Inflationary Cosmology and Fundamental Physics

by

Mark Peter Hertzberg

B.S. & M.S., University of Sydney (2004)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2010

Author .....
Department of Physics
May 12, 2010

Certified by .....
Frank Wilczek
Herman Feshbach Professor of Physics
Thesis Supervisor

Certified by .....
Max Tegmark
Professor of Physics
Thesis Supervisor

Accepted by .....
Krishna Rajagopal
Associate Department Head for Education

Inflationary Cosmology and Fundamental Physics

by

Mark Peter Hertzberg

Submitted to the Department of Physics
on May 12, 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis is a collection of several papers at the interface between cosmology, particle physics, and field theory. In the first half, we examine topics that are directly related to inflation: axions, string theory, and non-minimally coupled fields. In particular, we constrain the allowed parameter space of inflationary axion cosmology, identifying a classic window and an anthropic window; we discuss inflation in string theory, proving a no-go theorem for a class of string models; and we examine the quantum field theory governing inflation driven by non-minimally coupled fields, which is relevant to Higgs-inflation. In the second half, we examine other topics: oscillons, entanglement entropy, and the Casimir effect. In particular, we compute the quantum decay of oscillons, finding it to be dramatically different to the classical decay; we establish finite contributions to the entanglement entropy between a pair of regions, including a finite area law; and we compute the Casimir force in a closed geometry, finding an attractive force and invalidating claims of repulsion.

Thesis Supervisor: Frank Wilczek
Title: Herman Feshbach Professor of Physics

Thesis Supervisor: Max Tegmark
Title: Professor of Physics

To my parents
Ralph and Teresa Hertzberg

and to the memory of my Nan
Nora Roberts

Acknowledgments

First, to my supervisors, Prof. Max Tegmark and Prof. Frank Wilczek, I would like to thank them for all their help, support, and enthusiasm over the course of my PhD. Also to my academic advisor, Prof. Barton Zwiebach, and my thesis readers, Prof. Alan Guth and Prof. Roman Jackiw, for their advice. I would also like to thank my other collaborators, Andrea De Simone, Prof. Robert Jaffe, Prof. Shamit Kachru, Prof. Mehran Kardar, Onur Ozcan, Antonello Scardicchio, Jessie Shelton, and Prof. Washington Taylor, for their contributions to the papers in this thesis. Also to the administrative staff, Joyce Berggren, Scott Morley, Jean Papagianopoulos, and Charles Suggs, for their help. I would also like to thank all my friends, both at MIT and beyond, for their kindness and encouragement – there are too many of you to name here. Last, but not least, I would like to thank my parents, Ralph and Teresa, my brothers, Andrew and David, and Leah Jacobs, for all their love and support throughout this time.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cosmology and particle physics are closer today than ever before. This thesis is a collection of several papers at the interface between cosmology, particle physics, and field theory. In this opening chapter, we briefly introduce inflationary cosmology and some directly related topics: axions, string theory, and non-minimally coupled fields. We then move on to other topics: oscillons (which may also have cosmological implications), entanglement entropy, and the Casimir effect.

1.1 Inflationary Cosmology

Precision data [1, 2, 3] has revealed that on large scales our universe is approximately homogeneous, isotropic, and flat. Its properties are remarkably well described by the so-called ΛCDM model in which general relativity is assumed and $\sim 74\%$ of the universe is dark energy (Λ), $\sim 22\%$ is cold dark matter (CDM), $\sim 4\%$ is baryonic, plus a small amount of cosmic microwave background radiation (CMB). However, the universe is not perfectly homogeneous or isotropic. Instead there exists fluctuations in the matter density distribution and CMB at the $\sim 10^{-5}$ level on large scales. These fluctuations have been observed to be roughly scale-invariant, Gaussian, and of the adiabatic type - though small deviations from this are allowed by the current data.

This leads to several important questions, which may have an answer in particle physics, such as:

1. What caused the universe to be nearly homogeneous, isotropic, and flat?
2. What is the origin and nature of the $\sim 10^{-5}$ density fluctuations?
3. What is the nature of dark matter?
4. What is the origin of all matter and the baryon asymmetry?

The first two of these questions have a beautiful framework in which they may be addressed: inflation [4, 5, 6, 7]. This is the idea that a small patch of space in the early universe underwent a period of exponential growth - an approximate de Sitter phase. At the classical level this causes small departures from homogeneity and

isotropy to redshift away and drives the curvature towards zero. This also solves the horizon problem - it allows the early universe to be causally connected (as opposed to standard big bang cosmology) which is consistent with an approximately uniform CMB temperature across the sky. At the quantum level de Sitter fluctuations in light fields (including the metric) are established. These become density fluctuations in the late time universe.

The amplitude of such fluctuations, such as tensor modes or scalar modes (adiabatic or isocurvature) are highly model dependent, but are subject to improving observational constraints. For example, the amplitude of the tensor modes is proportional to the energy density of the inflationary model; so a direct detection of primordial gravity waves from inflation would probe the scale of inflation. On the other hand, a non-detection places an upper bound on the energy density of inflation. The (measured) amplitude of density fluctuations constrains a combination of the energy density of inflation and its rate of change during inflation.

Axions: Other light scalar fields can give rise to other forms of fluctuations. A particularly important example in particle physics is the QCD-axion [8, 9]; the leading candidate to solve the so-called strong CP-problem of the Standard Model. If the QCD-axion is present during inflation, then it establishes isocurvature fluctuations whose amplitude depends both on the scale of inflation and on the so-called Peccei-Quinn [10] scale that controls the axion's mass. Constraints on gravity waves and isocurvature modes therefore together constrain both the energy scale of inflation and the Peccei-Quinn scale of the axion. The constraints also depend substantially on the fraction of today's universe that is in the form of axions. Since axions are electrically neutral, stable, and very weakly interacting, they are a form of dark matter. Hence, apart from the connections to questions 1 and 2, this topic is also intimately connected to question 3 regarding the nature of dark matter. These constraints will be explored in detail in chapter 2.

The simplest way to implement an inflationary phase is to assume some scalar field ϕ exists which carries significant vacuum energy. This is provided by some potential function $V(\phi)$. Consider the following action (signature $-+++$)

$$S = \int d^4x\sqrt{-g}\left[\frac{1}{16\pi G}\mathcal{R} - \frac{1}{2}(\partial\phi)^2 - V(\phi)\right]. \tag{1.1}$$

where $\mathcal{R}$ is the Ricci scalar and G is Newton's constant. If the potential V includes a region over which it is positive and suitably flat, then it acts like a cosmological constant and can drive a de Sitter phase; this is the basic mechanism behind inflation. This phase should not last forever though. The potential should have a minimum about which ϕ eventually oscillates, which is the end of inflation. The shape of the potential determines the amplitude of fluctuations and its spectrum. In order for inflation to really answer questions 1 or 2, we must identify the microphysical origin of this field ϕ and the potential V. How is it connected to fundamental physics and what determines the details of the action including the gravity, kinetic, and potential sectors? We turn now to consider the possible embedding of inflation in particle physics.

String Theory: The leading candidate for a fundamental theory of quantum gravity is string theory. In order for both string theory and inflation to be correct, one should demonstrate explicitly that string theory offers a correct cosmological description of the universe and incorporates inflation. On the face of it this task seems relatively simple, since typical low energy string models include many scalar fields, which carry potential energy functions – the so-called "landscape" of string theory. But are these potential functions of the right shape to support inflation? We provide a pedagogical discussion of inflation in the context of string models in chapters 3 and 4. Further, we carry out a numerical search for inflation in a few simple string models in chapter 3, finding no inflating models. We then establish a no-go theorem for a class of models in chapter 4. This leads us to point out directions to make progress in the direction of connecting string theory and cosmological inflation in the future.

Non-Minimal Coupling: An entirely different approach to searching for inflation in string theory ("top-down approach") is to search for inflation models connected to low energy physics, such as the Standard Model of particle physics ("bottom-up approach"). In such an approach we would rely on the Higgs-boson to drive inflation. Since the Standard Model Higgs-boson has the potential function $V(\phi) \sim \lambda(\phi^2 - v^2)^2$, with $\lambda \sim 0.1$, it turns out that a phase of inflation can be realized, but the amplitude of density fluctuations far exceeds the required 10^{-5} level. One can improve the situation by including a non-minimal coupling between ϕ and gravity, i.e., by including $\xi\phi^2\mathcal{R}$ to the action (1.1). By taking $\xi \sim 10^4$ one can achieve the correct level of density fluctuations [11, 12]. However, with such a large coupling, one may wonder whether or not the quantum theory makes sense as an effective field theory. In chapter 5 we address this issue in detail, as well as models in which an $\mathcal{R}^2$ term is added to the action.

1.2 Other Topics

We now introduce some other topics. The first of which may well have some connections to cosmology that we briefly mention. The latter pair of topics do not directly apply to cosmology, though some cosmological connections do exist in the literature.

Oscillons: Inflationary models, axion models, Higgs models, etc, involve scalar fields evolving in some potential function that often supports coherently oscillating, localized clumps, known as oscillons [13, 14, 15, 16]. Studied classically, oscillons are known to be extremely long lived. They also arise from various initial states, such as thermal states, phase transitions, and possibly at the end of inflation. If such heavy clumps emerged in the early universe and then decayed, the decay products would be produced out of thermal equilibrium, a necessary condition for baryogenesis. Furthermore, oscillons may play some role during p/reheating of the universe at the end of inflation. Such possibilities would shed some light on question 4 posed earlier. However, these ideas have yet to be established. In fact a theory of oscillons in the framework of particle physics is yet to be developed. Most notably, little is known about the quantum mechanical stability of oscillons or the effects from coupling to other fields. In chapter 6 we examine these questions. We find oscillons radiate

quantum mechanically through channels that are typically suppressed classically.

Entanglement Entropy: An inherently quantum mechanical property of a system involving multiple degrees of freedom is entanglement. Consider a system that is not in a product state. For example, a pair of spins in the pure state $\frac{1}{\sqrt{2}}(|\uparrow\rangle|\uparrow\rangle + |\downarrow\rangle|\downarrow\rangle)$. The subsystem of the first spin cannot be fully described without reference to the second spin. If we ignore the second spin, by tracing over it, we lose knowledge of the subsystem as captured by some density matrix ρ. One measure of the loss of information is the entanglement entropy $S = -\mathrm{Tr}\,(\rho \ln \rho)$. In a field theory in flat space with $d > 1$, the entanglement entropy between a pair of regions typically scales with the area of the dividing boundary, but with a diverging constant of proportionality set by the cutoff on the field theory [17]. On the other hand, the entropy in $d = 1$ is finite [18, 19]. An interesting issue is whether there exist finite contributions to the entanglement entropy for $d > 1$. In chapter 7 we compute the entanglement entropy for a massive scalar field theory in arbitrary dimensions, and demonstrate that there are indeed finite contributions to the entropy which are cut-off independent and possibly measurable.

Casimir Effect: In 1948 Casimir [20] predicted that there exists an attractive force between a pair of parallel metallic plates in vacuum. This was experimentally verified in 1997 [21]. In quantum field theory this is best understood as a one-loop QED effect associated with the vacuum energy of the EM field defined on a space with conducting boundary conditions (this should not be confused with the vacuum energy of free space, which has no non-gravitational effects). When the divergent energy is differentiated with respect to the distance between the plates a finite force is obtained. Interestingly, the Casimir force is known to be strongly dependent on geometry, so one could imagine a situation in which the force is repulsive. Such claims abound in the literature in the context of closed geometries, such as a metallic cube. However, there are various divergent contributions to the vacuum energy which should be isolated and shown to be independent of distance in order to derive a finite Casimir force. In chapter 8 we demonstrate that this is impossible for certain closed geometries, such as a cube, but is possible for other closed geometries, such as a piston. We compute the Casimir force on the piston and find it to be attractive.

1.3 Organization of Thesis

Chapters 2 – 8 of this thesis are a reproduction of the seven papers [22, 23, 24, 25, 26, 27, 28], respectively. Chapters 2 – 5 explicitly relate to inflation; chapter 2 discusses inflationary axion cosmology, chapters 3 and 4 discuss inflation in string theory, and chapter 5 discusses inflation with non-minimally coupled fields. Chapters 6 – 8 are other field theory topics; chapter 6 discusses oscillons (this may well have cosmological implications, though this is not fully developed here), chapter 7 discusses entanglement entropy, and chapter 8 discusses the Casimir effect.

Let us clarify the contributions to the work in this thesis. In each paper, I performed the detailed calculations, wrote at least the bulk of the manuscript, and was the first author. The size of each collaboration varied from paper to paper: Ref. [22]

(chapter 2) was co-authored with Prof. Max Tegmark and Prof. Frank Wilczek; substantial guidance and input came from both. Ref. [23] (chapter 3) was co-authored with Prof. Max Tegmark, Prof. Shamit Kachru, Jessie Shelton and Onur Ozcan; substantial guidance came from Prof. Max Tegmark, many suggestions from Prof. Shamit Kachru, and initial calculations started by Onur Ozcan. Ref. [24] (chapter 4) was co-authored with Prof. Shamit Kachru, Prof. Washington Taylor, and Prof. Max Tegmark; the main result was a group effort. Ref. [25] (chapter 5) was sole-authored, with substantial input and suggestions from Prof. Frank Wilczek and Andrea De Simone. Ref. [26] (chapter 6) was sole-authored, with advice from Prof. Alan Guth. Ref. [27] (chapter 7) was co-authored with Prof. Frank Wilczek; the results were a joint effort. Ref. [28] (chapter 8) was co-authored with Prof. Robert Jaffe, Prof. Mehran Kardar, and Antonello Scardicchio; the main results were a group effort.

Bibliography

[1] D. N. Spergel et. al. "Wilkinson Microwave Anisotropy Probe (WMAP) Three Year Results: Implications for Cosmology," (2006), [arXiv:astro-ph/0603449v2].

[2] M. Tegmark, "Cosmological Constraints from the SDSS Luminous Red Galaxies," Phys. Rev. D74 123507 (2006) [arXiv:astro-ph/0608632v2].

[3] A.G. Sanchez et al., "Cosmological Parameters from Cosmic Microwave Background Measurements and the Final 2dF Galaxy Redshift Survey Power Spectrum," MNRAS **366**, 189 (2006).

[4] A. Guth, "The inflationary universe: A possible solution to the horizon and flatness problems", Phys. Rev. D, 23, 347 (1981).

[5] A. D. Linde, "A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems," Phys. Lett. B, 108, 389 (1982).

[6] A. Albrecht and P. J. Steinhardt, "Reheating an Inflationary Universe," Phys. Rev. Lett., 48, 1220 (1982).

[7] A. D. Linde, "Chaotic Inflation," Phys. Lett. B, 129, 177 (1983).

[8] S. Weinberg, "A new light boson?," Phys. Rev. Lett. 40, 223 (1978).

[9] F. Wilczek, "Problem of strong P and T invariance in the presence of instantons," Phys. Rev. Lett. 40, 279 (1978).

[10] R. D. Peccei and H. R. Quinn, "CP conservation in the presence of instantons," Phys. Rev. Lett. 38, 1440 (1977).

[11] D. S. Salopek, J. R. Bond and J. M. Bardeen, "Designing Density Fluctuation Spectra in Inflation," Phys. Rev. D **40**, 1753 (1989).

[12] F. L. Bezrukov and M. Shaposhnikov, "The Standard Model Higgs boson as the inflaton," Phys. Lett. B **659**, 703 (2008) [arXiv:0710.3755 [hep-th]].

[13] E. J. Copeland, M. Gleiser and H. R. Muller, "Oscillons: Resonant configurations during bubble collapse," Phys. Rev. D **52** (1995) 1920 [arXiv:hep-ph/9503217].

[14] E. Farhi, N. Graham, V. Khemani, R. Markov and R. Rosales, "An oscillon in the SU(2) gauged Higgs model," Phys. Rev. D **72** (2005) 101701 [arXiv:hep-th/0505273].

[15] E. W. Kolb and I. I. Tkachev, "Nonlinear axion dynamics and the formation of cosmological pseudosolitons", Phys. Rev. D, **49** 10 (1994).

[16] E. Farhi, N. Graham, A. H. Guth, N. Iqbal, R. R. Rosales and N. Stamatopoulos, "Emergence of Oscillons in an Expanding Background," Phys. Rev. D **77** (2008) 085019 [arXiv:0712.3034 [hep-th]].

[17] M. Srednicki, "Entropy and area," Phys. Rev. Lett. **71** (1993) 666 [arXiv:hep-th/9303048].

[18] C. G. . Callan and F. Wilczek, "On geometric entropy," Phys. Lett. B **333** (1994) 55 [arXiv:hep-th/9401072].

[19] C. Holzhey, F. Larsen and F. Wilczek, "Geometric and renormalized entropy in conformal field theory," Nucl. Phys. B **424** (1994) 443 [arXiv:hep-th/9403108].

[20] H. B. G. Casimir, K. Ned. Akad. Wet. Proc. **51**, 793 (1948).

[21] S. K. Lamoreaux, "Demonstration of the Casimir force in the 0.6 to 6 micrometers range," Phys. Rev. Lett. **78**, 5 (1997).

[22] M. P. Hertzberg, M. Tegmark and F. Wilczek, "Axion Cosmology and the Energy Scale of Inflation," Phys. Rev. D **78**, 083507 (2008) [arXiv:0807.1726 [astro-ph]].

[23] 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. D **76**, 103521 (2007) [arXiv:0709.0002 [astro-ph]].

[24] M. P. Hertzberg, S. Kachru, W. Taylor and M. Tegmark, "Inflationary Constraints on Type IIA String Theory," JHEP **0712**, 095 (2007) [arXiv:0711.2512 [hep-th]].

[25] M. P. Hertzberg, "On Inflation with Non-minimal Coupling," Submitted to JHEP (2010) [arXiv:1002.2995 [hep-ph]].

[26] M. P. Hertzberg, "Quantum Radiation of Oscillons," Submitted to Phys. Rev. D (2010) [arXiv:1003.3459 [hep-th]].

[27] M. P. Hertzberg and F. Wilczek "Finite Contributions to Entanglement Entropy," To be Published (2010).

[28] M. P. Hertzberg, R. L. Jaffe, M. Kardar and A. Scardicchio, "Casimir Forces in a Piston Geometry at Zero and Finite Temperatures," Phys. Rev. D **76**, 045016 (2007) [arXiv:0705.0139 [quant-ph]].

Chapter 2

Axion Cosmology and the Energy Scale of Inflation

We survey observational constraints on the parameter space of inflation and axions and map out two allowed windows: the classic window and the inflationary anthropic window. The cosmology of the latter is particularly interesting; inflationary axion cosmology predicts the existence of isocurvature fluctuations in the CMB, with an amplitude that grows with both the energy scale of inflation and the fraction of dark matter in axions. Statistical arguments favor a substantial value for the latter, and so current bounds on isocurvature fluctuations imply tight constraints on inflation. For example, an axion Peccei-Quinn scale of 10^{16} GeV excludes any inflation model with energy scale $> 3.8 \times 10^{14}$ GeV ($r > 2 \times 10^{-9}$) at 95% confidence, and so implies negligible gravitational waves from inflation, but suggests appreciable isocurvature fluctuations.

2.1 Introduction

Early universe inflation and the QCD axion provide explanations for otherwise mysterious features of the universe. Here we argue that assuming both at once leads to very significant constraints on their central parameters, and to highly falsifiable predictions.

2.1.1 Energy Scale of Inflation

Inflation is the leading paradigm for early universe phenomenology [1, 2, 3]. Its mechanism and the values of its central parameters are unknown, however. One central parameter is the energy scale of inflation E_I, defined as the fourth root of the inflationary potential energy density, evaluated when the modes that re-enter the horizon today left the horizon during inflation. E_I is subject to both theoretical and observational constraints, as illustrated in Figs. 2-1 & 2-2.

A multitude of inflation models involving a broad range of energy scales have been discussed in the literature, including chaotic inflation [4], brane inflation [5, 6] and

others [7]. However, very high E_I has been argued to be theoretically problematic, at least for single field slow-roll inflation, because it involves super-Planckian displacements of the inflaton field ϕ [8]. From Ref. [9] the region in which ϕ moves at least two Planck masses is $E \gtrsim 2.4 \times 10^{16}\,\mathrm{GeV}$. The intuition that high E_I is problematic seems borne out in many string theory models; an example is D-brane models [9].

Very low E_I has been argued to be theoretically problematic also. Naive consideration of families of potential energy functions suggests that $E_I \lesssim 2 \times 10^{16}\,\mathrm{GeV}$ ($r \lesssim 0.01$) is non-generic [10]. One of the most striking successes of high E_I potentials is that they can naturally predict $n_s \sim 0.96$, and generic low-energy potentials fail to make this prediction. Low E_I potentials have the slow-roll parameter ϵ exponentially small, so that the observation $n_s = 1 - 6\epsilon + 2\eta = 0.960 \pm 0.013$ [11] implies $\eta = -0.02 \pm 0.0065$, leaving us wondering why η is so small when it could just as well have been of order -1. This problem is not alleviated by anthropic considerations [12]. By using the observed value of density fluctuations, and setting $\epsilon < 10^{-4}$ as the boundary, $E \lesssim 6.7 \times 10^{15}\,\mathrm{GeV}$ defines this problematic low-scale region. These theoretical issues for inflation are indicated by the vertical regions in Fig. 2-1. Although there are inflation models in the literature at energy scales both above and below this naive window, the debate about whether they are generic continues.

With theory in limbo, we turn to observational guidance. High E_I implies a large amplitude for primordial gravitational waves (GWs). $E_I > 3.8 \times 10^{16}\,\mathrm{GeV}$ ($r > 0.22$) is ruled out by WMAP5 plus BAO and SN data [11], as indicated by the orange region of Fig. 2-2. Possible future searches for primordial GWs have rightly been a focus of attention. In this chapter we emphasize the additional information that can be learned from isocurvature fluctuations.

2.1.2 Axion Physics

The QCD Lagrangian accommodates a gauge invariant, Lorentz invariant, renormalizable term $\propto \theta\, \mathbf{E}^a \cdot \mathbf{B}^a$, with $\theta \in [-\pi, \pi]$, that manifestly breaks P and T symmetry. Precision bounds on the electric dipole moment of the neutron constrain $|\theta| \lesssim 10^{-10}$. The striking smallness of this parameter, which the standard model leaves unexplained, defines the strong P and T problem (a.k.a. "CP problem"). After introducing a new asymptotic (or alternatively, classical) Peccei-Quinn (PQ) symmetry [13] which is spontaneously broken, the effective θ becomes a dynamical variable, and relaxes toward extremely small values. The consequent approximate Nambu-Goldstone boson is the axion [14, 15].

