak, “Gravitational Lensing Effect on Cosmic Microwave Background Polarization,” *Phys. Rev.* **D58** (1998) 023003, [astro-ph/9803150](#).
- [150] E. Silverstein and A. Westphal, “Monodromy in the CMB: Gravity Waves and String Inflation,” *Phys. Rev.* **D78** (2008) 106003, 0803.3085.
- [151] L. McAllister, E. Silverstein, and A. Westphal, “Gravity Waves and Linear Inflation from Axion Monodromy,” 0808.0706.
- [152] X. Chen, R. Easther, and E. A. Lim, “Generation and Characterization of Large Non-Gaussianities in Single Field Inflation,” *JCAP* **0804** (2008) 010, 0801.3295.
- [153] J. Martin and C. Ringeval, “Superimposed Oscillations in the WMAP Data?,” *Phys. Rev.* **D69** (2004) 083515, [astro-ph/0310382](#).
- [154] J. Martin and C. Ringeval, “Exploring the superimposed oscillations parameter space,” *JCAP* **0501** (2005) 007, [hep-ph/0405249](#).

- [155] R. Easther, W. H. Kinney, and H. Peiris, “Observing trans-Planckian signatures in the cosmic microwave background,” *JCAP* **0505** (2005) 009, [astro-ph/0412613](#).
- [156] X. Chen, R. Easther, and E. A. Lim, “Large non-Gaussianities in single field inflation,” *JCAP* **0706** (2007) 023, [astro-ph/0611645](#).
- [157] J. Hamann, L. Covi, A. Melchiorri, and A. Slosar, “New constraints on oscillations in the primordial spectrum of inflationary perturbations,” *Phys. Rev.* **D76** (2007) 023503, [astro-ph/0701380](#).
- [158] R. Bean, X. Chen, G. Hailu, S. H. H. Tye, and J. Xu, “Duality Cascade in Brane Inflation,” *JCAP* **0803** (2008) 026, [0802.0491](#).
- [159] C. Pahud, M. Kamionkowski, and A. R. Liddle, “Oscillations in the inflaton potential?,” *Phys. Rev.* **D79** (2009) 083503, [0807.0322](#).
- [160] R. Flauger, L. McAllister, E. Pajer, A. Westphal, and G. Xu, “Oscillations in the CMB from Axion Monodromy Inflation,” [0907.2916](#).
- [161] D. Baumann and L. McAllister, “Advances in Inflation in String Theory,” [0901.0265](#).
- [162] K. Freese and D. Spolyar, “Chain inflation: ‘Bubble bubble toil and trouble’,” *JCAP* **0507** (2005) 007, [hep-ph/0412145](#).
- [163] K. Freese, J. T. Liu, and D. Spolyar, “Chain inflation via rapid tunneling in the landscape,” [hep-th/0612056](#).
- [164] D. Chialva and U. H. Danielsson, “Chain inflation revisited,” *JCAP* **0810** (2008) 012, [0804.2846](#).

- [165] J. Lesgourgues, A. A. Starobinsky, and W. Valkenburg, “What do WMAP and SDSS really tell about inflation?,” *JCAP* **0801** (2008) 010, 0710.1630.
- [166] R. Easther, B. R. Greene, W. H. Kinney, and G. Shiu, “Inflation as a probe of short distance physics,” *Phys. Rev.* **D64** (2001) 103502, hep-th/0104102.
- [167] N. Kaloper, M. Kleban, A. E. Lawrence, and S. Shenker, “Signatures of short distance physics in the cosmic microwave background,” *Phys. Rev.* **D66** (2002) 123510, hep-th/0201158.
- [168] P. D. Meerburg, J. P. van der Schaar, and P. S. Corasaniti, “Signatures of Initial State Modifications on Bispectrum Statistics,” *JCAP* **0905** (2009) 018, 0901.4044.
- [169] J. Polchinski, “String theory. Vol. 2: Superstring theory and beyond,” Cambridge, UK: Univ. Pr. (1998) 531 p.
- [170] S. Kachru, R. Kallosh, A. Linde, and S. P. Trivedi, “De Sitter vacua in string theory,” *Phys. Rev.* **D68** (2003) 046005, hep-th/0301240.
- [171] T. W. Grimm and J. Louis, “The effective action of $N = 1$ Calabi-Yau orientifolds,” *Nucl. Phys.* **B699** (2004) 387–426, hep-th/0403067.
- [172] X. G. Wen and E. Witten, “WORLD SHEET INSTANTONS AND THE PECCEI-QUINN SYMMETRY,” *Phys. Lett.* **B166** (1986) 397.
- [173] M. Dine and N. Seiberg, “Nonrenormalization Theorems in Superstring Theory,” *Phys. Rev. Lett.* **57** (1986) 2625.