The simplest axion models contain only one phenomenologically significant parameter: f_a, the scale at which the PQ symmetry breaks. The zero temperature Lagrangian for the complex field $\phi = \rho\, e^{i\theta}/\sqrt{2}$ is

$$\mathcal{L} = \frac{1}{2} f_a^2 (\partial_\mu \theta)^2 + \frac{1}{2} (\partial_\mu \rho)^2 - 2\Lambda^4 \sin^2(\theta/2) - \lambda(|\phi|^2 - f_a^2/2)^2$$

($\Lambda \approx 78\,\mathrm{MeV}$, ρ is irrelevant at low energies). Accelerator bounds require f_a to be well above the electroweak scale, and stellar astrophysics constraints place considerably

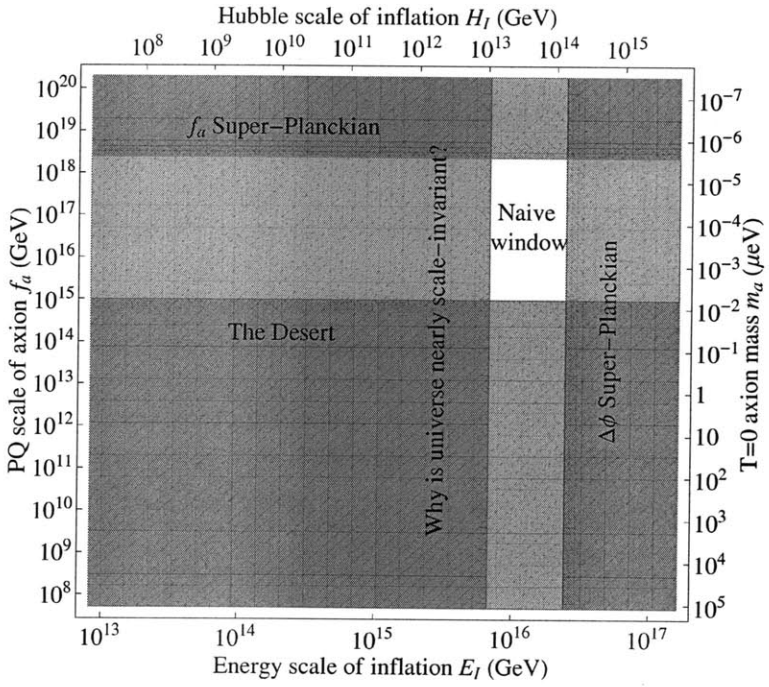

Figure 2-1: *Naive expectations* for the energy scale of inflation E_I and the axion PQ scale f_a. For $E_I \gtrsim 2.4 \times 10^{16}\,\mathrm{GeV}$, the inflaton must undergo super-Planckian excursions in field space (in single field models). For $E_I \lesssim 6.7 \times 10^{15}\,\mathrm{GeV}$, generic inflation potentials fail to reproduce the observed nearly scale invariant power spectrum. For $f_a \gtrsim 2.4 \times 10^{18}\,\mathrm{GeV}$, the PQ breaking scale is super-Planckian. For $f_a \lesssim 10^{15}\,\mathrm{GeV}$ (and $f_a \gg \mathrm{TeV}$), the PQ breaking is in the "desert" of particle physics and non-trivial to achieve in string theory. This leaves the region labeled "naive window".

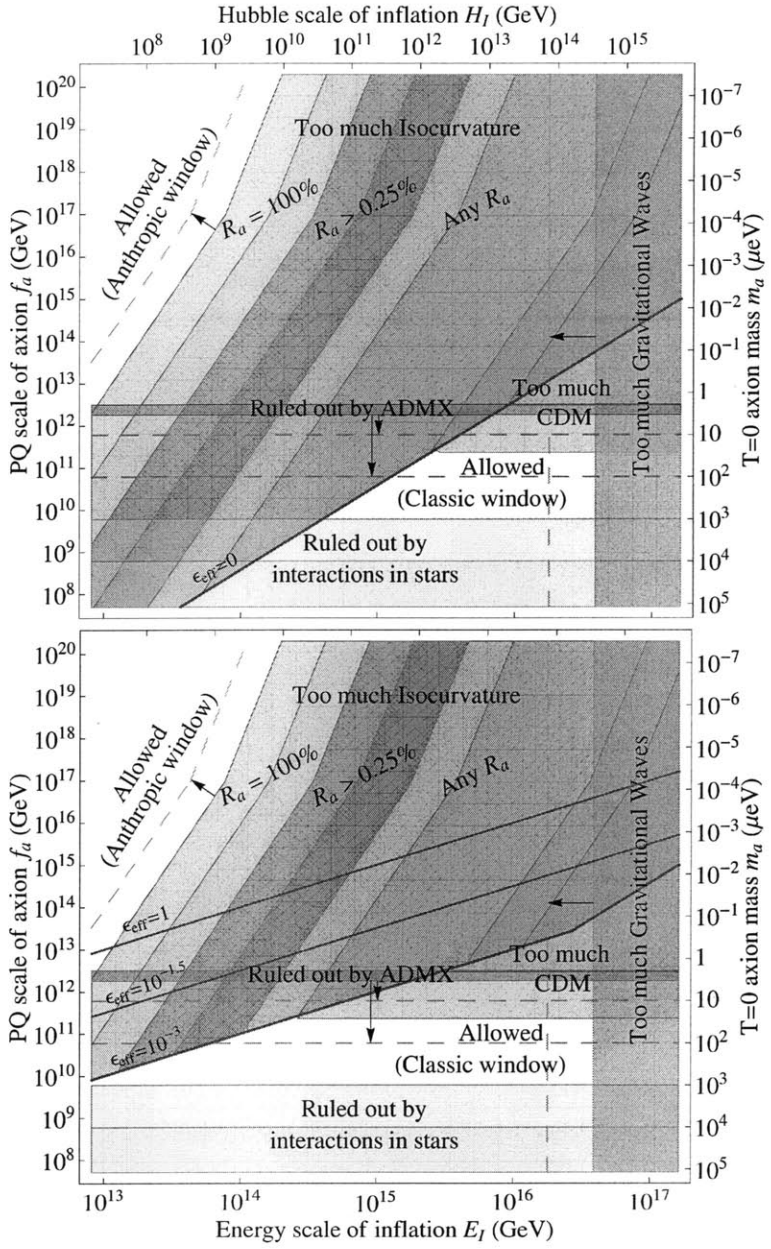

Figure 2-2: *Observational constraints* on the energy scale of inflation E_I and the axion PQ scale f_a are shown in the top (bottom) panel for inefficient (efficient) thermalization at the end of inflation. The thick red diagonal line is $f_a = \mathrm{Max}\{T_{\rm GH} = H_I/2\pi, T_{\rm max} = \epsilon_{\rm eff} E_I\}$, ($\epsilon_{\rm eff} \approx 0$ in top, $\epsilon_{\rm eff} = 10^{-3}$ in bottom, with $\epsilon_{\rm eff} = 10^{-1.5}$, 1 indicated). Above this line is the inflationary anthropic scenario and below this line is the classic scenario. The region in which there is too much isocurvature, $\alpha_a > 0.072$, depends on the axion fraction $R_a \equiv \xi_a/\xi_{\rm CDM}$ of the CDM; the purple region applies for any R_a, the blue region is for $R_a > 0.25\%$ (which is expected at 95% confidence), and the cyan region is for $R_a = 100\%$. The green region has too much axion CDM: $\xi_a > 2.9\,\mathrm{eV}$. Each constraint is divided into two parts: the darker part is for a conservative value $\chi = 1/20$ and the lighter part is for a moderate value $\chi = 1$. The orange region has excessive GWs amplitude: $Q_t > 9.3 \times 10^{-6}$. The yellow region has too much axion interaction in stars (darker is firmly ruled out, lighter is for some analyzes). The brown region is excluded by the laboratory ADMX search. The dashed cyan, orange, and brown lines are future targets for isocurvature, GWs, and ADMX searches, respectively.

higher limits. Given that electroweak values for f_a are ruled out, economy suggests that f_a could be associated with unification or Planck scales, rather than the "desert" of particle physics or super-Planckian scales, as indicated by the horizontal regions in Fig. 2-1. This intuition for f_a seems borne out in string theory, where f_a typically lies at or just above the GUT scale, and much lower values are non-trivial to achieve [16].[1] Such high values of f_a correspond to large contributions from axions to cold dark matter (CDM). Indeed, it is only after selection effects are taken into account that the ratio of axion density to entropy is small enough to be consistent with observations [18, 19]. When these effects are included, one finds that the expected density of dark matter in axions is close to the amount of dark matter actually observed [20].

2.1.3 Cosmological Observables, Summary

Quantum fluctuations in an effective inflaton field give rise to the standard adiabatic fluctuations that have grown into our cosmologically observed large-scale structure. If the PQ symmetry undergoes spontaneous symmetry breaking before the end of inflation, quantum fluctuations in the consequent light axion field give rise to isocurvature fluctuations. The amplitude of the isocurvature fluctuations grows with E_I, so upper bounds on the amplitude of isocurvature fluctuations imply upper bounds on E_I.

The purpose of this chapter is to delineate these bounds, extending earlier work on this subject such as [17, 21, 22, 23, 24, 25]. The bounds depend sensitively on the fraction R_a of CDM in the form of axions, which in turn depends not only on f_a, but also on the local initial misalignment angle $\theta_i \in [-\pi, \pi]$. These constraints are shown in Figs. 2-2, for different choices of the axion CDM fraction. We will estimate this fraction using statistical arguments.

In Section 2.2 we calculate the production and late time abundance ξ_a of axions and the amplitude α_a of isocurvature fluctuations, as well as reviewing the amplitude Q_t of primordial GWs. These three observables depend on two micro-physical parameters: the PQ scale f_a (or equivalently, the $T = 0$ axion mass m_a) and the energy scale of inflation E_I (or equivalently, the Hubble scale of inflation H_I), and on one "environmental" parameter: the misalignment angle θ_i. We summarize the final formulae here:

$$\xi_a = \bar{\Lambda}(\theta_i^2 + \sigma_\theta^2) f(\theta_i^2) \chi F, \tag{2.1}$$

$$\alpha_a = \frac{8}{25} \frac{(\bar{\Lambda}/\xi_m)^2}{\langle (\delta T/T)^2_{\rm tot} \rangle} \sigma_\theta^2 (2\theta_i^2 + \sigma_\theta^2) f(\theta_i^2)^2 \chi^2 F^2, \tag{2.2}$$

$$Q_t = \frac{H_I}{5\pi\, \bar{m}_{\rm Pl}} = \frac{E_I^2}{5\sqrt{3}\,\pi\, \bar{m}_{\rm Pl}^2}, \tag{2.3}$$

[1] For example, in weakly coupled heterotic string theory, the model-independent axion has its PQ scale given by $f_a = \alpha_U \bar{m}_{\rm Pl}/(2\pi\sqrt{2})$. A unified coupling $\alpha_U = 1/25$ then gives $f_a \approx 1.1 \times 10^{16}$ GeV [17, 16].

where

$$F \approx \begin{cases} 2.8 \left(\frac{\Lambda}{\Lambda_{\rm Q}}\right)^{2/3} \left(\frac{f_a}{\bar{m}_{\rm Pl}}\right)^{7/6} & \text{for } f_a \lesssim \hat{f}_a, \\ 4.4 \left(\frac{f_a}{\bar{m}_{\rm Pl}}\right)^{3/2} & \text{for } f_a \gtrsim \hat{f}_a, \end{cases} \tag{2.4}$$

$$\sigma_\theta = \gamma \frac{H_I}{2\pi f_a} = \gamma \frac{E_I^2}{2\sqrt{3}\,\pi f_a \bar{m}_{\rm Pl}} \tag{2.5}$$

(definitions are given below).

The most recent observational bounds from WMAP5 combined with other data are [11]

$$\xi_a \leq 2.9\,\text{eV}, \quad \alpha_a < 0.072, \quad Q_t < 9.3 \times 10^{-6}, \tag{2.6}$$

thereby constraining the two micro-physical parameters E_I and f_a.

These expressions for ξ_a and α_a only apply if the PQ symmetry undergoes spontaneous symmetry breaking before inflation and is not restored thereafter. This is true if f_a exceeds the Gibbons-Hawking temperature during inflation and the maximum post-inflationary thermalization temperature, as we discuss in Section 2.2.3. Inefficient thermalization leads to constraints displayed in Fig. 2-2 (top), while efficient thermalization leads to constraints displayed in Fig. 2-2 (bottom).

In Section 2.3, we present a statistical estimate for the axion abundance to place additional constraints on out parameter space. In Section 2.4, we conclude by discussing the implications for inflationary model building and future prospects.

2.2 Axion Cosmology

In this section, we review the production of axions in the early universe, their abundance in the late universe, and the amplitude of isocurvature fluctuations, following Refs. [24, 17], and explain and derive eqs. (2.1)–(2.5). We focus on axion production from the so called "vacuum misalignment" mechanism only. This provides the most conservative constraints. Additional production mechanisms, such as cosmic string decay, are subject to larger theoretical uncertainties (e.g., see [26]).

2.2.1 Onset of Axion Production

In an expanding flat FRW background at temperature T with Hubble parameter $H(T)$, the phase field θ of broken PQ symmetry satisfies the equation of motion

$$\ddot{\theta} + 3H(T)\dot{\theta} - \frac{\nabla^2\theta}{a^2} = -\frac{1}{f_a^2}\frac{\partial}{\partial\theta}V(\theta, T), \tag{2.7}$$

where dots indicate derivatives with respect to co-ordinate time. Here $V(\theta, T)$ is the temperature dependent potential induced by QCD instantons. At zero temperature, $V(\theta, 0) = \bar{\Lambda}^4(1 - \cos\theta)$, where $\bar{\Lambda} \approx 78\,\text{MeV}$ sets the scale of the vacuum energy of

QCD.[2] For small values of the axion field, the potential is approximately harmonic:

$$V(\theta, T) \approx \frac{1}{2} m_a(T)^2 f_a^2 \theta^2. \tag{2.8}$$

The mass is temperature dependent, with high and low T limits given by

$$m_a(T) \approx m_a(0) \begin{cases} b\left(\frac{\Lambda_Q}{T}\right)^4 & \text{for } T \gtrsim \bar{\Lambda}_Q, \\ 1 & \text{for } T \lesssim \bar{\Lambda}_Q, \end{cases} \tag{2.9}$$

where $\bar{\Lambda}_Q \sim 200\,\text{MeV}$ is the scale at which QCD becomes strongly coupled, $b = \mathcal{O}(10^{-2})$ depends on detailed QCD physics, and $m_a(0)$ is the zero temperature axion mass, related to the PQ scale f_a and $\bar{\Lambda}$ by $m_a(0) = \bar{\Lambda}^2/f_a$.[3] The temperature dependence is illustrated in Fig. 2-3.

In the early universe, the axion field is effectively massless, and so the right hand side of eq. (2.7) is negligible. Hence the zero mode of the axion field is essentially frozen due to Hubble friction. When the temperature T drops below T_{osc}, defined by

$$3H(T_{\text{osc}}) \approx m_a(T_{\text{osc}}), \tag{2.10}$$

the axion field will begin to oscillate, producing axions. Since this occurs during the radiation dominated era, we have[4]

$$H(T)^2 = \frac{1}{3\bar{m}_{\text{Pl}}^2} \frac{\pi^2}{30} g_*(T)\, T^4, \tag{2.11}$$

where $\bar{m}_{\text{Pl}} \approx 2.4 \times 10^{18}\,\text{GeV}$ and $g_*(T)$, the effective number of relativistic degrees of freedom, depends on whether T_{osc} occurs before or after the QCD phase transition: $g_*(T_{\text{osc}}) = 61.75$ for $T_{\text{osc}} \gtrsim \bar{\Lambda}_Q$ and $g_*(T_{\text{osc}}) = 10.75$ for $T_{\text{osc}} \lesssim \bar{\Lambda}_Q$. Eqs. (2.9–2.11) allow us to solve for T_{osc} in terms of f_a:

$$T_{\text{osc}} \approx \begin{cases} 0.36\, \Lambda_Q^{2/3} \Lambda^{1/3} \left(\frac{\bar{m}_{\text{Pl}}}{f_a}\right)^{1/6} & \text{for } f_a \lesssim \hat{f}_a, \\ 0.55\, \bar{\Lambda} \left(\frac{\bar{m}_{\text{Pl}}}{f_a}\right)^{1/2} & \text{for } f_a \gtrsim \hat{f}_a, \end{cases} \tag{2.12}$$

where $\hat{f}_a$, which reflects the break in eq. (2.9), is defined in eq. (2.16) below.

---

[2] $\bar{\Lambda}$ is set by $\bar{\Lambda}_Q$ and quark masses: $\bar{\Lambda}^2 = \frac{\sqrt{z}}{1+z} f_\pi m_\pi$, $z \equiv m_u/m_d \approx 0.56$.

[3] If there are N distinct vacua for θ, then we should replace f_a by f_a/N here and throughout the chapter. However, any $N > 1$ models are expected to have a large overabundance of energy density from domain walls, unless inflation intervenes.

[4] We assume here that the universe before BBN is adequately described by conventional physics. See Ref. [27] for other scenarios.

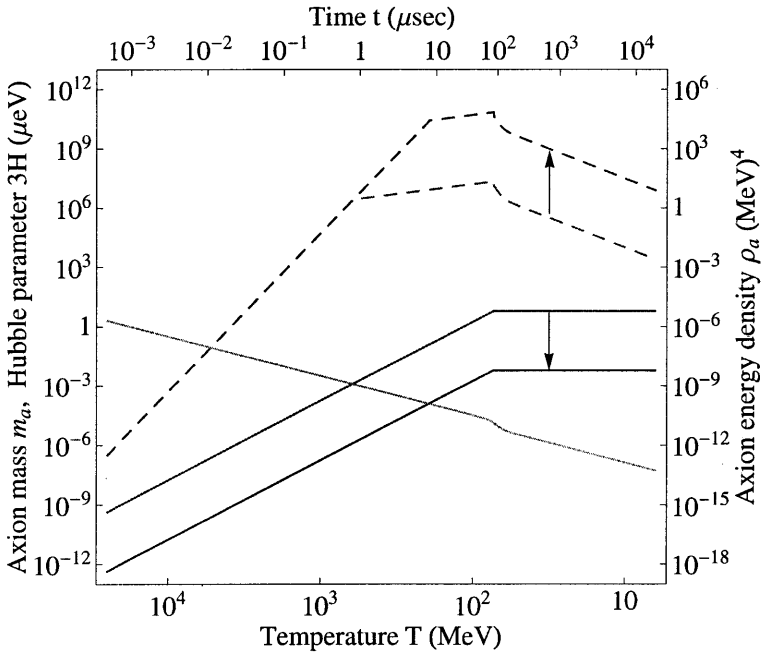

Figure 2-3: Axion mass m_a (solid) and axion energy density ρ_a (dashed), as a function of temperature T (or time t). The red and blue curves are for PQ breaking scales $f_a = 10^{12}\,\mathrm{GeV}$ and $f_a = 10^{15}\,\mathrm{GeV}$, respectively. The arrow indicates the effect of increasing f_a. The decreasing orange curve is the Hubble parameter $3H$. We have taken $\langle\theta^2\rangle = \pi^2/3$, although this is modified by the intervention of inflation, as explained in the text. For simplicity, we have here only kept track of the variation with temperature in the number of relativistic degrees of freedom before ($g_* = g_{*S} = 61.75$) and after ($g_* = g_{*S} = 10.75$) the QCD phase transition, and taken $\chi = 1$.

2.2.2 Density of Axions

At the onset of production (when $T = T_{\rm osc}$) the axion energy density is

$$\rho_a(T_{\rm osc}) \approx \frac{1}{2} m_a(T_{\rm osc})^2 f_a^2 \langle\theta^2\rangle f(\theta_i^2)\chi. \tag{2.13}$$

Here $\langle\theta^2\rangle$ is the spatial average over our Hubble volume of the square of the initial misalignment. In terms of its mean θ_i and standard deviation σ_θ, $\langle\theta^2\rangle = \theta_i^2 + \sigma_\theta^2$. If the axion field is established before (or during the early stages of) inflation, then spatial variations in θ are smoothed out over our Hubble volume ($\nabla^2\theta/a^2 \to 0$). Then $\theta_i = \langle\theta\rangle$ in our Hubble volume is an angle drawn from a uniform distribution: $\theta_i \in [-\pi, \pi]$, with a small variance that we discuss in the next subsection. On the other hand, if the axion field is established after inflation, then $\theta_i = \langle\theta\rangle = 0$, with variance $\sigma_\theta^2 = \pi^2/3$ due to small scale variations. $f(\theta_i^2)$ is a fudge factor acknowledging anharmonicity in the axion potential; for $\theta_i \to 0$, $f(\theta_i^2) \to 1$. Finally, χ is a dimensionless correction factor due to temperature dependence during formation. In our numerics, we take $\Lambda_{\rm Q} = 200\,{\rm MeV}$, $b = 0.018$, and absorb all theoretical uncertainties into χ. A conservative value is $\chi = 1/20$ and a more moderate value is $\chi = 1$;[5] both values are reported in Fig. 2.

By following the redshift as the universe expands from axion-formation to today, we can convert this initial energy density into a prediction for the present density, as illustrated in Fig. 2-3. Since we are focusing on the zero-mode, the axions form a non-relativistic Bose-Einstein condensate. At late times (say today's temperature T_0), the axion energy per photon is

$$\xi_a = \frac{\rho_a(T_0)}{n_\gamma(T_0)} = \frac{m_a(T_0)}{m_a(T_{\rm osc})} \frac{\rho_a(T_{\rm osc})}{n_\gamma(T_0)} \frac{s(T_0)}{s(T_{\rm osc})}, \tag{2.14}$$

where we have exploited the fact that entropy density $s(T)$ in a comoving volume is conserved. The entropy density is given by

$$s(T) = \frac{2\pi^2}{45} g_{*s}(T)\, T^3, \tag{2.15}$$

where $g_{*s}(T)$ is the effective number of relativistic degrees of freedom for $s(T)$ [30]. Note that $g_{*s}(T_{\rm osc}) = g_*(T_{\rm osc})$, $g_{*s}(T_0)$ comes from photons and neutrinos: $g_{*s}(T_0) = 2 + \frac{7}{8} \times 6 \times \frac{4}{11} = 3.91$, and $m_a(T_0) \approx m_a(0)$. Since the number density of photons $n_\gamma(T_0) = 2\zeta(3)T_0^3/\pi^2$ depends on temperature in the same way that the late time axion energy density $\rho_a(T_0)$ does, the quantity ξ_a is a temperature-independent, or equivalently time-independent, measure of the axion abundance. In contrast, the commonly used quantities Ω_a and $h^2\Omega_a$ do not tell us anything fundamental about our universe, since like T, they are effectively alternative time variables that evolve as our universe expands. These different measures of axion density are related by $\Omega_a h^2 \approx 0.0019\,(\xi_a/1\ {\rm eV})(T_0/1\ {\rm K})^3$, which at the present epoch ($T_0 = 2.725\,{\rm K}$) reduces

[5] See Refs. [28, 29] for precise estimates of axion abundance.

to $\Omega_a h^2 \approx \xi_a/26\,\mathrm{eV}$ (h is the dimensionless Hubble parameter).

Combining all this information yields eq. (2.1). Note that higher f_a (for a fixed value of $\langle\theta^2\rangle$) corresponds to higher axion energy density, as seen in Fig. 2-3. The reason for this is that higher f_a corresponds to lower $m_a(0)$, so that the onset of axion production, when $3H(T)$ has fallen to $m_a(T)$, occurs later. Hence there is less redshifting of the axion energy density after production (furthermore, $\rho_a(T_{\rm osc})$ is higher if $f_a < \hat{f}_a$).

We locate the boundary between the low and high f_a limits by equating the two expressions for ξ_a. They match when $f_a = \hat{f}_a$, where

$$\hat{f}_a \equiv 0.26 \left(\frac{\Lambda}{\bar{\Lambda}_{\rm Q}}\right)^2 \bar{m}_{\rm Pl}. \tag{2.16}$$

The observed total density of cold dark matter in our universe $\xi_{\rm CDM} \approx 2.9\,\mathrm{eV}$ implies $\xi_a \leq 2.9\,\mathrm{eV}$ [11].

2.2.3 Fluctuations from Inflation

During inflation, the universe undergoes an approximately de Sitter phase with Hubble parameter H_I. Quantum fluctuations during this phase induce several kinds of cosmological fluctuations.

- Adiabatic density fluctuations are generated, with an approximately scale invariant spectrum. The measured amplitude is $Q \approx 1.98 \times 10^{-5}$ [11].[6]

- Primordial gravity waves are generated, with an approximately scale-invariant spectrum whose amplitude is given in eq. (2.3). WMAP5 plus BAO and SN data imply the bound: $Q_t < 9.3 \times 10^{-6}$ (95%). Since Q is measured and Q_t is set by E_I, the tensor to scalar ratio $r \equiv (Q_t/Q)^2$ is often used to characterize the scale of inflation. Using the same notation as [11], it is bounded by $r < 0.22$ (95%).

- Any other light scalar fields, such as the axion, are imprinted with fluctuations during inflation, similarly to gravitons. The power spectrum of a canonically normalized scalar field ϕ, such as ϕ_a, in de Sitter space has a scale-invariant spectrum (e.g., see [31])

$$\langle|\delta\phi_a(k)|^2\rangle = \left(\frac{H_I}{2\pi}\right)^2 \frac{1}{k^3/2\pi^2}. \tag{2.17}$$

  It is essentially a thermal spectrum at Gibbons-Hawking temperature $T_{\rm GH} = H_I/2\pi$. Fluctuations in the misalignment angle in k-space are scaled as $\sigma_\theta = \sigma_a/f_a$, per eq. (2.5). We write the corresponding fluctuations in real space as $\sigma_a = \gamma\, H_I/2\pi$, where $\gamma = \mathcal{O}(1)$ is a dimensionless constant. Ref. [11] effectively takes $\gamma = 1$ and Ref. [22] argues that observations are sensitive to length scales

[6] Here $Q = \frac{2}{5}\bar{\Delta}_R(k = 0.002\,\mathrm{Mpc}^{-1})$ of Ref. [11].

corresponding to $\gamma \approx 4$, while in our figures we have taken a moderate value of $\gamma = 2$. These fluctuations provide a lower bound on ξ_a and, as we discuss in the next subsection, on isocurvature fluctuations.

In the preceding discussion, we assumed the existence of a light axion field during inflation. This is true only if PQ symmetry is broken before inflation. Furthermore, if PQ symmetry is restored after inflation, the fluctuations will be washed out. PQ symmetry can be restored either by the Gibbons-Hawking temperature during inflation, or by the maximum thermalization temperature after inflation $T_{\rm max}$.[7] To characterize the maximum thermalization temperature, we use a dimensionless efficiency parameter $\epsilon_{\rm eff}$ defined as $T_{\rm max} = \epsilon_{\rm eff} E_I$, with $0 < \epsilon_{\rm eff} < 1$, with $\epsilon_{\rm eff} \ll 1$ expected.

A robust criterion for the presence of the axion during inflation with fluctuations that survive is

$$f_a > {\rm Max}\{T_{\rm GH}, T_{\rm max}\}. \tag{2.18}$$

If this condition is satisfied, inflationary expansion implies that $\theta_i \in [-\pi, \pi]$ is drawn from a uniform distribution. By postulating that θ_i is atypically small in our neighborhood (i.e., in our Hubble volume) one can accommodate large f_a. This defines what we term the anthropic regime (see Fig. 2-2).

Alternatively, if $f_a < {\rm Max}\{T_{\rm GH}, T_{\rm max}\}$, then either there is no axion during inflation or its effects are washed out after inflation. In this case θ^2 fluctuates throughout our observable universe, with variance $\pi^2/3$, and there are no appreciable axion-induced isocurvature fluctuations. This defines what we term the classic regime (see Fig. 2-2).

2.2.4 Isocurvature Fluctuations

Fluctuations in the local equation of state $\delta(n_i/s) \neq 0$ at fixed total energy density $\delta\rho = 0$ are known as isocurvature fluctuations. (In contrast, fluctuations with $\delta(n_i/s) = 0$ and $\delta\rho \neq 0$ are known as adiabatic fluctuations.) Since the axion is essentially massless in the early universe, at temperatures much greater than the QCD phase transition ($T \gg \Lambda_{\rm Q}$), its energy density is entirely negligible at early times. Hence, at such early times, fluctuations in the number density of axions (established by de Sitter fluctuations during inflation) do not alter the energy density of the universe. Later, for temperatures below the QCD phase transition ($T \lesssim \Lambda_{\rm Q}$), the axion acquires a mass and a significant energy density (see Fig. 2-3), but any such fluctuations cannot alter the total energy density of the universe, by local conservation of energy. In the early radiation dominated era, this means that fluctuations in the axion energy density are compensated by fluctuations in photons and other relativistic fields. Hence, these are isocurvature fluctuations.[8]

[7] The maximum thermalization temperature should not be confused with the reheating temperature, which can be somewhat lower [32]. The maximum thermalization temperature is the maximum temperature of the thermal bath post-inflation, while the reheating temperature is the temperature at the end of the reheating phase, i.e., at the beginning of the radiation era.

[8] Later, around the onset of the matter dominated era, these isocurvature fluctuations are converted to adiabatic fluctuations, responsible for the familiar gravitational structures in our universe.

To quantify the amplitude of isocurvature fluctuations, it is useful to introduce the fractional change in the number density to entropy density ratio:

$$S_i = \frac{\delta(n_i/s)}{n_i/s} = \frac{\delta n_i}{n_i} - 3\frac{\delta T}{T}. \tag{2.19}$$

For adiabatic fluctuations, $S_i = 0$. We assume that this is true for all species other than the axion. Isocurvature fluctuations in the total energy density involve a sum over all massive species and radiation:

$$0 = \delta\rho_{\rm iso} = m_a \delta n_a + \sum_{i \neq a} m_i \delta n_i + 4\rho_r \frac{\delta T}{T}. \tag{2.20}$$

These two equations will be used to obtain an expression for the corresponding temperature fluctuations.

Initially the energy density of the axion field is a small fraction of the ambient total density, so eq. (2.20) gives $\delta T/T \ll \delta n_a/n_a$, and $S_a = \delta n_a/n_a$. Since $n_a \propto \theta^2$ (ignoring anharmonic effects), this implies

$$S_a = \frac{\theta^2 - \langle\theta^2\rangle}{\langle\theta^2\rangle}. \tag{2.21}$$

Assuming $\delta\theta \equiv \theta - \langle\theta\rangle$ is Gaussian distributed[9], we can calculate $\langle S_a^2\rangle$ in terms of $\theta_i = \langle\theta\rangle$ and $\sigma_\theta^2 = \langle(\delta\theta)^2\rangle$, as

$$\langle S_a^2\rangle = \frac{2\sigma_\theta^2(2\theta_i^2 + \sigma_\theta^2)}{(\theta_i^2 + \sigma_\theta^2)^2}. \tag{2.22}$$

Note that if $\theta_i^2 \ll \sigma_\theta^2$ then $\langle S_a^2\rangle = 2$, while if $\theta_i^2 \gg \sigma_\theta^2$ then $\langle S_a^2\rangle = 4\sigma_\theta^2/\theta_i^2$.

The most important axion induced temperature fluctuations are those on the largest scales. Such fluctuations enter the horizon well into the matter dominated era, where ρ_r can be ignored. This implies[10]

$$\left(\frac{\delta T}{T}\right)_{\rm iso} \approx -\frac{\xi_a}{3\,\xi_m} S_a, \tag{2.23}$$

where ξ_m is the total matter energy density per photon, whose measured value is $\xi_m = 3.5\,\text{eV}$ [11].

Following [11], we define α_a to be the fractional contribution to the CMB temperature power spectrum due to axion isocurvature:

$$\alpha_a \equiv \frac{\langle(\delta T/T)^2_{\rm iso}\rangle}{\langle(\delta T/T)^2_{\rm tot}\rangle}. \tag{2.24}$$

[9]This is a good assumption in the regime where the axions comprise a significant fraction of the dark matter, i.e., $\theta_i^2 \gg \sigma_\theta^2$.

[10]Due to the Sachs-Wolfe effect, there is a 20% enhancement to (2.23), but we will not go into those details here.

Using the relationship between S_a and $\delta T/T$ in eq. (2.23) and the preceding expression for $\langle S_a^2 \rangle$ we obtain

$$\alpha_a = \frac{8}{25} \frac{(\xi_a/\xi_m)^2}{\langle (\delta T/T)^2_{\rm tot} \rangle} \frac{\sigma_\theta^2 (2\theta_i^2 + \sigma_\theta^2)}{(\theta_i^2 + \sigma_\theta^2)^2}. \tag{2.25}$$

COBE measured (and WMAP confirmed) the root-mean-square total temperature fluctuations to be $\langle (\delta T/T)^2_{\rm tot} \rangle \approx (1.1 \times 10^{-5})^2$, averaged over the first few l. Using the expression for ξ_a given in eq. (2.1) we can write the isocurvature fluctuations as in eq. (2.2). This must be consistent with the latest observational bound $\alpha_a < 0.072$. Here we have used α_0 of Ref. [11], which assumes isocurvature fluctuations are uncorrelated from curvature ones.

2.3 Direct and Statistical Constraints

It is conceivable that the axion abundance is negligible (but see the following subsection). This scenario (case (i)) requires $\theta_i^2 \ll \sigma_\theta^2$. By demanding $\alpha_a < 0.072$ (the current isocurvature bound) and using eq. (2.2), we obtain the most conservative bound, studied in Ref. [17], corresponding to the purple region marked "Any R_a" in Fig. 2-2.

At the other extreme, if axions are the dominant form of dark matter in the universe (case iii), then $\theta_i^2 \gg \sigma_\theta^2$. Again demanding $\alpha_a < 0.072$ in eq. (2.2), with θ_i determined from eq. (2.1) with $\xi_a = \xi_{\rm CDM}$, the excluded region expands to include the cyan region marked "$R_a = 100\%$" in Fig. 2-2 (as well as the blue region marked "$R_a > 0.25\%$").

Each of these three regions are bifurcated by a line. In all three cases, the rightmost part gives the most robust constraint, coming from a conservative value $\chi = 1/20$, while the leftmost part extend the constraints using a more moderate (and more speculative) value $\chi = 1$. This comes from our uncertainty in the total axion abundance.

2.3.1 Statistics of a Two-Component Model

The viability of large f_a axion cosmology depends on taking selection effects seriously, since they produce a higher dark matter density ξ than observed in most Hubble volumes. In particular, the density of a typical galaxy scales as $\rho \sim \xi^4$. Taking into account that denser galaxies have fewer stable solar systems due to close encounters with other stars, etc., it has been found that typical stable solar systems in large f_a axion models reside in Hubble volumes where ξ is comparable to the observed value [20].

Here we draw out a statistical implication for the predicted axion abundance, if there is a second contributor to the dark matter density. Consider the hypothesis that the total CDM ($\xi_{\rm CDM}$) is comprised of axions (ξ_a) and some other component, say WIMPs (ξ_W): $\xi_{\rm CDM} = \xi_a + \xi_W$. The unknown separate axion and WIMP abundances should be drawn from prior distributions determined by underlying micro-physical

theories. For axions in the large f_a regime, above any inflation temperatures, this scenario implies that the initial misalignment angle θ_i is uniformly distributed. In the regime where the axion abundance is non-negligible ($\theta_i^2 \gg \sigma_\theta^2$), but still sufficiently small that we can ignore anharmonic effects ($\theta_i^2 \ll 1$), we have $\xi_a \propto \theta_i^2$. Since θ_i is uniformly distributed, it is simple to show that

$$p^{(a)}_{\rm prior}(\xi_a) \propto \frac{1}{\sqrt{\xi_a}}. \tag{2.26}$$

In contrast, we do not have a reliable prior distribution $p^{(W)}_{\rm prior}(\xi_W)$ for the WIMP.

As discussed in Ref. [20], selection effects depend only on the sum $\xi_{\rm CDM} = \xi_a + \xi_W$, so the total joint distribution for axions and WIMPs is

$$p(\xi_a, \xi_W) \propto p^{(a)}_{\rm prior}(\xi_a) p^{(W)}_{\rm prior}(\xi_W) p_{\rm selec}(\xi_{\rm CDM}) \tag{2.27}$$

As demonstrated in Ref. [20], the observed value of CDM $\xi_{\rm CDM} \approx 2.9\,{\rm eV}$ is nicely consistent with this distribution. Given this, we can focus on the remaining one-dimensional distribution for the axion:

$$p(\xi_a) \propto p^{(a)}_{\rm prior}(\xi_a) p^{(W)}_{\rm prior}(\xi_{\rm CDM} - \xi_a). \tag{2.28}$$

Unless $p^{(W)}_{\rm prior}$ is sharply peaked at $\xi_{\rm CDM}$, the axion prior (when integrated) disfavors very small values of ξ_a. For example, let us take the WIMP prior to be uniform. We can then make a prediction for the axion abundance at, say, 95% confidence. Defining $\hat{\xi}_a$ implicitly through

$$\int_0^{\hat{\xi}_a} p(\xi_a)\, d\xi_a = 0.05 \tag{2.29}$$

and solving eq. (2.29) using eq. (2.26), we find $\hat{\xi}_a = (0.05)^2 \xi_{\rm CDM} = 0.25\%\, \xi_{\rm CDM}$. This says that it statistically unlikely – at the 95% level – for axions to comprise less than 0.25% of the CDM of the universe (case(ii)).

By setting $\xi_a = 0.25\%\, \xi_{\rm CDM}$, we rule out the blue region marked $R_a > 0.25\%$ in Fig. 2-2 with high confidence. In other words, without assuming that axions comprise all the CDM, we find that on statistical grounds axions must comprise at least a non-negligible fraction of the universe's CDM, allowing us to extend the excluded region in Fig. 2-2 further towards the upper left.

2.3.2 Additional Constraints

The preceding applies in the $f_a > {\rm Max}\{T_{\rm GH}, T_{\rm max}\}$ regime, where the initial misalignment angle θ_i takes on a single constant value in our Hubble volume. For $f_a < {\rm Max}\{T_{\rm GH}, T_{\rm max}\}$, the misalignment angle varies on cosmologically small scales, with average $\langle\theta^2\rangle = \pi^2/3$. In this regime the isocurvature fluctuations are negligible. In this case, bounds arise from the requirement that the axion abundance is not greater than the observed total CDM abundance: $\xi_a \leq \xi_{\rm CDM} \approx 2.9\,{\rm eV}$. Using the upper expression for ξ_a in eq. (2.1), with $\theta_i^2 + \sigma_\theta^2 \to \langle\theta^2\rangle = \pi^2/3$, we find that

$f_a > 2.3 \times 10^{11}\,\chi^{-6/7}\,\mathrm{GeV}$ is ruled out. For the conservative value $\chi = 1/20$, this excludes the upper part of the green region marked "Too much CDM" in Fig. 2-2, and for the moderate value $\chi = 1$ this extends the exclusion to the lower part of the green region.[11]

Also, $m_a(0) > 10^3\,\mu\mathrm{eV}$ is firmly ruled out (and $m_a(0) > 10^4\,\mu\mathrm{eV}$ for some analyzes), since in this regime the coupling of axions to matter is too large, affecting the physics of stars, such as the cooling of red giants and the neutrino flux from SN 1987A [33] (yellow region at bottom of Fig. 2-2.) Furthermore, the ADMX search for axion dark matter in a microwave cavity detector has ruled out axions comprising the bulk of the halo dark matter in the following mass window: $1.9\,\mu\mathrm{eV} < m_a(0) < 3.3\,\mu\mathrm{eV}$ (brown band in Fig. 2-2) for so-called KSVZ axions, and the sub-window $1.98\,\mu\mathrm{eV} < m_a(0) < 2.17\,\mu\mathrm{eV}$ for so-called DFSV axions [34, 35]. The remaining white region is the allowed "classic window".

In Fig. 2-2 (top), corresponding to inefficient thermalization ($\epsilon_{\mathrm{eff}} \approx 0$), the boundary between the anthropic and classic regimes is $f_a = T_{\mathrm{GH}} = H_I/2\pi$. In Fig. 2-2 (bottom), corresponding to efficient thermalization ($\epsilon_{\mathrm{eff}} = 10^{-3}$, with $\epsilon_{\mathrm{eff}} = 10^{-1.5}$, 1 indicated), the boundary between the two regimes is $f_a = T_{\mathrm{max}} = \epsilon_{\mathrm{eff}} E_I$. Efficient thermalization thus opens up a larger "classic window", but the "anthropic window" is essentially unaltered.

2.3.3 Effect of Falling Density During Inflation

In our analysis, we have treated inflation as occurring at one rather well-defined Hubble scale. Although this is a good approximation in some inflation models, there are others giving an appreciable change in H between its value (say H_I) when the modes that are now re-entering our horizon left the horizon (55 or so e-foldings before the end of inflation), and its value (say H_{end}) at the end of inflation. This is particularly relevant to Fig. 2-2 (top), since it implies that the boundary between "anthropic" and "classic" regimes is blurred, since $T_{\mathrm{GH}} = H/2\pi$ is evolving. For high scale inflation, H typically changes by an amount of order the number of e-foldings, i.e., $\mathcal{O}(10^2)$, while for low scale inflation models, H typically changes very little.

If we consider $H_{\mathrm{end}} \ll H_I$, then the PQ symmetry can break *during* inflation. The resulting cosmology could be quite interesting with axion dark matter varying appreciably from one point in our Hubble volume to another, but is ruled out since $Q \sim 10^{-5}$. If PQ breaking occurs very close to 55 or so e-foldings before the end of inflation, then θ_i can be smoothed out on today's cosmological scales and make the analysis "anthropic". Otherwise, we expect the "classic" analysis to apply as usual, providing a ruled out green region in Fig. 2-2. Hence, we expect such corrections to the constraints to be reasonably minimal.

[11] If $f_a > \mathrm{Max}\{T_{\mathrm{GH}}, T_{\mathrm{max}}\}$, there is another region ruled out with too much CDM (green region above thick red line in Fig. 2-2.)

2.4 Discussion

We have surveyed observational constraints on the parameter space $\{E_I, f_a\}$ of inflation and axions, finding that most of it is excluded, leaving only two allowed regions that we term classic and anthropic windows. Part of the classic window $f_a \sim 10^{11}$ – 10^{12} GeV will be intensely explored by the ongoing ADMX experiment. The region indicated by the arrow to the horizontal brown lines in Fig. 2-2 to $m_a(0) = 10\,\mu$eV is expected to be explored by the end of ADMX Phase II, and onwards to $m_a(0) = 100\,\mu$eV some years thereafter [36]. In this window, comparatively little can be concluded about the scale of inflation. From Fig. 2-2 (top), taking $m_a(0) = 100\,\mu$eV and assuming $\xi_a > 0.25\%\,\xi_{\rm CDM}$, we rule out $8.4 \times 10^{13}\,{\rm GeV} \lesssim E_I \lesssim 1.3 \times 10^{15}\,{\rm GeV}$.[12] From Fig. 2-2 (bottom), the upper end of this ruled out region is reduced due to efficient post-inflation thermalization. Although we can rule out a range of low scale inflation models, these conclusions are not exceedingly strong.

On the other hand, a large f_a axion has strong implications for inflation. According to both Figs. 2-2 (top) & (bottom), if $f_a = 10^{16}\,{\rm GeV}$ then $E_I \gtrsim 5.5 \times 10^{14}\,{\rm GeV}$ ($r \gtrsim 9 \times 10^{-9}$) is ruled out at 95% confidence for the conservative value $\chi = 1/20$, and $E_I \gtrsim 2.6 \times 10^{14}\,{\rm GeV}$ ($r \gtrsim 4 \times 10^{-10}$) is ruled out for the moderate value $\chi = 1$. The geometric mean is $E_I \gtrsim 3.8 \times 10^{14}\,{\rm GeV}$ ($r \gtrsim 2 \times 10^{-9}$), which is reported in the abstract. This is incompatible with many models of inflation, including "classic" models with a single slow-rolling scalar field in a generic potential. For example, monomial potentials $V \propto \phi^p$ predict $r = 4\,p/N_e$, where N_e is the number of e-foldings of inflation from when it generated our horizon scale fluctuations to when it ended. For such models, N_e around 50 or 60 is expected, so any reasonable p is ruled out, including ϕ^2 chaotic inflation [4] and the stringy N-flation [37] and Monodromy [38] models. The same is true for exponential potentials $V \propto \exp(-\sqrt{2p}\,\phi/\bar{m}_{\rm Pl})$, which predict $r = 16p$.

Evidently there is considerable tension between the theoretically appealing large f_a and high-scale inflation scenario (see Fig. 2-1) and the observational constraints (see Fig. 2-2). Low-scale inflation may be emerging as favored from recent work in string theory. If we consider the small subspace (see [39, 40, 41]) of presently constructed string models that both inflate and agree with the observed values of Q and n_s, we are left with models that tend to be at rather small energies, typically $r < 10^{-8}$ for D-brane models and various other scenarios such as modular inflation [42]. There are also arguments for very low r in the simple KKLT framework discussed in Ref. [43]. This allows $f_a \sim 10^{16}\,{\rm GeV}$ to be marginally consistent with present isocurvature bounds. Although it is highly premature to conclude that very low energy scale is a generic feature of string realizations of inflation, it is intriguing that many string constructions have this feature. (See [37, 38] for interesting exceptions.)

The Planck satellite, CMBPol, and upcoming suborbital CMB experiments should probe well beyond the current bound on GWs of $r < 0.22$, perhaps reaching $r \sim 0.01$. This is indicated by an arrow toward the vertical dashed red line in Fig. 2-2. If gravity

[12]The quoted lower end of the ruled out region is the geometric mean of the conservative and moderate ξ_a-scenarios.

waves are observed in this regime, then the PQ scale f_a must be in the classic window.

Our considerations emphasize the fundamental importance of improving bounds on isocurvature fluctuations. For example, an order of magnitude improvement to $\alpha \sim 0.007$ would push the isocurvature bounds to the diagonal dashed cyan line in Fig. 2-2. (We have indicated the improvement for the case where axions comprise all the CDM: $\xi_a = \xi_{\rm CDM}$.) Detection of isocurvature fluctuations in this regime has three important implications:

1. It could be interpreted as evidence for the existence of the axion field, and assuming this:

2. It would probe low inflation scales E_I far beyond the scope of any foreseeable GW measurements.

3. It would be evidence that we live in a highly atypical Hubble volume, i.e., $\{E_I, f_a\}$ must be in the anthropic window.