- [174] E. Pajer, “Inflation at the Tip,” *JCAP* **0804** (2008) 031, 0802.2916.
- [175] S. Kachru *et al.*, “Towards inflation in string theory,” *JCAP* **0310** (2003) 013, hep-th/0308055.
- [176] P. Svrcek and E. Witten, “Axions in string theory,” *JHEP* **06** (2006) 051, hep-th/0605206.
- [177] M. Bertolini *et al.*, “Supersymmetric 3-branes on smooth ALE manifolds with flux,” *Nucl. Phys.* **B617** (2001) 3–42, hep-th/0106186.
- [178] D. Baumann, A. Dymarsky, I. R. Klebanov, J. M. Maldacena, L. P. McAllister, and A. Murugan, “On D3-brane potentials in compactifications with fluxes and wrapped D-branes,” *JHEP* **11** (2006) 031, hep-th/0607050.
- [179] D. Baumann, A. Dymarsky, I. R. Klebanov, and L. McAllister, “Towards an Explicit Model of D-brane Inflation,” *JCAP* **0801** (2008) 024, 0706.0360.
- [180] O. DeWolfe, S. Kachru, and M. Mulligan, “A Gravity Dual of Metastable Dynamical Supersymmetry Breaking,” *Phys. Rev.* **D77** (2008) 065011, 0801.1520.
- [181] D. J. Gross and J. H. Sloan, “The Quartic Effective Action for the Heterotic String,” *Nucl. Phys.* **B291** (1987) 41.
- [182] A. Kehagias and H. Partouche, “On the exact quartic effective action for the type iib superstring,” *Phys. Lett.* **B422** (1998) 109–116, hep-th/9710023.

- [183] T. Banks, M. Dine, P. J. Fox, and E. Gorbatov, “On the possibility of large axion decay constants,” *JCAP* **0306** (2003) 001, hep-th/0303252.
- [184] E. Witten, “Non-Perturbative Superpotentials In String Theory,” *Nucl. Phys.* **B474** (1996) 343–360, hep-th/9604030.
- [185] T. W. Grimm, “Non-Perturbative Corrections and Modularity in N=1 Type IIB Compactifications,” *JHEP* **10** (2007) 004, 0705.3253.
- [186] D. Robles-Llana, M. Rocek, F. Saueressig, U. Theis, and S. Vandoren, “Nonperturbative corrections to 4D string theory effective actions from $SL(2,Z)$ duality and supersymmetry,” *Phys. Rev. Lett.* **98** (2007) 211602, hep-th/0612027.
- [187] V. Balasubramanian, P. Berglund, J. P. Conlon, and F. Quevedo, “Systematics of Moduli Stabilisation in Calabi-Yau Flux Compactifications,” *JHEP* **03** (2005) 007, hep-th/0502058.
- [188] L. McAllister and E. Silverstein, “String Cosmology: A Review,” *Gen. Rel. Grav.* **40** (2008) 565–605, 0710.2951.
- [189] F. Quevedo, “Lectures on string/brane cosmology,” *Class. Quant. Grav.* **19** (2002) 5721–5779, hep-th/0210292.
- [190] A. Linde, “Inflation and string cosmology,” *ECONF* **C040802** (2004) L024, hep-th/0503195.
- [191] S. H. Henry Tye, “Brane inflation: String theory viewed from the cosmos,” *Lect. Notes Phys.* **737** (2008) 949–974, hep-th/0610221.