Isocurvature modes and tensor modes thus provide complementary constraints on fundamental physics, making it fruitful to study dark matter and inflation in a unified way.

Bibliography

[1] A. Guth, "The inflationary universe: A possible solution to the horizon and flatness problems," Phys. Rev. D, 23, 347 (1981).

[2] A. D. Linde, "A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems," Phys. Lett. B, 108, 389 (1982).

[3] A. Albrecht and P. J. Steinhardt, "Reheating an Inflationary Universe," Phys. Rev. Lett., 48, 1220 (1982).

[4] A. D. Linde, "Chaotic Inflation," Phys. Lett. B, 129, 177 (1983).

[5] G. Dvali and S.H. Tye, "Brane inflation," Phys. Lett. B450 (1999) 72 [arXiv:hep-ph/9812483].

[6] S. Kachru et al, "Towards inflation in string theory," JCAP 0310 (2003) 013 [arXiv:hep-th/0308055].

[7] A. R. Liddle and D. H. Lyth, "Cosmological inflation and large-scale structure," Cambridge Univ. Press, Cambridge (2000)

[8] D. Lyth, "What would we learn by detecting a gravitational wave signal in the cosmic microwave background anisotropy?" Phys. Rev. Lett. 78 (1997) 1861-1863 [arXiv:hep-ph/9606387v1].

[9] D. Baumann and L. McAllister, "A microscopic limit on gravitational waves from D-brane inflation," (2006) [arXiv:hep-th/0610285v1].

[10] L. A. Boyle, P. J. Steinhardt, and N. Turok, "Inflationary predictions for scalar and tensor fluctuations reconsidered," Phys. Rev. Lett. 96 (2006) 111301 [arXiv:astro-ph/0507455v3].

[11] E. Komatsu et al, "Five-year Wilkinson microwave probe (WMAP) observations: Cosmological interpretation."

[12] M. Tegmark "What does inflation really predict?" JCAP 0504 (2005) 001 [arXiv:astro-ph/0410281v2].

[13] R. D. Peccei and H. R. Quinn, "CP conservation in the presence of instantons," Phys. Rev. Lett. 38, 1440 (1977).

[14] S. Weinberg, "A new light boson?," Phys. Rev. Lett. 40, 223 (1978).

[15] F. Wilczek, "Problem of strong P and T invariance in the presence of instantons," Phys. Rev. Lett. 40, 279 (1978).

[16] P. Svrcek and E. Witten, "Axions in string theory," JHEP 0606 (2006) 051 [hep-th/0605206v2].

[17] P. Fox, A. Pierce, and S. Thomas, "Probing a QCD string axion with precision cosmological measurements," [arXiv:hep-th/0409059 v1].

[18] A. D. Linde, "Axions in inflationary cosmology," Phys. Lett. B 259, 38 (1991).

[19] F. Wilczek, "A model of anthropic reasoning, addressing the dark to ordinary matter coincidence," (2004) [hep-ph/0408167].

[20] M. Tegmark, A. Aguirre, M. J. Rees, and F. Wilczek, "Dimensionless constants, cosmology and other dark matters," Phys. Rev. D73 (2006) 023505 [arXiv:astro-ph/0511774].

[21] M. Turner and F. Wilczek, "Inflationary axion cosmology," PRL 66, 5 (1991).

[22] D. H. Lyth, "Axions and inflation: Vacuum fluctuations," Phys. Rev. D 45 3394 (1992).

[23] D. H. Lyth and E. D. Stewart, "Axions and inflation: String formation during inflation," Phys. Rev. D 46, 532 (1992).

[24] S. D. Burns, "Isentropic and isocurvature axion perturbations in inflationary cosmology," [arXiv:astro-ph/9711303v1].

[25] M. Beltran, J. Garcia-Bellido, and J. Lesgourgues, "Isocurvature bounds on axions revisited," Phys. Rev. D 75, 103507 (2007), [arXiv:hep-ph/0606107v3].

[26] P. Sikivie, "Axion Cosmology," [arXiv:astro-ph/0610440v2].

[27] D. Grin, T. L. Smith, and M. Kamionkowski, "Axion constraints in non-standard thermal histories," Phys. Rev. D 77, 085020 (2008), arXiv:0711.1352v2 [astro-ph].

[28] K. J. Bae, J. Huh, and J. E. Kim, "Update of axion CDM energy density," (2008) arXiv:0806.0497v2 [hep-ph].

[29] J. E. Kim and G. Carosi, "Axions and the strong CP problem," (2008) arXiv:0807.3125v1 [hep-ph].

[30] E. Kolb and M. Turner, "The early universe," Chap. 3, Basic Books (1990).

[31] A. Linde, "Particle physics and inflationary cosmology," (2005) [arXiv:hep-th/0503203v1].

[32] E. W. Kolb, A. Notari, and A. Riotto, "On the reheating stage after inflation," Phys. Rev. D 68 (2003) 123505, [hep-ph/0307241].

[33] Particle Data Group, http://pdg.lbl.gov/

[34] S. J. Asztalos et al, "Improved rf cavity search for halo axions," Phys. Rev. D 69, 011101(R) (2004).

[35] L. D. Duffy et al, "High resolution search for dark-matter axions," Phys. Rev. D 74, 012006 (2006).

[36] G. Rybka, Private communication.

[37] S. Dimopoulos, S. Kachru, J. McGreevy, and J. Wacker, "N-flation," (2005) [arXiv:hep-th/0507205v3].

[38] E. Silverstein and A. Westphal, "Monodromy in the CMB: Gravity waves and string inflation," arXiv:0803.3085v2 [hep-th] (2008).

[39] D. Baumann, A. Dymarsky, I. R. Klebanov, L. McAllister, and P. J. Steinhardt, "A delicate universe," arXiv:0705.3837v1 [hep-th] (2007).

[40] 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. D 76, 103521 (2007), arXiv:0709.0002v3 [astro-ph].

[41] M. P. Hertzberg, S. Kachru, W. Taylor, and M. Tegmark, "Inflationary constraints on type IIA string theory," JHEP 12 (2007) 095, arXiv:0711.2512v2 [hep-th].

[42] J. Conlon and F. Quevedo, "Kahler moduli inflation," JHEP 0601 (2006) 146, [arXiv:hep-th/0509012v2].

[43] R. Kallosh and A. Linde, "Landscape, the scale of SUSY breaking, and inflation," JHEP 0412 (2004) 004 [arXiv:hep-th/0411011v5].

Chapter 3

Searching for Inflation in Simple String Theory Models: An Astrophysical Perspective

Attempts to connect string theory with astrophysical observation are hampered by a jargon barrier, where an intimidating profusion of orientifolds, Kähler potentials, *etc.* dissuades cosmologists from attempting to work out the astrophysical observables of specific string theory solutions from the recent literature. We attempt to help bridge this gap by giving a pedagogical exposition with detailed examples, aimed at astrophysicists and high energy theorists alike, of how to compute predictions for familiar cosmological parameters when starting with a 10-dimensional string theory action. This is done by investigating inflation in string theory, since inflation is the dominant paradigm for how early universe physics determines cosmological parameters.

We analyze three explicit string models from the recent literature, each containing an infinite number of "vacuum" solutions. Our numerical investigation of some natural candidate inflatons, the so-called "moduli fields," fails to find inflation. We also find in the simplest models that, after suitable field redefinitions, vast numbers of these vacua differ only in an overall constant multiplying the effective inflaton potential, a difference which affects neither the potential's shape nor its ability to support slow-roll inflation. This illustrates that even having an infinite number of vacua does not guarantee having inflating ones. This may be an artifact of the simplicity of the models that we study. Instead, more complicated string theory models appear to be required, suggesting that identifying the inflating subset of the string landscape will be challenging.

3.1 Introduction

String theory is currently the most popular candidate for a consistent theory of quantum gravity, but the goal of confronting it with observation remains elusive. It is sometimes said that testing string theory requires prohibitively high energy accelerators, in order to probe the Planck scale predictions of the theory. There are non-trivial

tests of string theory at low energies, however, such as the requirement to have a solution with the standard model of particle physics. Furthermore, it is plausible that we can test string theory by turning to cosmology. The earliest moments of our universe involved extreme energies and the fingerprints of its birth are revealed today by precision measurements of the cosmic microwave background [1] and the large-scale structure of the universe [2, 3]. A highly non-trivial test of string theory then is whether it can reproduce our cosmology. With inflation emerging as the paradigm of early universe phenomena, string theory or any competing theory of quantum gravity must be able to realize this. Moreover, merely producing many e-foldings of inflation is not good enough: the details of inflation must give correct predictions for as many as eight cosmological parameters which have been measured or constrained [4].

Although there have been substantial efforts in the string theory literature aimed at identifying and counting long-lived potential energy minima (so-called vacua) [5, 6, 7, 8, 9, 10], the key cosmological observables depend also on the history of how our spacetime region evolved to this minimum by slow-roll inflation and/or tunneling. This chapter is aimed at discussing one of the simplest realizations of slow-roll inflation from string theory, where the inflaton fields are the so-called moduli which, loosely speaking, correspond to the size and shape of curled up extra dimensions (see Table 3.1) [11, 12]. An alternative scenario involves dynamical branes [13, 14, 15, 16, 17, 18, 19], with the most explicit models to date appearing in [20, 21]. Some other possibilities include [22, 23]. For recent reviews of inflation in string theory, see [24, 25, 26, 27].

This chapter is aimed at anyone who is intrigued by the possibility of connecting string theory and cosmology. We hope that it is accessible to non-string theorists, and have therefore tried hard to minimize string theory specific terminology and notation, referring the interested reader to more technical references for further detail. Table 3.1 provides a "Stringlish to English" reference dictionary for the most central string theory terms. Bridging the gap between string theory and observational astrophysics is important for both fields: not only does it offer potential tests of string theory as mentioned above, but it also offers an opportunity for cosmologists to move beyond the tradition of putting in inflaton potentials by hand.

3.1.1 Can String Theory Describe Inflation?

Answering the question of whether inflation can be embedded in string theory is very difficult. First of all, there is no known complete formulation of string theory, so the theory is not fully understood. In particular there does not seem to be a dynamical mechanism which selects the way in which the theory, which lives in 10 dimensions, should be compactified to 4 dimensions. Instead, there is apparently a "landscape" of possible 4-dimensional effective physics theories. Each so-called vacuum in the landscape corresponds to a stable or very long-lived configuration, containing, amongst other things, gravity, scalar fields (the above-mentioned moduli), and various potential energies. We would like to know if some of these vacua reproduce

Table 3.1: Dictionary of some basic string theory terminology.

Symbol	Name	Approximate meaning
α'	Regge parameter	Inverse string tension
l_s	String length	$= 2\pi\sqrt{\alpha'}$ (in our convention)
κ_{10}	10-d gravitational strength	$= \sqrt{8\pi G_{10}} = l_s^4/\sqrt{4\pi}$, gravitational strength in 10 dims
$\bar{m}_{\rm Pl}$	(Reduced) Planck mass	$= 1/\sqrt{8\pi G}$, mass scale of quantum gravity in 4 dims
ϕ	Dilaton	Scalar field that rescales the strength of gravity
a_i	Axions	Pseudo-scalars that appear in the 4-d theory
b_i	Geometric moduli	Scalar fields describing ϕ and the size & shape of the compact space
	– Dilaton modulus	$\sim e^{-\phi}$ (explicit form is model dependent)
	– Kähler moduli	Scalar fields that specify the *size* of the compact space
	– Complex structure moduli	Scalar fields that specify the *shape* of the compact space
ψ_i	Complex moduli	$= a_i + i\, b_i$
$\boldsymbol{\psi}$	Complex inflaton vector	$= (\psi_1, ..., \psi_n)$, the complex moduli-vector that can evolve during inflation
$\boldsymbol{\phi}$	Real inflaton vector	$= (a_1, b_1, ..., a_n, b_n)$, the real moduli-vector that can evolve during inflation
g_s	String coupling	$= e^{\phi}$, the string loop expansion parameter
F_p	p-form field strength	Generalized electromagnetic field strength carrying p-indices
f_p	Flux	$\propto \int F_p$, (normally integer valued) equivalent to a generalized electric or magnetic charge, but can arise purely due to non-trivial topology
g_{10}/R_{10}	10-d string metric/Ricci scalar	Metric/Ricci scalar in the fundamental 10-d action in string frame
g_4/R_4	4-d string metric/Ricci scalar	Metric/Ricci scalar in the effective 4-d action in string frame
g_E/R_E	4-d Einstein metric/Ricci scalar	Metric/Ricci scalar in the effective 4-d action after a conformal transformation to Einstein frame
g_6	Metric on compact space	2nd block of $g_{10} = \mathrm{diag}(g_4, g_6)$, describing the geometry of compact space
Vol	6-d volume of compact space	$= \int_{\rm cs} d^6x\sqrt{g_6}$ (cs $\equiv$ compact space)
T^6	6-d torus	A 6-d manifold that is Riemann flat, defined by periodic identifications
K	Kähler potential	Scalar function whose Hessian matrix is the metric on moduli space
W	Superpotential	Scalar function that describes the interactions between moduli set up by fluxes etc in a supersymmetric theory in 4 dims
V	Supergravity potential energy	Potential energy function governing the fields a_i, b_i in 4 dims as set up by the supergravity formula in eq. (3.11)
$\bar{V}$	Potential energy	$= \bar{m}_{\rm Pl}^4 V/4\pi$, potential energy function in 4 dims in conventional units
ϵ	First slow-roll parameter	See eq. (3.18), quantifies the magnitude of the 1st derivatives of V
η	Second slow-roll parameter	See eq. (3.19), quantifies the minimum of the 2nd derivatives of V
-	D(irichlet)p-brane	A $(p+1)$-d object that contributes positive energy and can source F_{p+2}
-	O(rientifold)p-plane	A $(p+1)$-d plane that can contribute negative energy; it arises at fixed points in so-called orientifold models

our observed large and rather uniform patch of 3+1-dimensional spacetime. We return below to the process by which this 4-dimensional picture emerges from the 10-dimensional picture. What is exciting is that these ingredients in 4 dimensions are precisely those used in inflationary model building.

There are several difficulties that must be overcome to build reasonable 4 dimensional models. Firstly, it is difficult to stabilize the moduli. One reason this is problematic is that one of the moduli corresponds to the size of the compact space. If it were not stabilized then the field may roll to very large values and the "compact space" would de-compactify. Furthermore, stabilizing all moduli is important to reproduce our universe which exhibits (approximate) Poincaré invariance, and a notable absence of long-range fifth forces. Authors have discussed various ingredients that may be included for such stabilization, *e.g.*, non-perturbative effects that go by names such as gluino condensation and instantons. A second difficulty arises due to supersymmetry; most well-understood vacua have negative cosmological constants, *i.e.*, correspond to anti de Sitter spaces (we will discuss this issue in some detail later on). One resolution of these problems was provided by KKLT [28], who included non-perturbative phenomena for stabilization and broke supersymmetry to achieve a 4-dimensional solution with positive cosmological constant. (Earlier constructions of de Sitter vacua in non-critical string theory appeared in [29]).

Within this framework (where non-perturbative corrections play an important role, and a supersymmetry breaking sector is incorporated to generate positive vacuum energy) various plausible models of inflation using moduli fields have been suggested in the literature: "N-flation" [30, 31], "Kähler moduli inflation" [32, 33, 34], "Inflating in a Better Racetrack" [35], and using brane moduli "KKLMMT" and related scenarios [19, 20, 21, 36]. These models, however, share a common property: they are not entirely explicit constructions, though steady progress in that direction has been made. This leads us to the obvious and important question: *Can we realize inflation explicitly and reliably in string theory?*

3.1.2 Explicit String Theory Inflation

One of the difficulties with making fully explicit models of inflation has been that most of the methods of moduli stabilization involve an interplay of classical effects in the potential (which are easily computable), and quantum effects whose existence is well established, but for which precise computations are often difficult.

However, recently, models which stabilize all moduli using classical effects alone have been constructed. These manage to stabilize all moduli in a regime where all approximations are parametrically under control [37, 38, 39, 40, 41]. These examples are all explicit stable compactifications. They primarily achieve this stability by using potential energy contributions from generalized electric and magnetic fields (so-called fluxes; see Table 3.1) whose combined energy are minimized when the moduli fields take some particular values. To borrow the language of quantum field theory, the potential functions in these models are generated at "tree-level", and quantum corrections are shown to be small. This stabilization of all moduli at tree-level is

what distinguishes these models from their earlier counterparts.[1]

In this chapter, we take three such recently found models and analyze each from the point of view of inflation. Specifically, we consider the models of DeWolfe, Giryavets, Kachru, & Taylor (DGKT) [39], Villadoro & Zwirner (VZ) [40], and Ihl & Wrase (IW) [41]. All of these arise in the string theory known as type IIA. Each of these models possesses an infinite number of vacua, distinguished by fluxes.

We wish to examine whether the tree-level potential for moduli fields in these models can support inflation. A well-known challenge for generating inflation within string theory is that generic potentials will not be sufficiently flat, a point we will expand on below. However, one might hope that when vast or infinite numbers of vacua are available, some of them would by chance have sufficiently flat directions to support inflationary slow-roll even if generic ones do not.[2] One of our key findings below is that in the case of the (simplest moduli in the simplest examples of) IIA flux vacua that we study here, the distributions of the quantities relevant for inflation are narrow enough that the large number of vacua does not help. Instead, the candidate inflaton potentials have the same shape in many (sometimes all) of the vacua, differing only in overall normalization. So, somewhat surprisingly, our search below does not turn up a single vacuum supporting inflation.

We hasten to emphasize that since these models are in many ways the simplest possible models in their class (involving the simplest compactification geometry, the six-torus, in a crucial way), and since we focus on a simple subset of the moduli (the "untwisted moduli") even in these models, our results should only be viewed as a first pass through this class of models. It is possible (but by no means certain) that more generic vacua in this class (based on compactification manifolds which have more complicated geometry, or based on studies of other moduli) would yield different results. More generally, flux potentials in other classes of vacua may well have broader distributions of the relevant physical quantities for inflation, allowing one to tune fluxes to achieve inflation.[3]

The rest of this chapter is organized as follows. Section 3.2 is a basic review of string compactification aimed at the non-specialist. Here we review the process of moving from the 10-dimensional theory to the 4-dimensional theory in fairly simple terms, showing how the 4-dimensional picture has the ingredients of inflation (gravity as well as kinetic and potential energies for scalar fields). We then show how the familiar slow-roll conditions for inflation become slightly generalized due to the non-

[1]It should be noted that the parametric control (which arises at very large values of fluxes) comes with various features which are undesirable for phenomenology: the extra dimensions become large at large flux values, the moduli masses become small, and the coupling constants become extremely weak. So for real world model-building, one would place a cutoff on the flux values, and lose parametric control. However, as simple and explicit examples of stable compactifications, these examples provide a useful setting to address theoretical questions, like our question about explicit computable models of moduli inflation.

[2]This of course depends on the extent to which the inflationary slow-roll parameters ϵ and η vary as the fluxes are changed: if they densely sample a wide range including $\epsilon < 1$, $|\eta| < 1$, then flux tuning can allow inflation, otherwise even large numbers of vacua may not help.

[3]Very concrete reasons to expect that the distributions are broader in IIB flux vacua are described in §6.2 of [39], for instance.

standard kinetic terms from string theory. In Section 3.3, we present and analyze the three explicit models analytically and numerically. We summarize our conclusions in Section 3.4.

3.2 String Theory and Dimensional Reduction

In this Section we give a gentle introduction and review of the study of compactification in string theory, with a focus on the ingredients that are relevant for the specific models we will investigate in Section 3.3. Much more complete and technical reviews are given in [5, 6, 7, 8, 9, 10], while a qualitative introductory review appears in [42]. We begin by mentioning the basic ingredients of type IIA string theory, with a focus on fluxes.

3.2.1 Supergravity

String theory is believed to be a consistent theory of quantum gravity. One curious feature of the theory is that this *consistency* suggests a special role for 10 dimensions, where consistent string theories are in correspondence with the so-called "maximal supergravities".[4] Furthermore, a remarkable feature of the theory is that its dynamics in 10 dimensions can be *derived*, rather than *guessed*, by demanding consistency. For comparison, consider the familiar case of a charged point particle moving in a background curved space-time $g_{\mu\nu}$ with a background electromagnetic field strength $F_{\mu\nu}$. There is no reasonable way to uniquely derive the dynamical equations governing the time-evolution of $g_{\mu\nu}$ and $F_{\mu\nu}$ from any consistency arguments about the behavior of the point particle. However, in the case of the string, this is precisely what happens.

In this chapter, we will focus on what is known as the type IIA string theory. It can be derived that part of the 10-dimensional action governing gravity and the field strength of the string in this theory is [43]

$$S = \frac{1}{2\kappa_{10}^2}\int d^{10}x\sqrt{-g_{10}}\,e^{-2\phi}\Big(R_{10} + 4(\partial_\mu\phi)^2 - \frac{1}{2}F_{\mu\nu\rho}F^{\mu\nu\rho}\Big). \tag{3.1}$$

Here R_{10} is the 10-dimensional Ricci scalar, ϕ is a scalar field known as the "dilaton", and $F_{\mu\nu\rho}$ is a generalized electromagnetic field strength; it carries 3 indices (making it a so-called 3-form) since it is sourced by the $(1+1)$-dimensional string, just like the familiar electromagnetic field strength $F_{\mu\nu}$ carries 2 indices (a 2-form) since it is sourced by a $(0+1)$-dimensional point particle. The overall pre-factor sets the gravitational strength in 10 dimensions $\kappa_{10}^2 = 8\pi G_{10}$. It is related to the so-called Regge-parameter α' (units of length-squared) by $2\kappa_{10}^2 = (2\pi)^7\alpha'^4$. The inverse of the Regge-parameter is the string tension $(= 1/2\pi\alpha')$; the tension of a string is an absolute constant and is analogous to the mass of a particle. Furthermore, α' is related to the string length by (in our convention) $l_s = 2\pi\sqrt{\alpha'}$. Later we will see

[4] Here we are referring to the $\mathcal{N}=1$ and $\mathcal{N}=2$ theories in 10-dimensions, where $\mathcal{N}$ is the number of supersymmetries.

that it is convenient to measure a number of dimensionful quantities in units of l_s. Table 3.1 provides a hopefully useful dictionary of key string theory notation and the symbols used in this chapter, including a summary of the above.

Let us summarize: In eq. (3.1) we see that the 10-dimensional universe of string theory contains gravity and a field strength $F_{\mu\nu\rho}$, and that they appear in the same way as gravity and electromagnetism do in 4 dimensions. We further note that there exists the dilaton ϕ which is non-minimally coupled to gravity. Because the coefficient of $F_{\mu\nu\rho}F^{\mu\nu\rho}$ is proportional to $e^{-2\phi}$, one identifies $g_s \equiv e^{\phi}$ as the string coupling; it is the string loop expansion parameter analogous to e in electromagnetism.