- [192] J. M. Cline, “String cosmology,” [hep-th/0612129](#).
- [193] C. P. Burgess, “Lectures on Cosmic Inflation and its Potential Stringy Realizations,” *PoS P2GC* (2006) 008, [0708.2865](#).
- [194] R. Kallosh, “On Inflation in String Theory,” *Lect. Notes Phys.* **738** (2008) 119–156, [hep-th/0702059](#).
- [195] O. DeWolfe, A. Giryavets, S. Kachru, and W. Taylor, “Type IIA moduli stabilization,” *JHEP* **07** (2005) 066, [hep-th/0505160](#).
- [196] T. Banks and K. van den Broek, “Massive IIA flux compactifications and U-dualities,” *JHEP* **03** (2007) 068, [hep-th/0611185](#).
- [197] M. P. Hertzberg, M. Tegmark, S. Kachru, J. Shelton, and O. Ozcan, “Searching for Inflation in Simple String Theory Models: An Astrophysical Perspective,” *Phys. Rev.* **D76** (2007) 103521, [0709.0002](#).
- [198] M. P. Hertzberg, S. Kachru, W. Taylor, and M. Tegmark, “Inflationary Constraints on Type IIA String Theory,” *JHEP* **12** (2007) 095, [0711.2512](#).
- [199] G. Villadoro and F. Zwirner, “N = 1 effective potential from dual type-IIA D6/O6 orientifolds with general fluxes,” *JHEP* **06** (2005) 047, [hep-th/0503169](#).
- [200] M. Ihl and T. Wrase, “Towards a realistic type IIA T**6/Z(4) orientifold model with background fluxes. I: Moduli stabilization,” *JHEP* **07** (2006) 027, [hep-th/0604087](#).

- [201] E. Silverstein, “Simple de Sitter Solutions,” *Phys. Rev.* **D77** (2008) 106006, 0712.1196.
- [202] S. S. Haque, G. Shiu, B. Underwood, and T. Van Riet, “Minimal simple de Sitter solutions,” *Phys. Rev.* **D79** (2009) 086005, 0810.5328.
- [203] E. Palti, G. Tasinato, and J. Ward, “WEAKLY-coupled IIA Flux Compactifications,” *JHEP* **06** (2008) 084, 0804.1248.
- [204] C. Caviezel *et al.*, “On the Cosmology of Type IIA Compactifications on SU(3)- structure Manifolds,” *JHEP* **04** (2009) 010, 0812.3551.
- [205] R. Flauger, S. Paban, D. Robbins, and T. Wrase, “Searching for slow-roll moduli inflation in massive type IIA supergravity with metric fluxes,” *Phys. Rev.* **D79** (2009) 086011, 0812.3886.
- [206] J. Scherk and J. H. Schwarz, “How to Get Masses from Extra Dimensions,” *Nucl. Phys.* **B153** (1979) 61–88.
- [207] N. Kaloper, R. R. Khuri, and R. C. Myers, “On generalized axion reductions,” *Phys. Lett.* **B428** (1998) 297–302, hep-th/9803066.
- [208] N. Kaloper and R. C. Myers, “The O(dd) story of massive supergravity,” *JHEP* **05** (1999) 010, hep-th/9901045.
- [209] M. Ihl, D. Robbins, and T. Wrase, “Toroidal Orientifolds in IIA with General NS-NS Fluxes,” *JHEP* **08** (2007) 043, 0705.3410.
- [210] D. Robbins and T. Wrase, “D-Terms from Generalized NS-NS Fluxes in Type II,” *JHEP* **12** (2007) 058, 0709.2186.

- [211] S. Hellerman, J. McGreevy, and B. Williams, “Geometric Constructions of Nongeometric String Theories,” *JHEP* **01** (2004) 024, hep-th/0208174.
- [212] A. Dabholkar and C. Hull, “Duality twists, orbifolds, and fluxes,” *JHEP* **09** (2003) 054, hep-th/0210209.
- [213] J. Shelton, W. Taylor, and B. Wecht, “Nongeometric Flux Compactifications,” *JHEP* **10** (2005) 085, hep-th/0508133.
- [214] J. Shelton, W. Taylor, and B. Wecht, “Generalized flux vacua,” *JHEP* **02** (2007) 095, hep-th/0607015.
- [215] G. Aldazabal and A. Font, “A second look at N=1 supersymmetric AdS_4 vacua of type IIA supergravity,” *JHEP* **02** (2008) 086, 0712.1021.
- [216] B. Wecht, “Lectures on Nongeometric Flux Compactifications,” *Class. Quant. Grav.* **24** (2007) S773–S794, 0708.3984.
- [217] T. W. Grimm and J. Louis, “The effective action of type IIA Calabi-Yau orientifolds,” *Nucl. Phys.* **B718** (2005) 153–202, hep-th/0412277.
- [218] M. Grana, R. Minasian, M. Petrini, and A. Tomasiello, “A scan for new N=1 vacua on twisted tori,” *JHEP* **05** (2007) 031, hep-th/0609124.
- [219] J. Erler and A. Klemm, “Comment on the generation number in orbifold compactifications,” *Commun. Math. Phys.* **153** (1993) 579–604, hep-th/9207111.