Now we must mention some other features of string theory in 10 dimensions that we did not include in eq. (3.1). First of all, it turns out that the 3-form $F_{\mu\nu\rho}$ (which we will later denote simply F_3, and which is often denoted H_3 in the string literature) is *not* the only field strength that appears in string theory. Rather, there are also various other fields; various so-called p-forms with p indices, where p takes various integer values, and whose interactions are also uniquely determined by consistency. In addition, there are extended objects of various dimensionality in the theory known as Dirichlet branes and orientifold planes, which are charged under these p-forms [44]. Also, there are fermions which give rise to a collection of terms to be added to eq. (3.1), since we are describing a supersymmetric theory, but we have set their values to zero here. We focus on cosmologies that have maximal space-time symmetry (Minkowski, anti de Sitter, de Sitter) which means that the vacuum expectation values of the fermion fields must vanish. Finally, what appears in eq. (3.1) is only the first term in a perturbative expansion in powers of α' and g_s. For length scales large compared to the string length l_s and for small g_s we can ignore such corrections; this is known as the supergravity approximation.

3.2.2 Compactification and Fluxes

Calabi-Yau manifolds

Of most interest to us is what this theory predicts in 4 dimensions. Currently there is no background independent formulation of string theory, so the *compactification* of the 10-dimensional geometry to 4 large dimensions is specified by hand and is not unique. The most commonly studied compact spaces are Calabi-Yau manifolds (e.g., see [7, 43] for a technical definition). They are useful for at least two reasons: Firstly, Calabi-Yau manifolds preserve some unbroken supersymmetry which allows for better computational control. Secondly, and most importantly for us, Calabi-Yau manifolds are spaces that possess a metric that is Ricci flat (has $R_{\mu\nu} = 0$ like any vacuum metric). This is very convenient in finding solutions to the 10-dimensional equations of motion. The simplest example is that of the torus T^6, which is not only Ricci flat, but also flat (with vanishing Riemann tensor). In this chapter we will focus on this space (T^6), since this has been studied the most intensely in the literature. It is also a very useful pedagogical device, and we will make some comments below on the connection of our results to more general compactifications.

Orbifolds and Orientifolds

Although understanding them in detail is not central to following our examples below, let us briefly mention orbifolding [45] and orientifolding [46, 47, 48], two technical operations that string theorists perform on the compact space, since they occur in all the models we will investigate. It often proves to be important to reduce the number of points in a manifold by declaring some of them identical. In technical jargon, one forms the quotient space with some finite symmetry group of the manifold, for example $T^6/\mathbb{Z}_p$, where $\mathbb{Z}_p$ is the group of integers modulo p. The specific $\mathbb{Z}_p$ symmetry is model dependent. This defines a so-called orbifold. Certain toroidal orbifolds are of interest since they are a special singular limit of some (non-toroidal) Calabi-Yau.[5] Also, by performing additional discrete operations one can form what is known as an orientifold. It is related to forming unoriented strings out of oriented strings. It will be important for us in what follows that at fixed points of the group action on the internal manifold, in orientifold models, one gets so-called *orientifold planes* (or O-planes). These O-planes provide a negative contribution to the vacuum energy (see ahead to eq. (3.13)) and this is important for stabilization.

Both of these operations, orbifolding and orientifolding, serve the purpose of allowing for chiral fermions and reducing the amount of supersymmetry in 4 dimensions. For suitable choices of symmetry groups, there can however be a residual amount of supersymmetry in 4 dimensions.[6]

Moduli

In general, there are scalar fields characterizing the size and shape of any compact manifold: Kähler moduli (roughly specifying size) and complex structure moduli (roughly specifying shape). Table 3.1 summarizes all fields whose dynamics we will keep track of in 4 dimensions: besides gravity $g_{\mu\nu}$, we have "geometric" moduli: a dilaton ϕ, Kähler moduli (also known as "radions"), and complex structure moduli. In addition, each of these geometric moduli are accompanied by a field that is generically referred to as an "axion". The reason these are called axions is not important here, but suffice it to say that they are all pseudo-scalars and some are coupled to a generalized $\mathbf{E}\cdot\mathbf{B}$ term in the action, reminiscent of the axion of quantum chromodynamics (see, e.g., [49, 50]). If we denote the various geometric moduli by b_i ($i = 1, 2, ...$) and the axions by a_i, then these two degrees of freedom can be put into a complex pair $\psi_i = a_i + i\, b_i$.[7] We will see that this construction of forming a complex scalar is quite useful.[8] We will group all these complex fields into a single vector $\boldsymbol{\psi}$, which will act as our complex inflaton vector. When separated into its real components, we denote this $\boldsymbol{\phi}$; our real inflaton vector.

[5] E.g., the orbifold $T^4/\mathbb{Z}_2$ is a limit of the Calabi-Yau K3.

[6] For example, in commonly studied orientifolds of Calabi-Yau manifolds, the $\mathcal{N} = 2$ theory in 10 dimensions becomes an $\mathcal{N} = 1$ theory in 4 dimensions in type II string theory.

[7] In the string literature, there are many symbols used for the different moduli, such as T, U, v, etc, but we will just use the common notation $\psi_i = a_i + i\, b_i$ for all moduli.

[8] In fact this construction is integral to $\mathcal{N} = 1$ supersymmetric models, where such pairs are unified in a chiral multiplet, a representation of the supersymmetry algebra.

Fluxes and Potential Energies

Strictly speaking, moduli are defined as those scalars that have vanishing potential. Without including any extra ingredients (such as field strengths), the above mentioned fields would indeed be massless and free. This is very problematic. For example, if the radions are freely propagating fields then the size of the compact space could take on any value, including unacceptably large values. Indeed, there are constraints from 5-th force experiments showing that these fields must be stabilized with a large effective mass (moderately large compared to the inverse millimeter scale to which gravity has been tested, and huge compared to today's Hubble scale), *i.e.*, that there must be contributions to the potential energy density of the form $m_i^2 b_i^2$ (where the coefficients m_i are large). Furthermore, we are interested in whether any of these scalars could be the inflaton. Since a free field by definition is one that does not feel a potential, it cannot possibly drive inflation.

However, an important feature of string theory is the existence of various field strengths, and these induce interactions for the moduli. We have already introduced the field strength $F_{\mu\nu\rho}$, hereafter abbreviated F_3. We will focus on what is known as type IIA string theory in this chapter, in which there are also forms with even numbers of indices, such as F_2 and F_4, but more general forms F_p occur in other models. In our 3+1 large dimensions, Lorentz invariance prevents any cosmological field strengths, however such restrictions do not apply to the components of F_p in the compact space.[9] Assuming $p \leq 6$, then the fields satisfy

$$\frac{1}{l_s^{p-1}} \int F_p = f_p, \tag{3.2}$$

where the integral is over some p-dimensional internal manifold of the compact space. Such integrals appear when we compactify the theory. Here f_p is an integer, corresponding to a generalized Dirac charge quantization condition. These quantized integrals of the field strengths are known as 'fluxes'. They correspond to wrapped field lines in the compact space. Such fields can be thought of as being sourced by generalized electric and magnetic charges provided by the various branes of the theory [7]. Note, however, that this is just an incomplete analogy, since the fluxes we are referring to here thread topologically non-trivial internal submanifolds of the compact space; therefore, Gauss' law does not require charges to source the flux. (There are, however, other space-filling branes in the theory, which will appear in the models). Since there is an energy cost associated with deformations of the compact space in the presence of field strengths, these fluxes induce a potential energy $V = V(\psi)$ for the moduli. This potential is necessary for stabilization, and we will investigate if it can also drive inflation.

To get an idea of the form that these energies will take, consider a generalized

[9] In the presence of field strengths in the compact space, the Ricci tensor on the compact space is in general non-zero, and so the space is strictly no longer Calabi-Yau. However, when the moduli masses are light compared to the inverse size of the compact space (the "Kaluza Klein scale") then this *back-reaction* is small, and we can continue to treat the compact space as Calabi-Yau. This is a property of the models we study.

electric field $\mathbf{E}_p$ set up by a stationary source; a point source for $\mathbf{E}_2$, a string for $\mathbf{E}_3$, a membrane for $\mathbf{E}_4$ etc. If there are f_p units of charge contained in a compact space of size r, then the energy density is given roughly by (ignoring factors of l_s)

$$|\mathbf{E}_p|^2 \sim \frac{f_p^2}{r^{2p}}, \tag{3.3}$$

which reduces to the familiar result f_2^2/r^4 for a point source. The total energy density will involve a sum of such terms. eq. (3.12) below illustrates the form this takes when expressed in a 4-dimensional action.

3.2.3 The 4-Dimensional Action and Slow-Roll

Here our aim is to move from the 10-dimensional theory to the 4-dimensional theory.

Integrating Out the Compact Space

To understand the 4-dimensional action, let us begin by focusing on the gravity sector. For simplicity, we will assume that the 10-dimensional metric is in block diagonal form: $g^{10}_{AB} = \mathrm{diag}(g^4_{\mu\nu}, g^6_{ab})$, i.e., that it separates into a piece governing the 4 large space-time dimensions and a piece governing the 6 compact dimensions. Furthermore, let us assume that the Ricci scalar is independent of the compact coordinates (the usual assumption in Kaluza-Klein models), so that we can integrate over d^6x. This assumption is valid when one considers an (approximately) "unwarped compactification," as we do in this chapter. The relevant piece in the action is the first term in eq. (3.1), involving the 10-dimensional Ricci scalar R_{10} and the dilaton (scalar field). This gives

$$\int d^{10}x\sqrt{-g_{10}}\, e^{-2\phi} R_{10} = \int d^4x\sqrt{-g_4}\,\mathrm{Vol}\, e^{-2\phi} R_{10} \tag{3.4}$$

where Vol is the 6-dimensional volume of the compact space. The particular form of Vol is model dependent, and the relationship between the 10-dimensional Ricci scalar R_{10} and the 4-dimensional Ricci scalar R_4 is also model dependent through the form of the metric on the compact space. However, for this class of models it is true that

$$R_{10} = R_4 + \mathrm{func}(\text{compact space fields}) \tag{3.5}$$

Since both the volume Vol and the dilaton ϕ are allowed to be dynamical, the action in eq. (3.4) is evidently not in canonical form. In order to bring the action into canonical form, we introduce the so-called Einstein metric $g^E_{\mu\nu}$, which is defined via a conformal transformation as

$$g^E_{\mu\nu} \equiv \frac{\mathrm{Vol}\, e^{-2\phi}}{\bar{m}^2_{\mathrm{Pl}}\,\kappa^2_{10}} g^4_{\mu\nu} \tag{3.6}$$

where $g^4_{\mu\nu}$ is the 4-dimensional string metric – the metric in the "string frame" of eq. (3.1). In this transformation we have introduced the (reduced) Planck mass

$\bar{m}_{\rm Pl} = 1/\sqrt{8\pi G} \approx 2 \times 10^{18}\,\text{GeV}$, where G is the 4-dimensional Newton constant. The gravity sector, written in terms of the corresponding Einstein Ricci scalar R_E, then appears in canonical form

$$\frac{1}{2\kappa_{10}^2}\int d^{10}x\sqrt{-g_{10}}e^{-2\phi}R_{10} = \int d^4x\sqrt{-g_E}\left(\frac{1}{16\pi G}R_E + \cdots\right) \tag{3.7}$$

and this is referred to as the "Einstein frame".

Let us clarify a feature of the conformal transformation. Recall that during inflation, both Vol and ϕ may be dynamical, since they in fact depend on the inflaton vector ψ. Well after inflation, we expect these fields to be stabilized at some fixed values: $\langle\text{Vol}\rangle$ and $\langle\phi\rangle$. At such values we require that the conformal transformation in eq. (3.6) be the identity transformation. This implies a relationship between κ_{10}, $\langle\text{Vol}\rangle$, $\langle\phi\rangle$, and $\bar{m}_{\rm Pl}$:

$$\bar{m}_{\rm Pl}^2 = \frac{\langle\text{Vol}\rangle\, e^{-2\langle\phi\rangle}}{\kappa_{10}^2}. \tag{3.8}$$

Note that the Planck mass is defined in terms of the fields at their *stabilized values*, so it is a constant. Since it is natural for $\langle\text{Vol}\rangle$ to be given in units of l_s^6 and since $\kappa_{10}^2 = l_s^8/4\pi$, one can in fact use this equation to determine the string length l_s in terms of the Planck length for any particular model.

Kinetic Energy

In general, the kinetic energy of the moduli is not in canonical form. Recall that the moduli are a combination of not only the dilaton, but also the Kähler and complex structure moduli which describe the size and shape of the particular Calabi-Yau compact space. An important property of every Calabi-Yau (related to the underlying supersymmetry it preserves) is that there exists a scalar function K of the moduli, known as the Kähler potential, whose Hessian matrix is the metric on moduli space:

$$K_{i\bar{j}} \equiv K_{,i\bar{j}} = \frac{\partial^2 K}{\partial\psi^i\partial\psi^{\bar{j}}}, \tag{3.9}$$

where ψ^i is a generic name for a complex modulus (axion partnered with geometric moduli), and barred variables denote the complex conjugate. Contracting space-time derivates of the moduli with this metric gives the kinetic energy in the Einstein frame:

$$T = -\bar{m}_{\rm Pl}^2 K_{i\bar{j}}\partial_\mu\psi^i\partial^\mu\psi^{\bar{j}}. \tag{3.10}$$

There is an implicit sum over all i, j. Since we shall deal with dimensionless Kähler potentials, the factor of $\bar{m}_{\rm Pl}^2$ is necessary on dimensional grounds. In the limit in which we shall work, K does *not* depend on fluxes, which means that there is one Kähler potential for each of the three models that we will investigate, applicable to any vacuum in their respective landscapes. In other words, fluxes affect only the potential energy, not the kinetic energy.

Potential Energy

In addition to the Kähler potential K, there exists another object which contains information about the particular compactification. For supersymmetric compactifications, of the type we will focus on, this object is the so-called superpotential W. W is a complex analytic function of the complex moduli ψ^i and also depends on the fluxes. If we turn on fluxes, then there are in general energies induced associated with distortions of the compact space and displacements of the dilaton and axions. These interactions are all contained in W. The potential energy term in the 4 dimensional Lagrangian density is given by

$$V = e^K \left(D_i W K^{i\bar{j}} \overline{D_j W} - 3|W|^2 \right), \tag{3.11}$$

which is sometimes referred to as the supergravity formula. Here $D_i W \equiv \partial_i W + W \partial_i K$ and $K^{i\bar{j}}$ is the matrix inverse of $K_{i\bar{j}}$. Again there is an implicit sum over all i, j. A reader familiar with supersymmetry in 4 dimensions will recognize that these are the so-called "F terms" and that there are no so-called "D terms".

In order to develop some intuition, we would like to get an idea of the typical form of V. In eq. (3.3) we estimated the energy density established by a p-form field strength, but must now take into account the integration over the compact space and the conformal transformation. For a compact space of size r, the contribution to V is given roughly by (ignoring factors of l_s)

$$\begin{aligned} \Delta V &\sim f_3^2 \frac{e^{2\phi}}{r^{12}} \quad \text{for } F_3, \\ \Delta V &\sim f_p^2 \frac{e^{4\phi}}{r^{6+2p}} \quad \text{for } F_{p\neq 3}. \end{aligned} \tag{3.12}$$

We also mention an estimate for the contribution to V from N_1 D6-branes and N_2 O6-planes

$$\begin{aligned} \Delta V &\sim N_1 \frac{e^{3\phi}}{r^9} \quad \text{for D6-branes}, \\ \Delta V &\sim -N_2 \frac{e^{3\phi}}{r^9} \quad \text{for O6-planes}. \end{aligned} \tag{3.13}$$

Note that the O6-plane makes a negative contribution.

In the models studied, when calculating the potential from the supergravity formula V, we will for simplicity work in units where the string length $l_s = 1$. This must be rescaled in order to obtain the potential in Einstein frame $\bar{V}$ in conventional units:

$$\bar{V} = \frac{\bar{m}_{\rm Pl}^4}{4\pi} V. \tag{3.14}$$

(This comes from: $\bar{m}_{\rm Pl}^4 \kappa_{10}^2 = \bar{m}_{\rm Pl}^4 l_s^8 / 4\pi = \bar{m}_{\rm Pl}^4 / 4\pi$.) In Sections 3.3.2 – 3.3.5 we will just refer to V, rather than $\bar{V}$.

Putting it All Together

Altogether, the effective 4-dimensional action in the Einstein frame is a familiar sum of a gravity term, kinetic energy, and potential energy, i.e.,

$$S = \int d^4x\sqrt{-g_E}\left[\frac{1}{16\pi G}R_E - \bar{m}_{\rm Pl}^2 K_{i\bar{j}}\partial_\mu\psi^i\partial^\mu\psi^{\bar{j}} - \bar{V}(\boldsymbol{\psi})\right]. \tag{3.15}$$

Keeping the kinetic term general, the Euler-Lagrange equations of motion for a flat Friedmann-Robertson-Walker (FRW) universe $g^E_{\mu\nu} = \mathrm{diag}(-1, a(t)^2, a(t)^2, a(t)^2)$ are in terms of K and its derivatives:

$$\ddot{\psi}^i + 3H\dot{\psi}^i + \Gamma^i_{jk}\dot{\psi}^j\dot{\psi}^k + K^{i\bar{j}}\bar{V}_{,\bar{j}}/\bar{m}_{\rm Pl}^2 = 0, \tag{3.16}$$

$$H^2 = \frac{8\pi G}{3}\left[\bar{m}_{\rm Pl}^2 K_{i\bar{j}}\dot{\psi}^i\dot{\psi}^{\bar{j}} + \bar{V}(\boldsymbol{\psi})\right], \tag{3.17}$$

where $H = \dot{a}/a$ is the Hubble parameter and $\Gamma^i_{jk} = K^{i\bar{n}}K_{jk,\bar{n}}$ is the Christoffel symbol on moduli space.

To investigate inflation, we must compute the slow-roll parameters. The first slow-roll condition is that the kinetic energy in (3.17) should be small compared to the potential energy when the acceleration term in (3.16) is negligible. The second condition is that the acceleration is and remains small along the slow-roll direction (quantified by differentiating and demanding self-consistency). These two conditions define the two slow-roll parameters ϵ and η and correspond to the requirements that $\epsilon < 1$ and $|\eta| < 1$. We find that

$$\epsilon = \frac{K^{i\bar{j}}V_{,i}V_{,\bar{j}}}{V^2}\left(= \frac{g^{ab}V_{,a}V_{,b}}{2V^2}\right), \tag{3.18}$$

$$\eta = \text{min eigenvalue}\left\{\frac{g^{ab}\left(V_{,bc} - \Gamma^d_{bc}V_{,d}\right)}{V}\right\}. \tag{3.19}$$

Here η (and ϵ) is written in terms of the metric g_{ab} governing real scalar fields ϕ^a: $K_{i\bar{j}}\partial_\mu\psi^i\partial^\mu\psi^{\bar{j}} = \frac{1}{2}g_{ab}\partial_\mu\phi^a\partial^\mu\phi^b$, with $\phi^{2i-1} = a^i$ and $\phi^{2i} = b^i$. Note that we can use either V or $\bar{V}$ in the slow roll conditions, as $V \propto \bar{V}$ (see eq. (3.14)). In regions where inflation occurs, these three functions (V, ϵ, η) can be used to predict several cosmological parameters, as detailed in Appendix D. A comparison between the theoretical predictions and observational data provides a precision test of the model.

3.2.4 de Sitter Vacua

There are two good reasons to want de Sitter "vacua", namely that there are at least two eras of our universe that are approximately de Sitter; during inflation, which exhibits slow-roll, and at late times, which appears consistent with a positive cosmological constant. This chapter is focussed on the former epoch. If we have a region in moduli space that is de Sitter, namely a region in which the gradient of the

4-dimensional potential V (from eq. (3.15)) is zero with $\bar{V} > 0$, we are a significant step closer to realizing inflation. In such a region (which may be a single point) we at least know that the first slow-roll parameter $\epsilon = 0$, although we may still face the so-called η-problem [51] (for a recent discussion in the string theory context, see [52]).

Let us make some comments about supersymmetric vacua and anti de Sitter (AdS) space. The condition for supersymmetry (SUSY) is that the covariant derivative of the superpotential vanishes, i.e.,

$$D_i W = \partial_i W + W \partial_i K = 0 \tag{3.20}$$

for all i. It is simple to show from eq. (3.11) that at any such SUSY point the supergravity potential V is stationary. This constitutes some "vacuum" (stable or unstable) of the theory. At such a point the potential is

$$V_{\text{SUSY}} = -3e^K |W|^2, \tag{3.21}$$

so we see it is necessarily non-positive. It may happen that $W = 0$, corresponding then to Minkowski space. But what is *much more common* is for $W \neq 0$, corresponding then to anti-de Sitter space.

In fact it is known that any vacuum that is supersymmetric (in supergravity, superstring theory, M-theory, etc) is necessarily non de Sitter. But it may be the case that in such a model, a non-SUSY minimum in the space of scalar field expectation values is de Sitter (spontaneously broken SUSY). Achieving this is not easy, as described by various 'no-go theorems'. In particular, under mild assumptions on the nature of the compact space (namely that it is non-singular etc.), one can show that inclusion of fluxes alone does not allow one to find any de Sitter vacua [53, 54].

There are, however, other structures besides fluxes in string theory, e..g, D-branes and O-planes. In [55] it is shown that the argument of [54] may be extended to include most forms of D-branes, but it cannot be extended to include O-planes. The energies associated with such structures were described in eqn. (3.13). Recent work has shown that the realization of de Sitter vacua is possible with such ingredients, as in the constructions of [29]. However, we do not find any de Sitter vacua in the simple IIA models we study.

Given that we are considering models with AdS vacua, what is the implication for inflation? Suppose that the inflaton eventually settles down to some such AdS vacuum. There the potential has a (negative) value that we will call the 'cosmological constant'. This may well compromise any chance to obtain many e-folds of inflation, which requires $\bar{V} > 0$. However, we can imagine a priori a scenario where this is *not* catastrophic for inflation: We will see later that we have fluxes that can be used to dial the cosmological constant toward zero. Hence the depth of AdS space can be tuned very small. Furthermore, well away from the SUSY vacuum there are regions in moduli space were the potential is large and *positive*. Then, as long as $\bar{V}$ during inflation is much greater than the depth of the AdS minimum, it is plausible that inflation could be realized.

At the end of inflation one should in principle enter the radiation era. Normally

this occurs through the decay of the inflaton to various fields including the standard model particles. However, the standard model is not contained in the models that we investigate, so this is an issue that we do not tackle. Furthermore, we do not address the late-time problem of the smallness of the (positive) cosmological constant. (A popular explanation of the smallness of the cosmological constant appeals to the existence of exponentially many vacua realizing different vacuum energies, see [56]).

3.3 Type IIA Models

Here we investigate the cosmology of three explicit models. Many choices of compactification are possible. However, the torus is flat and is perhaps the most well studied compact manifold in the literature, so we will focus on (orbifolds and orientifolds of) this. We will investigate the resulting inflation picture for three explicit models: DGKT [39], VZ [40], and IW [41]. In this Section we use units $l_s = 1$.

3.3.1 Diagonal Torus Models

For clarity, let us describe the properties of the Kähler potential and the slow-roll conditions in more detail for a particular class of examples. The first two torus models to be discussed have the property that the Kähler potential is the logarithm of a product of moduli. Writing

$$\psi_i = a_i + ib_i \tag{3.22}$$

for all moduli[10], we have

$$K = -\ln\left(\prod_i b_i^{n_i}\right) + \text{const} \tag{3.23}$$

where n_i are $\mathcal{O}(1)$ integers (e.g., in the VZ model described below, there are 7 moduli ($i = 1, \ldots, 7$) with $n_i = 1$). The kinetic energy for such models is then

$$T = -\bar{m}_{\rm Pl}^2 \sum_i \frac{n_i}{4} \frac{(\partial_\mu a_i)^2 + (\partial_\mu b_i)^2}{b_i^2} \tag{3.24}$$

In this case, the equations of motion for the moduli are

$$0 = \ddot{b}_i + 3H\dot{b}_i + \frac{\dot{a}_i^2 - \dot{b}_i^2}{b_i} + \frac{1}{\bar{m}_{\rm Pl}^2} \frac{2b_i^2}{n_i} \frac{\partial \bar{V}}{\partial b_i}, \tag{3.25}$$

$$0 = \ddot{a}_i + 3H\dot{a}_i - 2\frac{\dot{a}_i \dot{b}_i}{b_i} + \frac{1}{\bar{m}_{\rm Pl}^2} \frac{2b_i^2}{n_i} \frac{\partial \bar{V}}{\partial a_i}. \tag{3.26}$$

[10]More precisely, we study all moduli that arise from metric deformations of the torus, the dilaton, and their superpartners. We neglect so-called "twisted moduli" or "blow-up modes" originating from singularities of the orbifold group action, though we briefly discuss them in Section 3.3.5.

The first slow-roll parameter for inflation then takes the form

$$\epsilon = \frac{1}{V^2}\sum_i \frac{b_i^2}{n_i}\left[\left(\frac{\partial V}{\partial a_i}\right)^2 + \left(\frac{\partial V}{\partial b_i}\right)^2\right]. \tag{3.27}$$

It is important to take note of the form of the Kähler potential; it is independent of all axions (general fact). Hence if we shift an axion by a constant, it has no effect on the kinetic energy. Also, if we rescale ψ_i by a real number, the kinetic energy is also unchanged. In short,

$$a_i \to d_i a_i + c_i, \quad b_i \to d_i b_i \tag{3.28}$$

leaves the kinetic energy unchanged for any constants $c_i, d_i \in \mathbb{R}$. In turn, the form of the slow-roll parameters ϵ and η are unaffected. These shift and scaling symmetries allow one to eliminate some flux parameters that appear in the superpotential. In the first model, we will see that these symmetries allow all fluxes to be absorbed into an overall multiplicative factor, while in the second and third models we will have one additional non-trivial flux parameter to dial. There will in general be ambiguities associated with positive/negative values of the fluxes that one should keep careful track of.

In order to emphasize that the Lagrangian in 4 dimensions is reminiscent of that in standard inflation, let us perform a field redefinition for the simple case where we ignore the axions and focus on $\bar{V}(b_i)$. By defining

$$\phi_i \equiv \sqrt{\frac{n_i}{2}}\,\bar{m}_{\rm Pl}\log b_i, \tag{3.29}$$

the kinetic energy is put in canonical form and the action in eq. (3.15) becomes

$$S = \int d^4x\sqrt{-g_E}\left[\frac{1}{16\pi G}R_E - \sum_i \frac{1}{2}(\partial_\mu\phi_i)^2 - \bar{V}\left(e^{\frac{\sqrt{2}\,\phi_i}{\sqrt{n_i}\,\bar{m}_{\rm Pl}}}\right)\right]. \tag{3.30}$$

Note that the argument of $\bar{V}$ is now an exponential. The first slow-roll parameter then takes the canonical form for multi-field inflation:

$$\epsilon = \frac{\bar{m}_{\rm Pl}^2}{2}\frac{|\nabla_\phi V|^2}{V^2}. \tag{3.31}$$

We point out that this is only true when ignoring the axions and relies upon the assumed simple form of the Kähler potential.

3.3.2 The Model of DeWolfe, Giryavets, Kachru, and Taylor (DGKT)

In May 2005, DeWolfe et.al [39] (DGKT) found an explicit *infinite* class of stable vacua in type IIA string theory. In their model they found that all moduli are stabilized

by including the 3-form, 4-form (and less importantly the 2-form) fluxes of type IIA, and also including a 0-form flux. The 0-form flux plays the role of a mass-term in the theory. Its presence induces several extra pieces into the action, which can only be derived from so-called M-theory. This framework is known as "massive type IIA supergravity".

Starting from the torus, they built the orbifold $T^6/\mathbb{Z}_3$ and projected it to the orientifold $T^6/\mathbb{Z}_3^2$. In addition, they introduced a static $(6+1)$-dimensional plane that carries charge (an O6-plane), for the purpose of satisfying a constraint known as a tadpole condition. We will focus here on reporting the salient features of the geometry after these technical operations have been performed; the reader is referred to the original paper [39] for details.

The torus of DGKT takes on essentially the simplest possible form (we will see more complexity in the later models of VZ and IW). The orbifolding and orientifolding act to reduce the number of degrees of freedom of the metric on the compact space to just 3. Here we neglect the moduli which arise at the orbifold fixed points (the so-called "blow up modes" or "twisted sector" moduli).[11] These 3 are the (untwisted) Kähler moduli of the theory. There are no complex structure structure moduli left. The metric on the compact space and the volume are given by:

$$(ds^2)_6 = \sum_{i=1}^{3} \gamma_i \left[(dx^i)^2 + (dy^i)^2\right], \tag{3.32}$$

$$\text{Vol} = \int_{T^6/\mathbb{Z}_3^2} d^6x\sqrt{g_6} = \frac{1}{8\sqrt{3}}\gamma_1\gamma_2\gamma_3 = b_1b_2b_3. \tag{3.33}$$

Here the elements of the metric are called γ_i. The volume is proportional to the determinant of the square root of the metric $(\gamma_1\gamma_2\gamma_3)$, the factor of $1/8\sqrt{3}$ comes from performing the peculiar integration over $T^6/\mathbb{Z}_3^2$, but is not important for us. What is important is the identification of the good Kähler co-ordinates b_1, b_2, b_3 whose product is the volume[12] (up to a prefactor, they are just the components of the metric).

Let us summarize the moduli of this model. As mentioned, all complex structure moduli are projected out by the orientifolding, leaving 4 moduli: 3 Kähler moduli $\psi_i = a_i + ib_i$, $i = 1,2,3$ and an axio-dilaton $\psi_4 = a_4 + ib_4$ ($b_4 = e^{-\phi}\sqrt{\text{Vol}}/\sqrt{2}$).[13] We note that the axio-dilaton appears in the compactification, not through an explicit appearance in the compact metric, but through its direct appearance in the action, as discussed in Section 3.2.

The Kähler potential takes on the form promised in Section 3.3.1, namely the logarithm of the product of geometric moduli. All that is left is to specify the values

[11] We will briefly discuss the proper inclusion of the blow-up modes in 3.3.5. They alter the discussion in various important ways, but do not at first sight seem to change our conclusions.

[12] In the DGKT paper: $\text{Vol} = \kappa b_1b_2b_3$, with $\kappa = 81$. By rescaling $b_i \to \kappa^{-1/3}b_i$, $i = 1,2,3$ we obtain eq. (3.33) and κ is eliminated.

[13] The canonical model-independent axion is $\xi = 2a_4$.

of n_i and the constant. One finds that

$$K = -\ln\left(32\, b_1\, b_2\, b_3\, b_4^4\right). \tag{3.34}$$

The superpotential is set by the interactions: DGKT turned on fluxes coming from F_3, F_2, F_4, and a zero form F_0, a so-called mass term, as well as an F_6. By studying the work of Grimm and Louis [37] they find

$$W = \frac{f_6}{\sqrt{2}} + \sum_{i=1}^{3} \frac{f_{4,i}}{\sqrt{2}}\,\psi_i - \frac{f_0}{\sqrt{2}}\,\psi_1\,\psi_2\,\psi_3 - 2\,f_3\,\psi_4, \tag{3.35}$$

the flux integers f_6, $f_{4,i}$, f_0, f_3 arising from F_6, F_4, F_0, F_3, respectively. We have turned F_2 off, as all results are qualitatively similar, although it is simple to include.

As mentioned, DGKT are able to satisfy the "tadpole condition" by including an O6-plane. In order to so, the following relationship between two of the flux integers must hold:

$$f_0\, f_3 = -2. \tag{3.36}$$

At this point there are several flux integers in the problem. However, we can simplify the problem greatly by exploiting the shift and scaling symmetries that we discussed in Section 3.3.1. Let us perform the following transformations of our fields:

$$\begin{aligned}
\psi_i &\to \frac{1}{|f_{4,i}|}\sqrt{\frac{|f_{4,1}\, f_{4,2}\, f_{4,3}|}{|f_0|}}\,\psi_i \quad (i = 1,2,3), \\
\psi_4 &\to \frac{1}{|f_3|}\sqrt{\frac{|f_{4,1}\, f_{4,2}\, f_{4,3}|}{|f_0|}}\,\psi_4 + \frac{1}{2\sqrt{2}}\frac{f_6}{f_3}
\end{aligned} \tag{3.37}$$

which leaves the form of the kinetic terms invariant. In terms of these new variables, the superpotential becomes

$$W = \sqrt{\frac{|f_{4,1}\, f_{4,2}\, f_{4,3}|}{|f_0|}}\left(\sum_{i=1}^{3} \frac{\hat{f}_{4,i}}{\sqrt{2}}\,\psi_i - \frac{\hat{f}_0}{\sqrt{2}}\,\psi_1\,\psi_2\,\psi_3 - 2\,\hat{f}_3\,\psi_4\right), \tag{3.38}$$

where the 'hat' fluxes are just the signs of the fluxes, e.g., $\hat{f}_0 \equiv f_0/|f_0|$. This has a very interesting form: apart from an overall multiplicative factor, the superpotential is *independent of the magnitude of the fluxes* (although their sign will be important).

We now have all the tools we need; the Kähler potential and the superpotential. Using these, we can compute the 4-dimensional potential V using the supergravity formula (3.11). Focusing on the symmetric case, i.e., $\psi_1 = \psi_2 = \psi_3$, we find

$$\begin{aligned}
V \;=\; & V_{\text{flux}}\big[2(3\,a_1 + 2\sqrt{2}\,a_4)^2 - 4\,\delta\,a_1^3(3a_1 + 2\sqrt{2}\,a_4) \\
& +\; 2\,a_1^6 + 6\,b_1^2 + 4\,b_4^2 - 12\,\delta\,a_1^2\,b_1^2 + 6\,a_1^4\,b_1^2 + 6\,a_1^2\,b_1^4 \\
& +\; 2\,b_1^6 - 8\sqrt{2}\,b_1^3 b_4\big]/(32\,b_1^3\,b_4^4),
\end{aligned} \tag{3.39}$$

where

$$V_{\text{flux}} \equiv \frac{|f_0|^{5/2}|f_3|^4}{|f_{4,1}f_{4,2}f_{4,3}|^{3/2}} \tag{3.40}$$

is an overall multiplicative scale that depends on the fluxes. Note that since f_0 and f_3 are tightly constrained by the tadpole condition (3.36), V_{flux} is bounded from above and approaches 0 as $f_{4,1}f_{4,2}f_{4,3} \to \infty$. Also, $\delta \equiv \hat{f}_0\hat{f}_{4,1}\hat{f}_{4,2}\hat{f}_{4,3} = \pm 1$, delineates two independent families of V. The more general result, without simplifying to the symmetrical case, is given in Appendix A eq. (3.80).

Here we make a parenthetical comment: One can perform direct dimensional reduction from the 10-dimensional action without using the Kähler potential or superpotential. In the DGKT paper this is done explicitly with the axions (a_i, a_4) set to their SUSY values. Having set the axions to their SUSY values, the natural (and perhaps most intuitive) co-ordinates are then the original fields: dilaton ϕ and radions b_i. They find

$$V = \frac{f_3^2}{4}\frac{e^{2\phi}}{\text{Vol}^2} + \frac{1}{4}\left(\sum_{i=1}^{3} f_{4,i}^2 b_i^2\right)\frac{e^{4\phi}}{\text{Vol}^3} + \frac{f_0^2}{4}\frac{e^{4\phi}}{\text{Vol}} - 2\frac{e^{3\phi}}{\text{Vol}^{3/2}}, \tag{3.41}$$

where the first 3 terms come from fluxes: 3-form, 4-form, and 0-form, respectively, and the final term comes from the O6-plane. The first 3 terms take on the form we indicated in eq. (3.12) for p-forms. The final term carries a minus sign, since O6-planes carry *negative* tension, as we indicated in eq. (3.13). This term is crucial to achieve stability. In this form it is not clear that the fluxes scale out, however. By rewriting this in terms of the variables b_i, $b_4 = e^{-\phi}\sqrt{\text{Vol}}/\sqrt{2}$, and then scaling according to eq. (3.37), it is simple to show that one recovers a simpler version of (3.39), one with a_i and a_4 set to zero.

We plot V in Fig. 3-1. In order to discuss the properties of this potential, let us begin by discussing its supersymmetric properties. The SUSY vacuum lies at

$$a_1 = 0,\ b_1 \approx 1.29,\ a_4 = 0,\ b_4 \approx 1.22. \tag{3.42}$$

For $\delta = -1$, this has a corresponding positive definite mass matrix and is clearly stable. However, for $\delta = +1$ the mass matrix has negative eigenvalues and so is tachyonic, as reported by DGKT. Nevertheless, it is stable as it satisfies the Breitenlohner-Freedman bound [57] which states that tachyonic vacua can be stable if the cosmological constant is large and negative. In Fig. 3-1 (top), we set the axions to zero and plot V as a a function of b_1 and b_4 in the vicinity of the SUSY vacuum. We see that with respect to these two co-ordinates, the potential has a regular stable minimum. Note also, that with $a_1 = a_4 = 0$ then the values of the potential in (3.39) for $\delta = \pm 1$ coincide. In Fig. 3-1 (bottom), we plot V with b_1 and b_4 fixed at their SUSY values, and allow a_1 and a_4 to vary. We have plotted the case $\delta = +1$ as its behavior is the most interesting.

Now, let us investigate the potential further away from the SUSY point. We find that in the $\delta = +1$ case (tachyonic), there is a second stationary point of the potential.

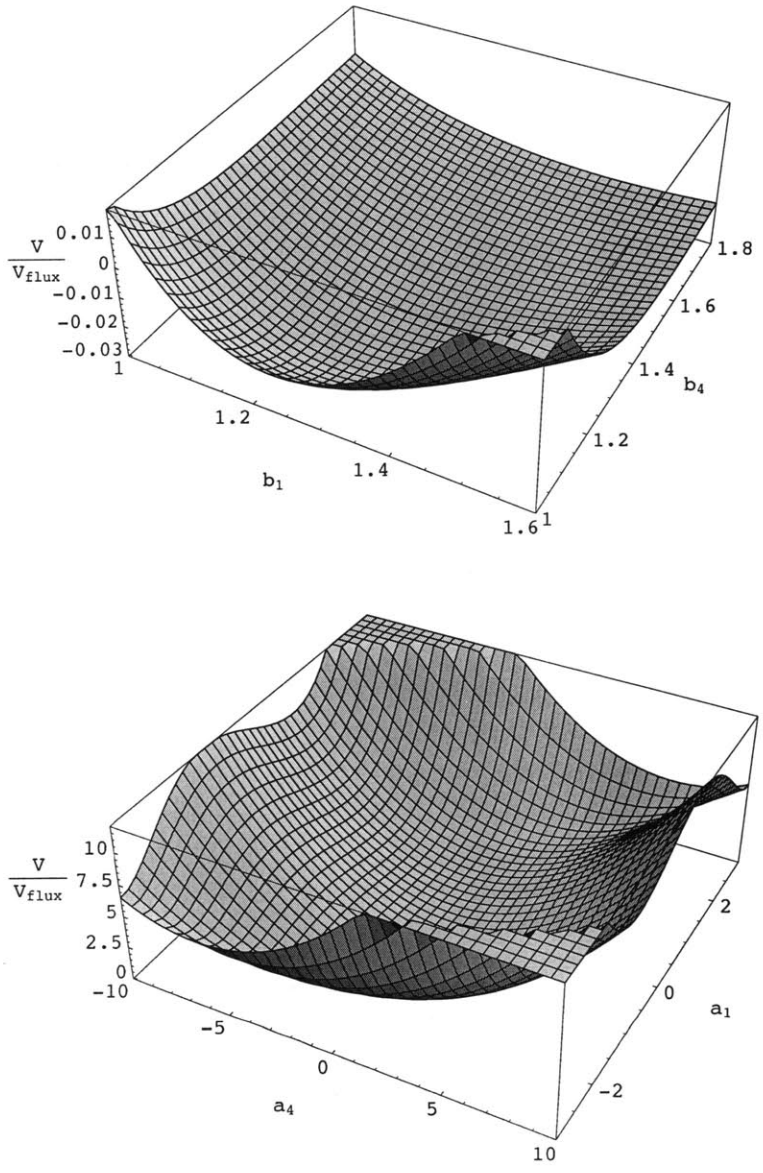

Figure 3-1: Top: The potential $V = V(b_1, b_4)$ with axions at their SUSY values: $a_1 = a_4 = 0$. Bottom: The potential $V = V(a_1, a_4)$ with $\delta = +1$ and geometric moduli at their SUSY values: $b_1 \approx 1.291$, $b_4 \approx 1.217$.

It is non-supersymmetric, and lies at:

$$a_1 \approx \pm 0.577,\ b_1 \approx 1.15,\ a_4 \approx \mp 0.544,\ b_4 \approx 1.09. \tag{3.43}$$

Given these two stationary points of V (one SUSY, one non-SUSY) we choose to plot V as a function of λ, where λ is a parameter that linearly interpolates between these two points. With the SUSY point denoted by a vector of moduli $\psi_{i,susy}$ and the other (non-SUSY) stationary point denoted by a vector $\psi_{i,stat}$, we form the interpolating vector:

$$\psi_i(\lambda) \equiv (1-\lambda)\psi_{i,susy} + \lambda\psi_{i,stat} \tag{3.44}$$

so that $\lambda = 0$ is the SUSY vacuum and $\lambda = 1$ is the second stationary point. We plot this in Fig. 3-2 (top). Also, in Fig. 3-2 (bottom), we plot V as a function of b_4 with all other moduli at their SUSY values. As already stated, $b_4 \approx 1.217$ is the SUSY point (a minimum with respect to b_4) and this exists on the left hand side of the figure with the potential much lower than shown. However, the interesting feature is that for $b_4 \approx 7.912$ there exists a local maximum (with $V > 0$) with respect to this modulus (but not stationary with respect to the other moduli) and then V approaches zero from above as $b_4 \to \infty$. There is quite similar behavior when one plots V versus the radial modulus, with the dilaton fixed.

Let us recapitulate the salient features of this class of vacua. We have come to an important realization: the potential V is of the form $V = V_{\text{flux}}(f_i)\,\text{func}(\psi^i)$, where f_i are flux integers and $\text{func}(\psi^i)$ is some function of the (rescaled) moduli, *independent of fluxes.* Hence, apart from the overall multiplicative scale (which is proportional to the cosmological constant) all vacua look the same. This means that the slow-roll parameters ϵ and η are independent of the fluxes. So for this model inflation is realized by all or none of the flux vacua.

Of course we wish to know if the potential is sufficiently flat in some region to exhibit slow-roll. Here we see a general barrier to this. Note that the potential is a polynomial in $\{a_i, b_i, 1/b_i\}$. Naively, this may look as if it allows for inflation due to some form of power law potential, e.g., the potential is quadratic in a_4, so this may look like Linde's $\sim \phi^2$ "chaotic inflation" [58, 59]. However, by inspecting the form of eq. (3.27), we see that this is not at all the case. The factor of b_i^2 in the summand changes the picture significantly. It means that the typical contribution to ϵ is not $\mathcal{O}(\phi^{-2})$ but $\mathcal{O}(1)$, and cannot be tuned small by taking ϕ large, as in chaotic inflation. In fact, an extensive numerical search of moduli space (detailed below) suggests that $\epsilon > 1$ whenever $V > 0$ (of course $\epsilon \to 0$ at the stationary point(s) of the potential, but $V < 0$ there). With the axions set to zero, it is simple to analytically prove the non-existence of inflation. With axions non-zero, we produced vast tables of ϵ supporting this result. We will give a representative plot of ϵ in the upcoming VZ model (see Fig. 3-4).

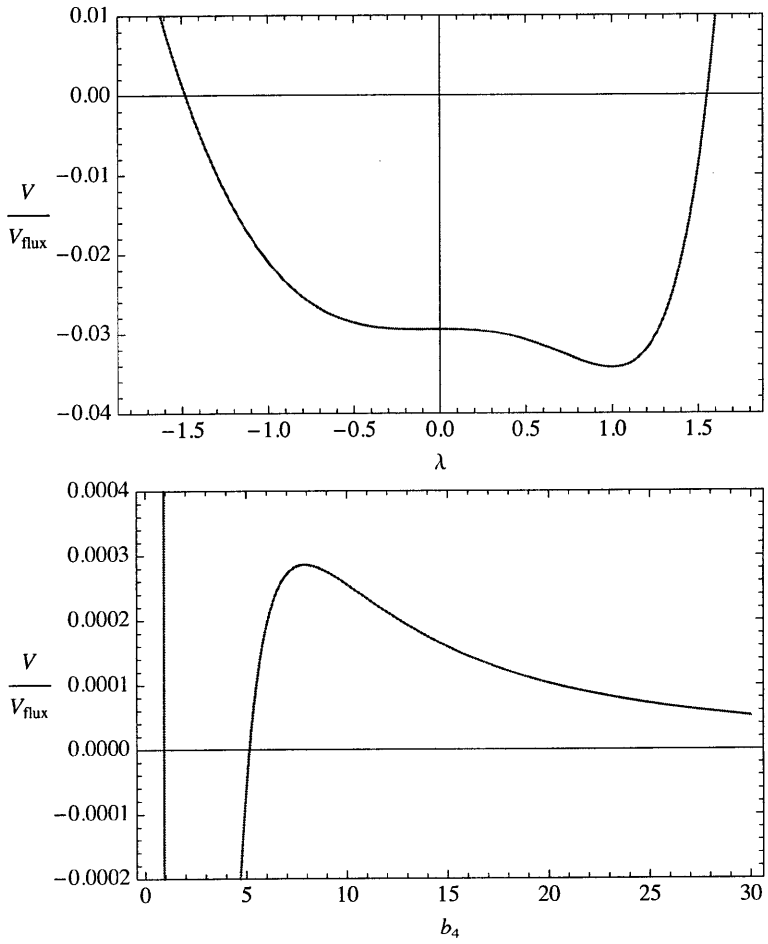

Figure 3-2: Top: We plot $V = V(\lambda)$ by interpolating between the two stationary points of the potential, which exists for the 'lower case'. $\lambda = 0$ corresponds to the (tachyonic) SUSY vacuum and $\lambda = 1$ corresponds to a local (non-SUSY) minimum. Bottom: We plot $V = V(b_4)$, focusing on large b_4, with all other moduli fixed at their SUSY values.