- [220] S. Reffert, “Toroidal orbifolds: Resolutions, orientifolds and applications in string phenomenology,” `hep-th/0609040`.
- [221] S. Reffert, “The Geometer’s Toolkit to String Compactifications,” `0706.1310`.
- [222] G. Villadoro and F. Zwirner, “On general flux backgrounds with localized sources,” *JHEP* **11** (2007) 082, `0710.2551`.
- [223] A. Tomasiello, “New string vacua from twistor spaces,” *Phys. Rev.* **D78** (2008) 046007, `0712.1396`.
- [224] P. Koerber, D. Lust, and D. Tsimpis, “Type IIA AdS4 compactifications on cosets, interpolations and domain walls,” *JHEP* **07** (2008) 017, `0804.0614`.
- [225] C. Caviezel *et al.*, “The effective theory of type IIA AdS4 compactifications on nilmanifolds and cosets,” *Class. Quant. Grav.* **26** (2009) 025014, `0806.3458`.
- [226] A. Bergman and D. Robbins, “Ramond-Ramond Fields, Cohomology and Non-Geometric Fluxes,” `0710.5158`.
- [227] L. Covi *et al.*, “de Sitter vacua in no-scale supergravities and Calabi-Yau string models,” *JHEP* **06** (2008) 057, `0804.1073`.
- [228] L. Covi *et al.*, “Constraints on modular inflation in supergravity and string theory,” *JHEP* **08** (2008) 055, `0805.3290`.
- [229] P. J. Steinhardt and D. Wesley, “Dark Energy, Inflation and Extra Dimensions,” `0811.1614`.

- [230] I. R. Klebanov and A. Murugan, “Gauge/Gravity Duality and Warped Resolved Conifold,” *JHEP* **03** (2007) 042, [hep-th/0701064](#).
- [231] L. A. Pando Zayas and A. A. Tseytlin, “3-branes on resolved conifold,” *JHEP* **11** (2000) 028, [hep-th/0010088](#).
- [232] S. Kuperstein, “Meson spectroscopy from holomorphic probes on the warped deformed conifold,” *JHEP* **03** (2005) 014, [hep-th/0411097](#).
- [233] P. Ouyang, “Holomorphic D7-branes and flavored $N = 1$ gauge theories,” *Nucl. Phys.* **B699** (2004) 207–225, [hep-th/0311084](#).
- [234] P. Candelas and X. C. de la Ossa, “Comments on Conifolds,” *Nucl. Phys.* **B342** (1990) 246–268.
- [235] N. Barnaby, C. P. Burgess, and J. M. Cline, “Warped reheating in brane-antibrane inflation,” *JCAP* **0504** (2005) 007, [hep-th/0412040](#).
- [236] L. Kofman and P. Yi, “Reheating the universe after string theory inflation,” *Phys. Rev.* **D72** (2005) 106001, [hep-th/0507257](#).

Vita

Raphael Manfred Flauger was born in Nürnberg, Germany on July 7, 1979, the son of Manfred Flauger and Mechthild Flauger. After receiving the *Abitur* from the Leibniz-Gymnasium in Altdorf/Nbg., Germany, in June 1998, he entered the German Air Force in July 1998 to perform his compulsory military service. In October 1998 he enrolled in a long distance education program in mechanical engineering at the Technische Universität Dresden. After the completion of his military service in April 1999, he enrolled in the physics program at the Ludwig-Maximilians Universität Würzburg, Germany, where he received the *Vordiplom* in August 2000. He continued his education in Würzburg until July 2001 before joining the Weinberg Theory Group at The University of Texas at Austin to receive a Master of Arts in Physics under the supervision of Sonia Paban in August 2002. He continued his studies at Imperial College London where he received a Master of Science in Quantum Fields and Fundamental Forces under the supervision of Arkady Tseytlin in September 2003. Since then he has been a PhD student in the Weinberg Theory Group at the University of Texas under the supervision of Steven Weinberg. From January until August 2008, he was a Graduate Fellow at the KITP in Santa Barbara, CA.

Permanent address: Nibelungenstraße 6
90592 Schwarzenbruck
Germany

This dissertation was typeset with L^AT_EX[†] by Raphael Manfred Flauger.

[†]L^AT_EX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's T_EX Program.