3.3.3 The Model of Villadoro and Zwirner (VZ)

In March 2005, Villadoro and Zwirner (VZ) [40] constructed a class of orientifold compactifications based on toroidal orbifolds, where the dilaton and all the moduli associated with the torus are stabilized through the inclusion of p-form field strength fluxes (as discussed earlier) and other sorts of fluxes, simply referred to as *general fluxes*. Their model is strongly motivated by the work of Derendinger et al. [38]. In particular, VZ include what are known as Scherk-Schwarz geometrical fluxes[14] which provides a large class of vacua. With many fluxes in the model, there are a number of Bianchi identities and tadpole constraints that the fluxes must satisfy. This is achieved by including D6-branes and O6-planes. The interested reader is referred to the original paper [40] for details.

As originally studied by Derendinger et. al. [38] the orbifold is $T^6/\mathbb{Z}_2$. A further $\mathbb{Z}_2$ projection is performed to obtain an O6 orientifold. This particular orientifold permits 6 degrees of freedom in the metric on the compact space. The torus takes the form $T_6 = T_2 \times T_2 \times T_2$ and possesses a diagonal metric. Then, without loss of generality, the 6-dimensional metric can be parameterized by 6 variables γ_i and β_i (i=1,2,3) as follows:

$$(ds^2)_6 = \sum_{i=1}^{3} \left((\gamma_i/\beta_i)\,(dx^i)^2 + (\gamma_i\,\beta_i)\,(dy^i)^2 \right), \tag{3.45}$$

$$\text{Vol} = \int_{T^6/\mathbb{Z}_2^2} d^6x\sqrt{g_6} = \gamma_1\gamma_2\gamma_2 = b_1 b_2 b_3. \tag{3.46}$$

The form of the volume explains the choice in decomposing the metric as above, namely that the product of the γ_i is proportional to the volume, as it was for DGKT. In turn, we again identify 3 good Kähler co-ordinates b_1, b_2, b_3. The β_i, on the other hand, are related to the complex structure and axio-dilaton moduli.

Let us now provide the full list of moduli (again ignoring "blow-up modes"). In this case there are seven complex moduli that survive the orientifold projection: 3 Kähler moduli $\psi_i = a_i + ib_i$, $i = 1,2,3$, an axio-dilaton $\psi_4 = a_4 + ib_4$ ($b_4 = e^{-\phi}\sqrt{\text{Vol}}/\sqrt{\beta_1\,\beta_2\,\beta_3}$), and 3 complex structure moduli $\psi_i = a_i + ib_i$, $i = 5,6,7$ ($b_5 = e^{-\phi}\sqrt{\text{Vol}}\sqrt{\beta_2\,\beta_3/\beta_1}$ etc).

The Kähler potential takes on an extremely simple form: in the notation of Section 3.3.1 it has all 7 $n_i = 1$. Explicitly, it is[15]

$$K = -\ln\left(b_1\,b_2\,b_3\,b_4\,b_5\,b_6\,b_7\right) \tag{3.47}$$

By way of comparison to the DGKT model, it is as though $b_4^4 \to b_4\,b_5\,b_6\,b_7$, in order to accommodate the 3 complex structure moduli that appear here.

[14] Geometric flux here refers to a particular kind of topologically non-trivial alteration of the metric on the compact space which yields a contribution to the scalar potential analogous to the contributions from the p-form fluxes.

[15] We follow the convention of Ref. [40] where an overall factor of 2^7 was removed from the argument of the logarithm, since it can be simply reabsorbed into W.

The superpotential incorporates geometric flux, in addition to the familiar p-form flux, and satisfies the 'tadpole condition' with D-branes. The superpotential, as derived in [38], is given by:

$$\begin{aligned} W &= f_{111} - f_{112}(\psi_1 + \psi_2 + \psi_3) + f_{222}\,\psi_1\,\psi_2\,\psi_3 \\ &+ f_{122}(\psi_1\,\psi_2 + \psi_1\,\psi_3 + \psi_2\,\psi_3) - f'_{111}\,\psi_4 \\ &+ f'_{112}\,\psi_4(\psi_1 + \psi_2 + \psi_3) - f_{114}(\psi_5 + \psi_6 + \psi_7) \\ &+ f_{124}(\psi_1(\psi_6 + \psi_7) + \psi_2(\psi_5 + \psi_7) + \psi_3(\psi_5 + \psi_6)) \\ &- f_{113}(\psi_1\,\psi_5 + \psi_2\,\psi_6 + \psi_3\,\psi_7) \end{aligned} \tag{3.48}$$

Here we have designated the fluxes by f_{ijk}, which correspond to different choices of p-form and geometric flux wrapped on various cycles of the torus. There are constraints that the fluxes f_{ijk} must satisfy, namely

$$f_{124}(f_{124} - f_{113}) = 0, \;\; f'_{112}(f_{124} - f_{113}) = 0. \tag{3.49}$$

VZ find a family of SUSY vacua by choosing the following parameterization of fluxes:

$$\begin{aligned} &f_{111} = -15p_1, \;\; f_{112} = \frac{3p_2}{q_1}, \;\; f_{122} = \frac{p_1}{q_1^2}, \\ &f_{222} = -\frac{3p_2}{q_1^3}, \;\; f'_{111} = -\frac{2p_2}{q_2}, \;\; f'_{112} = -\frac{2p_1}{q_1 q_2}, \\ &f_{114} = -\frac{2p_2}{q_3}, \;\; f_{113} = f_{124} = -\frac{6p_1}{q_1 q_3}. \end{aligned} \tag{3.50}$$

We note that the f_{ijk} are actually non-integer. Here we do not record the conditions that p_1, p_2, q_1, q_2, q_3 must satisfy, but refer the reader to [40]. We do note that $\{q_1, q_2, q_3\} \in \mathbb{R}^+$. What is important is that this designates an infinite family of vacua with fluxes parameterized by the set of five parameters p_1, p_2, q_1, q_2, q_3. So we have started with a superpotential with 9 fluxes: $f_{111}, \ldots, f_{124}$, one has been eliminated by the conditions (3.49) ($f_{113} = f_{124}$), three have been eliminated by demanding that the SUSY condition (3.20) be satisfied for each ψ_i, leaving five independent parameters.

For this family of vacua it is rather straightforward to show that we can scale out the fluxes q_1, q_2, q_3 by making the following rescaling of our fields:

$$\begin{aligned} &\psi_i \to q_1\,\psi_i \;\; (i = 1, 2, 3), \;\; \psi_4 \to q_2\,\psi_4, \\ &\psi_i \to q_3\,\psi_i \;\; (i = 5, 6, 7). \end{aligned} \tag{3.51}$$

This leaves only p_1 and p_2 of which we can scale out one of them, leaving only their ratio as a tunable parameter

$$s \equiv \frac{p_1}{p_2} \tag{3.52}$$

($p_2 = 0$ can be handled separately).

Now let us focus on the symmetric case, in which $\psi_1 = \psi_2 = \psi_3$ and $\psi_5 = \psi_6 = \psi_7$,

and keep track of the fields ψ_1, ψ_4, ψ_5. We find that W is simplified to

$$\begin{aligned} W &= -15p_1 - 9p_2\psi_1 + 3p_1\psi_1^2 - 3p_2\psi_1^3 \\ &\quad +2p_2(\psi_4 + 3\psi_5) - 6p_1\psi_1(\psi_4 + 3\psi_5). \end{aligned} \tag{3.53}$$

We note that since W only depends on a linear combination of ψ_4 and ψ_5, namely $\psi_4 + 3\psi_5$, the potential V only depends on the same linear combination of the corresponding axions, namely $\hat{a}_4 \equiv a_4 + 3a_5$.

By using (3.11), it is a straightforward matter to obtain the potential. The result is a rather long expression that we report in Appendix B eq. (3.81). The leading prefactor

$$V_{\text{flux}} \equiv \frac{p_2^2}{q_1^3\, q_2\, q_3^3} \tag{3.54}$$

is an overall multiplicative scale that depends on the fluxes. At fixed s there exists a family of solutions for p_1, p_2, q_1, q_2, q_3 for which $V_{\text{flux}} \to 0$ parametrically. However, one should note the explicit appearance of $s = p_1/p_2$ in the potential, which controls its *shape*. We mention that without loss of generality we can focus on s non-negative, since $s \to -s$ and $a_i \to -a_i$ leaves V unchanged.

By solving the equations $D_i W = 0$, one can show that the SUSY vacuum lies at:

$$a_1 = a_4 + 3\,a_5 = 0, \;\; b_1 = b_4 = b_5 = \sqrt{\frac{5}{3}}, \tag{3.55}$$

for all s. We note that this (AdS) SUSY vacuum is tachyonic but stable, as is satisfies the Breitenlohner-Freedman bound [57]. The potential here takes on the value

$$V_{\text{SUSY}} = -V_{\text{flux}} \frac{432\sqrt{15}}{125} \left(1 + 15s^2\right). \tag{3.56}$$

Now, an important special case is when $p_1 = p_2$ ($s = 1$), since as VZ describe, this provides this $\mathcal{N} = 1$ supergravity theory with an interpretation in terms of an $\mathcal{N} = 4$ supergravity theory with extended (gauged) symmetry. This may be of some interest [60]. In this case, the potential in eq. (3.81) may be simplified to

$$\begin{aligned} V &= V_{\text{flux}}(36\tilde{a}_1^6 + 108b_1^2\tilde{a}_1^4 + 144\tilde{a}_4\tilde{a}_1^4 + 1280\tilde{a}_1^3/3 \\ &+ 108b_1^4\tilde{a}_1^2 + 144\tilde{a}_4^2\tilde{a}_1^2 + 144\tilde{a}_4 b_1^2\tilde{a}_1^2 + 144b_4^2\tilde{a}_1^2 \\ &+ 432b_5^2\tilde{a}_1^2 + 2560\tilde{a}_4\tilde{a}_1/3 + 36b_1^6 + 48\tilde{a}_4^2 b_1^2 + 48b_1^2 b_4^2 \\ &- 432b_1^2 b_5^2 - 576b_1^2 b_4 b_5 + 102400/81)/(b_1^3\, b_4\, b_5^3). \end{aligned} \tag{3.57}$$

with $\tilde{a}_1 \equiv a_1 - 1/3$ and $\tilde{a}_4 = a_4 + 3a_5 + 4/3$. In addition to the SUSY vacua, we find

three additional AdS vacua given by

$$\begin{aligned} a_1 &= a_4 + 3a_5 = \frac{1}{3},\ b_1 \approx 1.38,\ b_4 = b_5 \approx 1.26 \\ a_1 &= a_4 + 3a_5 = \frac{1}{3},\ b_1 \approx 1.38,\ b_4 \approx 2.87,\ b_5 \approx 0.958 \\ a_1 &= 1,\ a_4 + 3a_5 = -4,\ b_1 = b_4 = b_5 = \frac{2}{\sqrt{3}} \end{aligned} \tag{3.58}$$

In Fig. 3-3 (top) we plot V as a function of a_1 and a_4 with other parameters fixed at their SUSY values. We see that V is relatively flat along each axis, while V is steep along diagonals.

Now, this potential contains one modulus that is not stabilized. One linear combination of the axions is left exactly massless at the SUSY vacuum. We we will return to this later in the discussion. This is a result of the fact that the superpotential in eq. (3.53) only depends on the combination $a_4 + 3\,a_5$. A plot of V as a function of a_4 and a_5, with all other moduli fixed at their SUSY values, is given in Fig. 3-3 (bottom). We see the flat 'valley'. The existence of such a flat direction certainly seems useful from the point of view of inflation, however one should recall that this flat direction emanates from an AdS vacuum.

For slow-roll it is again evident that this is very difficult, due to the argument presented at the end of Section 3.3.2, namely that the characteristic value of ϵ in these types of tree-level toroidal models is $\mathcal{O}(1)$. However, it is important to investigate the effect of our tunable parameter s. To get a flavor of its effect, in Fig. 3-4 (top) we plot $\epsilon = \epsilon(a_1)$ for $2 < s < 20$. We see that $\epsilon > 1$ in this region. Indeed our numerical studies indicate (detailed below) that there is no inflating region anywhere in moduli space. Again this is based on the results of vast tables of ϵ over moduli space. For a more conventional representation, in Fig. 3-4 (bottom) we plot ϵ as a function of a pair of moduli, namely a_1 and b_1, with $s = 1$. Here ϵ is large in all regions in which $V > 0$. ($\epsilon \to \infty$ as $V \to 0$ and $\epsilon \to 0$ at the AdS minimum). The plot displays a dip in ϵ as $a_1 \to -1$, $b_1 \to 0$. At this point we find $\epsilon \to 4$.

3.3.4 The Model of Ihl and Wrase (IW)

In the previous two models, the Kähler potential took on the form of eq. (3.23), which we referred to as diagonal torus models. It followed from this that the first slow-roll parameter took on the form as given in eq. (3.27). For potentials V that were rational in the moduli, this meant that $\epsilon = \mathcal{O}(1)$ was quite natural. We would like then to investigate more complicated models in which this does not occur. In April 2006 Ihl and Wrase [41] obtained an explicit example of this nature. Their work is strongly motivated by the work of DGKT. Indeed they also consider massive type IIA supergravity. However, unlike the DGKT model, we find that one tunable parameter remains in the potential, as we found in the VZ model.

The orientifold is $T^6/\mathbb{Z}_4$. Unlike the torii of DGKT and VZ, this orientifold *does not* permit the decomposition of T^6 to $T_2 \times T_2 \times T_2$ with identical T_2s. Instead the T_2s

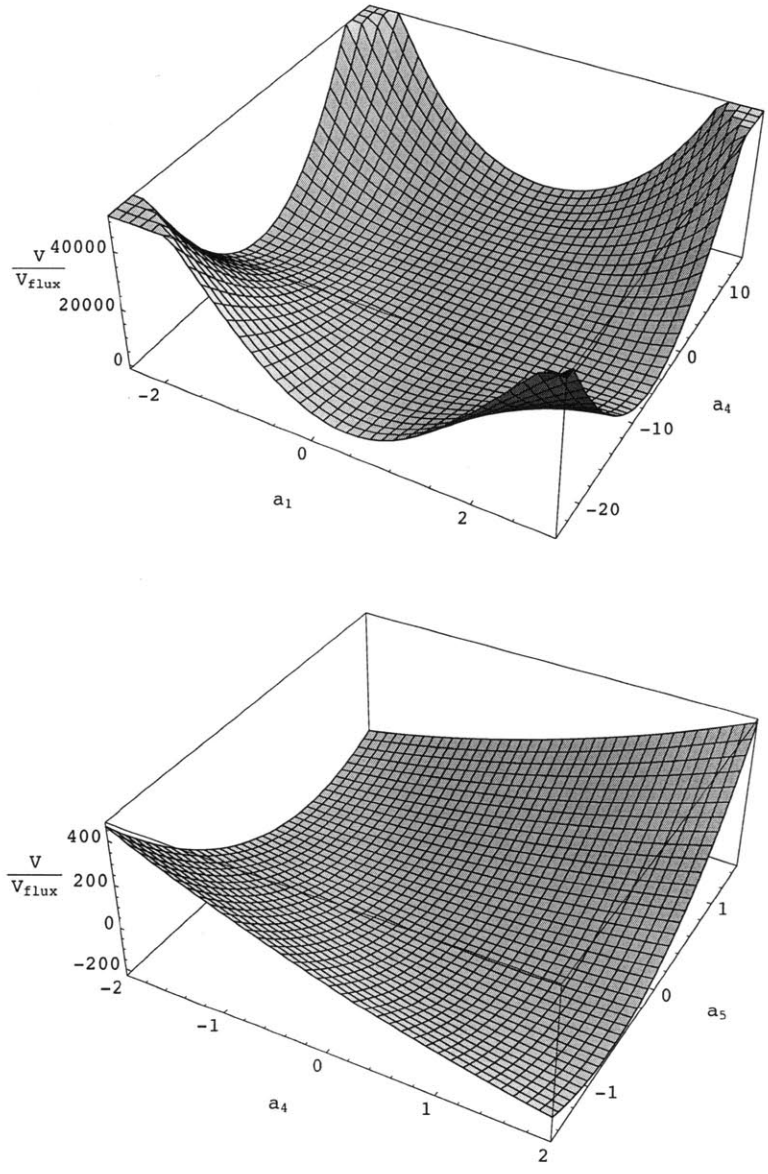

Figure 3-3: Top: The potential $V = V(a_1, a_4)$ with $a_5 = 0$ and other moduli taking on their SUSY values. Bottom: The potential $V = V(a_4, a_5)$ with other moduli taking on their SUSY values.

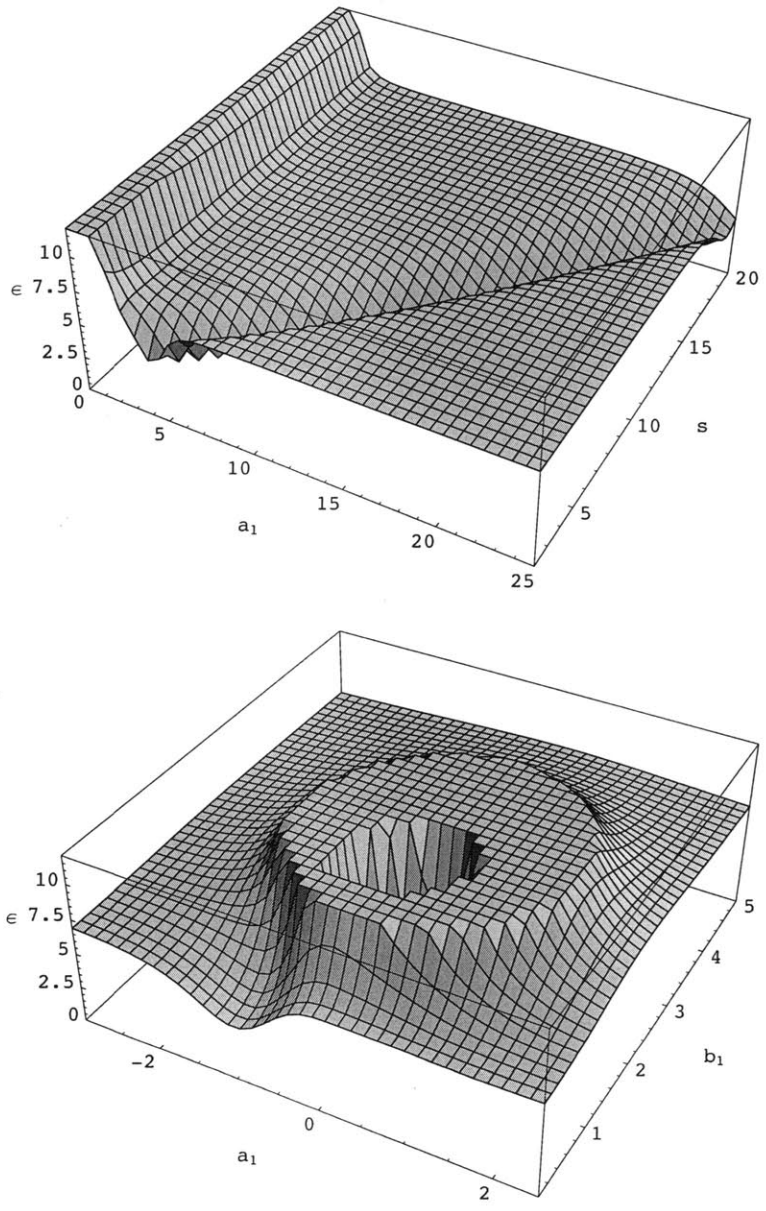

Figure 3-4: Top: The slow-roll parameter $\epsilon(a_1, s)$ with other moduli taking on their values from the third set of eq. (3.58). Bottom: The slow-roll parameter $\epsilon(a_1, b_1)$ with $s = 1$ and other moduli taking on their SUSY values.

must have *different* metrics. The interested reader is referred to the original paper [41] for details.

The metric on the compact space is somewhat more complicated than our previous models. When expressed in terms of the most useful co-ordinates (those that are readily related to Kähler and complex structure moduli) the metric on the compact space is non-diagonal. Here there are 4 independent degrees of freedom that appear explicitly in the metric. Denoting them as usual by γ_i, the metric is given by

$$(ds^2)_6 = \sum_{i=1}^{3} \gamma_i \left[(dx^i)^2 + (dy^i)^2\right] + 2\,\gamma_4 \left(dx^1\,dx^2 + dy^1\,dy^2 - \sum_{i,j=1}^{2} \epsilon_{ij}\,dx^i\,dy^j \right), \quad (3.59)$$

$$\begin{aligned} \text{Vol} &= \int_{T^6/\mathbb{Z}_4} d^6x\sqrt{g_6} = U_2\gamma_3(\gamma_1\gamma_2 - 2\gamma_4^2)/4 \\ &= b_3(b_1b_2 - b_4^2/2), \end{aligned} \quad (3.60)$$

where ϵ_{ij} is the Levi-Civita symbol defined by $\epsilon_{12} = -\epsilon_{21} = 1$, $\epsilon_{11} = \epsilon_{22} = 0$. The form of the volume requires a little explanation: The square-root of the determinant of the metric is easily shown to be $\gamma_3(\gamma_1\gamma_2 - 2\gamma_4^2)$, and indeed Vol is proportional to this. There is also a factor of U_2, which is related to the canonical complex structure moduli. It is known as a 'pure type' contribution that is in some sense hidden in the metric. The interested reader is referred to footnote 11 of the IW paper for clarification. Again we have written the volume in terms of some b_i that are good Kähler co-ordinates.[16] Note that the volume is *not* simply a product of the Kähler co-ordinates.

In summary, after the orientifolding there remains *one* complex structure modulus U_2. In total we have 4 Kähler moduli $\psi_i = a_i + ib_i$, $i = 1, \ldots, 4$ and 2 other moduli which mix the axio-dilaton and the complex structure modulus: $\psi_5 = a_5 + ib_5$ ($b_5 = e^{-\phi}\sqrt{\text{Vol}}/\sqrt{U_2}$) and $\psi_6 = a_6 + ib_6$ ($b_6 = 2\sqrt{U_2}\,e^{-\phi}\sqrt{\text{Vol}}$).[17] (We again ignore the "blow-up modes").

Here the Kähler potential *does not* take the form of the previous 2 models, i.e., it is not of the form of a logarithm of a product of geometric moduli and so is not quite of the form discussed in Section 3.3.1. There is a modification due to the non-trivial form of the volume, namely $\text{Vol} = b_3(b_1b_2 - b_4^2/2)$. The Kähler potential is

$$K = -\ln\left(2\,b_3(b_1\,b_2 - b_4^2/2)b_5^2\,b_6^2\right) \quad (3.61)$$

— see eq. (3.23) for comparison.

The ingredients for the superpotential are just the same as the DGKT model. A 0-form, 3-form, 4-form (and an unimportant 2-form) are included. The superpotential

[16] In the IW paper: $\text{Vol} = \kappa\,b_3(b_1b_2 - b_4^2/2)$. By rescaling $b_i \to \kappa^{-1/3}b_i$, $i = 1, 2, 3, 4$ we obtain eq. (3.60) and κ is eliminated.

[17] The canonical axions are $\xi_5 = 2a_5$ and $\xi_6 = 2a_6$.

is a simple modification of the DGKT model, namely

$$W = \frac{f_6}{\sqrt{2}} + \sum_{i=1}^{4} \frac{f_{4,i}}{\sqrt{2}}\,\psi_i - \frac{f_0}{\sqrt{2}}\,\psi_3(\psi_1\,\psi_2 - \psi_4^2/2) - 2\,f_3(\psi_5 + \psi_6), \tag{3.62}$$

which is to be compared to eq. (3.35). We note that there is one additional flux component: $f_{4,4}$, which is due to the presence of a 4th Kähler modulus ψ_4.

Furthermore, just as in the DGKT model, an 06-plane is introduced in order to satisfy the tadpole condition. This occurs in precisely the same way as before (see eq. (3.36)), i.e., $f_0\,f_3 = -2$.

Again let us exploit all existing shift and scaling symmetries. We perform the following transformations on our fields:

$$\begin{aligned}
\psi_i &\to \frac{1}{|f_{4,i}|}\sqrt{\frac{|f_{4,1}\,f_{4,2}\,f_{4,3}|}{|f_0|}}\,\psi_i \;\; (i=1,2,3),\\
\psi_4 &\to \sqrt{\frac{|f_{4,3}|}{|f_0|}}\,\psi_4,\\
\psi_5 &\to \frac{1}{|f_3|}\sqrt{\frac{|f_{4,1}\,f_{4,2}\,f_{4,3}|}{|f_0|}}\,\psi_5 + \frac{1}{2\sqrt{2}}\frac{f_6}{f_3},\\
\psi_6 &\to \frac{1}{|f_3|}\sqrt{\frac{|f_{4,1}\,f_{4,2}\,f_{4,3}|}{|f_0|}}\,\psi_6,
\end{aligned} \tag{3.63}$$

which was chosen in such a way as to leave the kinetic terms invariant. This allows one to rewrite the superpotential as

$$\begin{aligned}
W = \;& \sqrt{\frac{|f_{4,1}\,f_{4,2}\,f_{4,3}|}{|f_0|}}\Big(\sum_{i=1}^{3}\frac{\hat{f}_{4,i}}{\sqrt{2}}\,\psi_i + \frac{\hat{f}_{4,4}}{\sqrt{2}}\,t\,\psi_4\\
& -\frac{\hat{f}_0}{\sqrt{2}}\,\psi_3(\psi_1\,\psi_2 - \psi_4^2/2) - 2\,\hat{f}_3(\psi_5+\psi_6)\Big).
\end{aligned} \tag{3.64}$$

Here there exists one combination of the fluxes that does *not* scale out:

$$t \equiv \frac{|f_{4,4}|}{\sqrt{|f_{4,1}\,f_{4,2}|}}. \tag{3.65}$$

This is a result of the non-trivial ("intersection") form for the volume, which puts ψ_4 on a different footing from the other Kähler moduli.

We turn now to the 4-dimensional potential V. In the presence of all axions, the result is somewhat complicated: see Appendix C eq. (3.82). Here we note that a consistent solution is found with all (shifted) axions vanishing, and so we will focus on this case: $a_1 = \ldots = a_6 = 0$. Also note that ψ_5 and ψ_6 are treated on equal footing. Inspired by this fact, we will concentrate on the case $\psi_5 = \psi_6$. This sets

$U_2 = 1/2$, as reported in the IW paper. We also make a final set of multiplicative transformations by ± 1, namely: $b_1 \to \hat{f}_{4,1}\, b_1$, $b_2 \to \hat{f}_{4,1}\, b_2$, $b_4 \to \hat{f}_{4,4}\, b_4$, which *still* preserves the form of the kinetic energy. We find

$$\begin{aligned} V &= V_{\text{flux}}[2(b_1^2 + b_2^2 + b_3^2) + 2\,\delta_{12}\, b_4^2 + 16\, b_5^2 - 16\sqrt{2}\, b_1\, b_2\, b_3\, b_5 \\ &+ 8\sqrt{2}\, b_3\, b_4^2\, b_5 + 2\, b_1^2\, b_2^2\, b_3^2 - 2\, b_1\, b_2\, b_3^2\, b_4^2 + b_3^2\, b_4^4/2 + 4(b_1 + \delta_{12}\, b_2) b_4\, t \\ &+ (b_4^2 + 2\, b_1\, b_2)\, t^2]/(2\, b_3 (b_1\, b_2 - b_4^2/2) b_5^4), \end{aligned} \tag{3.66}$$

where

$$V_{\text{flux}} \equiv \frac{|f_0|^{5/2} |f_3|^4}{|f_{4,1}\, f_{4,2}\, f_{4,3}|^{3/2}} \tag{3.67}$$

as we defined it for the DGKT model. Note however, that this is not the only piece that depends on the magnitude of the fluxes, since the flux parameter t also appears in (3.66). So there is one combination of the fluxes that describes the *shape* of the potential. We have defined $\delta_{12} \equiv \hat{f}_{4,1}\hat{f}_{4,2} = \pm 1$ which delineates two families of potentials. We should also keep track of the physical constraints that the area of the third torus and the compact volume Vol are both positive, so $b_3 > 0$ and $b_1 b_2 - b_4^2/2 > 0$.

All stationary points are AdS (even the non-SUSY ones). The interested reader is referred to the IW paper [41] for a detailed description of the locations of these stationary points. Here we begin by noting that when $\delta_{12} = +1$ and $t = 0$, there is a SUSY AdS minimum, which coincides with that of the DGKT model:

$$b_1 = b_2 = b_3 \approx 1.29,\ b_4 = 0,\ b_5 \approx 0.609. \tag{3.68}$$

(compare to eq. (3.42) with renaming of variables; b_5 of IW replaced by $b_4/2$ of DGKT.)

Since the Kähler potential is not simply the logarithm of a product of moduli, ϵ is *not* given by eq. (3.27). Instead we revert to eq. (3.18). We emphasize that the particular transformations we have performed on the ψ_i have left the form of $K^{i\bar{j}} V_i V_{\bar{j}}$ unchanged. With this in mind, we find the following:

$$\epsilon = \frac{1}{V^2}\left\{ \sum_{i,j=1}^{4} M_{ij} \left(\frac{\partial V}{\partial a_i}\frac{\partial V}{\partial a_j} + \frac{\partial V}{\partial b_i}\frac{\partial V}{\partial b_j} \right) + \sum_{i=5}^{6} \frac{b_i^2}{2} \left[\left(\frac{\partial V}{\partial a_i} \right)^2 + \left(\frac{\partial V}{\partial b_i} \right)^2 \right] \right\}, \tag{3.69}$$

where

$$M \equiv \begin{pmatrix} b_1^2 & b_4^2/2 & 0 & b_1 b_4 \\ b_4^2/2 & b_2^2 & 0 & b_2 b_4 \\ 0 & 0 & b_3^2 & 0 \\ b_1 b_4 & b_2 b_4 & 0 & b_1 b_2 + b_4^2/2 \end{pmatrix}, \tag{3.70}$$

and V given in terms of our rescaled variables, i.e., by eq. (3.66) for the simple vanishing axion case, and by eq. (3.82) for the general non-vanishing axion case. In Fig. 3-5 we give a representative plot of a piece of moduli space. We see significant variation as we change the flux parameter from $t = 1$ in (top) to $t = 10$ in (bottom).

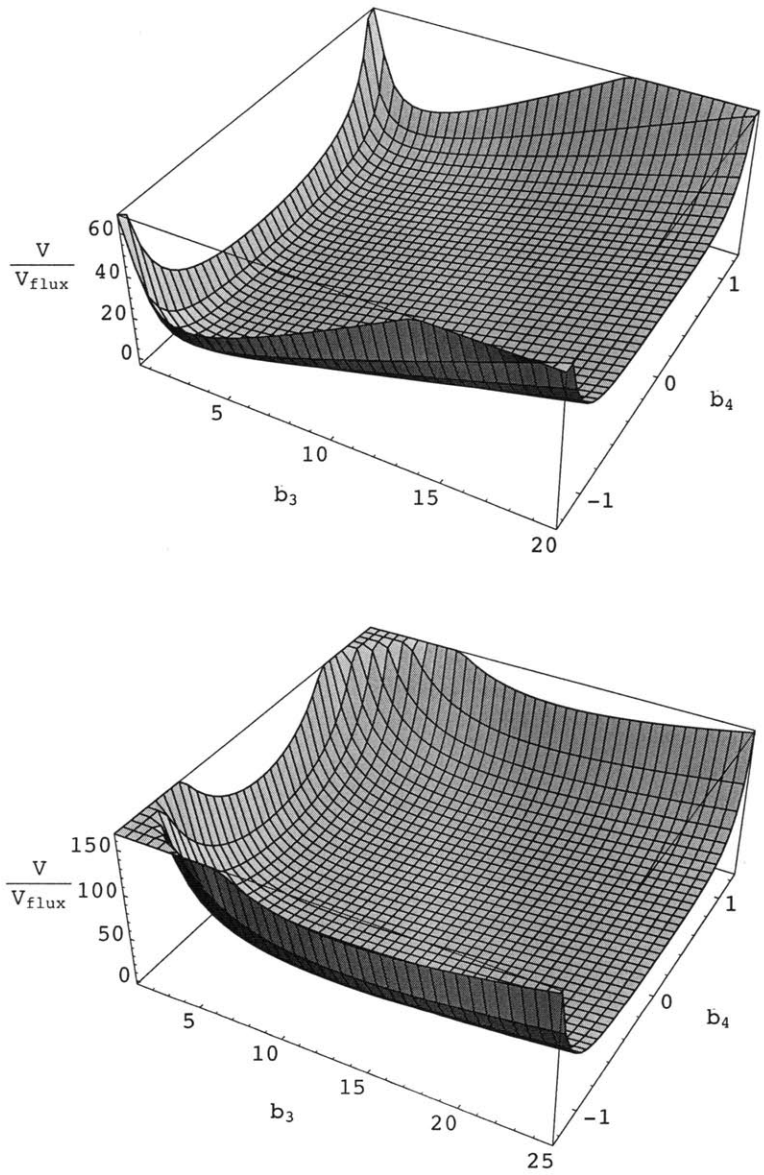

Figure 3-5: The potential $V = V(b_3, b_4)$ with $b_1 = b_2 = b_5 = 1$ and $\delta_{12} = -1$. Top: $t = 1$. Bottom: $t = 10$.

In this plot we have ensured that b_4 has remained in the physical region given by $b_1 b_2 - b_4^2/2 > 0$. We note that the potential becomes singular at this boundary; this follows again from the non-trivial form of Vol. Our numerical investigations into the slow-roll parameter ϵ have again yielded $\epsilon > 1$ whenever $V > 0$, although we have not investigated the full moduli space – we did not include all axions in our search.

3.3.5 Comments on Blow-up Modes

In the three models that we have investigated, we have ignored a class of moduli known as "twisted moduli" or "blow-up modes". Recall that apart from the dilaton, the geometric moduli describe the size and shape of the compact space, i.e., its *geometry*. These are the Kähler and complex structure moduli. For a smooth compact space, this is fully general. However, the models investigated here are not smooth; they are all *orbifolds*, which have fixed points. These fixed points correspond to conical singularities. In the large volume limit, these conical singularities are 'blown-up' and smoothed out. The effective 4-dimensional description then captures this aspect of the geometry by a modulus for each fixed point; the so called blow-up modes.

These blow-up modes can be included in the analysis in a straightforward fashion through the use of the Kähler potential K and superpotential W. Let us give an explicit example; the DGKT model. Here there are 9 fixed points, and so there are 9 blow-up modes. We call these: $\psi_i = a_i + i\, b_i$ for $i = 5, \ldots, 13$. The expression for the volume in eq. (3.33) is modified to

$$\text{Vol} = b_1 b_2 b_3 - \frac{1}{54} \sum_{i=5}^{13} b_i^3. \tag{3.71}$$

The Kähler potential (3.34) and the superpotential (3.35) are modified to

$$K = -\log\left[32 \left(b_1 b_2 b_3 - \frac{1}{54} \sum_{i=5}^{13} b_i^3\right) b_4^4\right], \tag{3.72}$$

$$W = \frac{f_6}{\sqrt{2}} + \sum_{i=1}^{3} \frac{f_{4,i}}{\sqrt{2}} \psi_i + \sum_{i=5}^{13} \frac{f_{4,i}}{\sqrt{2}} \psi_i - \frac{f_0}{\sqrt{2}} \left(\psi_1 \psi_2 \psi_3 - \frac{1}{54} \sum_{i=5}^{13} \psi_i^3\right) - 2 f_3 \psi_4. \tag{3.73}$$

In principle we could now explore this larger moduli space for inflation. However, it is numerically difficult; we have moved from 4 complex moduli (axio-dilation plus three Kähler moduli) to 13 complex moduli through the addition of 9 complex blow-up modes. Instead of a full investigation into the effects of dynamical blow-up modes, we shall freeze the blow-up modes at some vacuum expectation value (vev). Such a vev is explicitly found in the DGKT paper. We then explore the effect of a non-zero vev for the blow-up modes on the original moduli. Suppose the b_i are frozen in at some value, which we characterize as: $B \equiv \frac{1}{54} \sum_{i=5}^{13} b_i^3$. With B taken as a constant, our Kähler potential becomes:

$$K = -\log\left[32 \left(b_1 b_2 b_3 - B\right) b_4^4\right]. \tag{3.74}$$

Also, since we are treating the blow-up modes as constants, the superpotential is, for all intents and purposes, unchanged from its value in eq. (3.35). This is because we can always shift the real part (axion) of ψ_4 to eliminate any constants. We perform the same scalings as before in eqns. (3.37), with an extra shift on ψ_4 to eliminate any constants, giving eq. (3.38). Under such field redefinitions we introduce $\bar{B}$, defined such that: $(b_1 b_2 b_3 - B) \to (b_1 b_2 b_3 - \bar{B})$.

Let us focus on the case in which the axions are vanishing and $b_1 = b_2 = b_3$, leaving 2 moduli: b_1 and b_4. We find the potential:

$$\begin{aligned} V &= V_{\text{flux}}\Big[\left(6\,b_1^2 + 4\,b_4^2 + 2\,b_1^6 - 8\sqrt{2}\,b_1^3\,b_4\right) \\ &+ \bar{B}\left(12\sqrt{2}\,b_4 - 6\,b_1^3 + (4\delta_a - 6)\,/b_1 + 4\sqrt{2}\,\delta_b\,b_4/b_1^2\right) \\ &+ \bar{B}^2\left(6 + 4\,\delta_b/b_1^2 + 6/b_1^4\right)\Big]/[32(b_1^3 - \bar{B})b_4^4], \end{aligned} \tag{3.75}$$

where $\delta_a \equiv \hat{f}_1\hat{f}_2 + \hat{f}_2\hat{f}_3 + \hat{f}_3\hat{f}_1$, $\delta_b \equiv \hat{f}_0(\hat{f}_1 + \hat{f}_2 + \hat{f}_3)$, and V_{flux} is given in eq. (3.40). There are physical constraints: $0 < \bar{B} < b_1^3$. Note that the presence of the $\bar{B}$ parameter breaks the scaling of the model that occurs in the absence of blow-up modes, i.e., scaling only occurs in the $\bar{B} \to 0$ limit. This is because the nonvanishing blow-up modes introduce a non-trivial intersection form, as we encountered previously in the IW model, which prevents the flux parameters from being scaled out completely. However, several flux parameters can still be eliminated for finite $\bar{B}$. Note that in the limit $\bar{B} \to 0$ this potential gives precisely the potential in eq. (3.39) with $a_1 = a_4 = 0$.

Our numerical investigation into V of eq. (3.75) has again yielded no inflating region, despite the presence of the tunable parameter $\bar{B}$.

3.4 Discussion and Conclusions

We have presented an explicit investigation into three explicit string models. Although this represents only a rather small part of the landscape, this acts as a starting point for further investigation into moduli driven inflation. The non-string theorist should note that despite the inherent complexity of string theory, M-theory etc, it is possible to strip down the physics in 4 dimensions to familiar territory. eq. (3.15) gives a familiar 4-dimensional action for n scalar fields minimally coupled to gravity. We note, however, that the kinetic energy is in general non-canonical since $K_{i\bar{j}}$ is typically not equal to $\delta_{i\bar{j}}$ and furthermore the geometric moduli and axions appear in the action differently.[18] We have proceeded in the usual fashion to check for inflation by examining the slow-roll conditions (3.18)–(3.19).

We have not found inflation in any of the specific models presented. In the absence of blow-up modes, the DGKT, VZ and IW models involved 8, 14 and 12 real-valued inflaton fields, respectively, making a full numerical exploration of the inflaton potential $V(\phi)$ (which can also be flux dependent) computationally challenging. We

[18] If we ignore the axions and focus on a certain class of simple models, we can perform field redefinitions to obtain a canonical action, as given in eq. (3.30).

have therefore performed as comprehensive a search as feasible given our available resources:

- For the DGKT model we derived an analytic expression for the 8-dimensional potential $V(\phi)$, finding $V(\phi)$ to be flux independent (up to an overall scale), and searched the 8-dimensional moduli space for vacua, finding one AdS vacuum in addition to the known SUSY vacuum from [39].

- For the VZ model, we derived an analytic expression for the 14-dimensional potential $V(\phi)$, finding that $V(\phi)$ depended on the fluxes via a single parameter s (up to an overall scale). We found the potential to be invariant under permutation of two triplets of complex moduli, and searched the full 6-dimensional subspace corresponding to $\psi_1 = \psi_2 = \psi_3$, $\psi_5 = \psi_6 = \psi_7$ for vacua for the cases $s \in \{0, 1/2, 1, 2, 5, \infty\}$, finding three new AdS vacua in addition to the known SUSY vacuum from [40].

- For the IW model, we derived an analytic expression for the 12-dimensional potential $V(\phi)$, finding that $V(\phi)$ depended on the fluxes via a single parameter t (up to an overall scale). We searched the 6-dimensional subspace corresponding to vanishing axions, finding no new vacua in addition to the five AdS vacua reported by [41] for various flux sign combinations.

We performed this search for vacua using the numerical packages Mathematica[19] and Singular[20] to algebraically solve (using Gröbner bases) the set of high order coupled polynomial equations that follow from setting $\nabla V = 0$.

We then performed a numerical investigation of the slow-roll conditions, evaluating slow roll parameters for the three models on a multi-dimensional grid (of dimension 8, 6, and 6,[21] respectively), involving of order 10^9 grid points each, and found $\epsilon > 1$ for all grid points where $V > 0$. Although we cannot claim to have a *proof* of the non-existence of inflation in these models, as the moduli space is rather large and the potentials V are rather complicated, we do suspect this to be true. We also performed a partial investigation into the consequences of (frozen) blow-up modes, as described in Section 3.3.5, again finding no inflation.

In the type of models presented we have identified at least three obstacles to realizing inflation: the vacua are AdS, there is a logarithmic Kähler potential K, and suitable field redefinitions allow one to scale many of the fluxes out of the potential. None of these features forbid a realization of inflation. Each is probably a reflection of the simple starting point we took, studying models closely based on toroidal compactification and focussing on the moduli of the underlying torus. It is certainly known that each of these three points may be avoided in other regions of the landscape. Nevertheless, our result does underscore that slow-roll inflation may be a rare and delicate phenomenon in the landscape. We will now discuss each of the three obstacles in turn.

[19] http://www.wolfram.com

[20] http://www.singular.uni-kl.de

[21] In fact we did a little more than this: In the IW model we did not *fully* include the axions, which would be a 12 dimensional space, but did so *partially*.

3.4.1 The Potential Energy Challenge

As we discussed earlier in Section 3.2.4 it is somewhat difficult to realize de Sitter vacua with $V > 0$ in string theory. If we break supersymmetry, then existing analyses suggest that such vacua are rare, but plentiful in absolute number. Let us truncate our discussion here to supersymmetric vacua, which we know must not be de Sitter. A good starting point would be Minkowki vacua which are allowed. Again focusing on toroidal orientifolds in type IIA string theory, a detailed investigation is given in [61], in which a host of fluxes are included. In addition to the geometric fluxes that we have described, they turn on so-called non-geometric fluxes [62], and additionally turn on fluxes associated with S-duality (strong – weak coupling duality). In this framework, although they are non-generic, Minkowski vacua are explicitly found (see also [63]). This may be a good starting point for considering inflation models where the inflaton eventually settles down to zero energy density. However, the Minkowski vacua given in [61] are not under good perturbative control. In other words, it is expected that there are large α' and g_s (loop) corrections to the potential. This is in contrast to the models we have investigated in this chapter. In each case we could dial the fluxes in a particular fashion so that all quantum corrections were small. This justifies the supergravity treatment and makes the results of our investigation particularly informative.

3.4.2 The Kinetic Energy Challenge

Let us turn to the form of the kinetic energy, which is governed by the Kähler potential K. As we have pointed out several times, in supergravity models this is typically logarithmic. The Hessian matrix of K determines the form of the kinetic terms. The tree-level form of this for the diagonal torus model is given in eq. (3.24). At the level of supergravity (i.e., ignoring quantum corrections) this form is rather generic for non-torus models also [64]. So for instance, this kind of metric on moduli space will generically occur for the volume modulus b. Let us write this form as:

$$T \sim -\partial_\mu(\log b)\partial^\mu(\log b) \tag{3.76}$$

(suppressing factors of $\bar{m}_{\rm Pl}$). Although we shall not go through the explicit details here, this is fairly simple to show from the fact that in performing the dimensional reduction from 10 to 4 dimensions, we pick up factors of the volume modulus. In order to move to the Einstein frame, we must then compute the transformation of the Ricci scalar, which is a contraction of the Riemann tensor. Since the mixed derivative terms of the Riemann tensor are proportional to the Christoffel symbols, and since the Christoffel symbols essentially perform *logarithmic* derivatives of the metric, the result in (3.76) follows.

For models involving fluxes etc., it is rather generic that the potential V be given (at large volume) by some rational function of b (or more generally of the full set of Kähler moduli). We have given several explicit examples of this in this chapter. Let

us drastically simplify the form by writing

$$V \sim \sum_i c_i b^{-k_i} \tag{3.77}$$

where the k_i are positive integers and the c_i are some coefficients.[22] We note that by defining $\phi \equiv \log b$, the kinetic energy is in canonical form and $V \sim \sum_i c_i e^{-k_i \phi}$. (This property was alluded to earlier in eq. (3.30)). Of course exponentials have slow-roll parameters $\epsilon \sim \eta \sim k^2$, which are not generically small for positive integers k. A recent discussion of how one can perhaps construct working models by fine-tuning similar potentials with several terms appears in [66]. We point out that in [66], only the volume modulus is considered and other moduli are treated as fixed; we have seen explicit examples in our models where although one partial derivative of the potential may be small, another one will often be large, hence making ϵ large and spoiling inflation.

α' Corrections

Let us comment now on the effect that α' corrections have. According to [67], in type IIB string theory there exists an α' correction to the Kähler potential from an $\mathcal{O}(R^4)$ term in the 10-dimensional action. In units where $2\pi\alpha' = 1$, the piece coming from the volume modulus is found to be

$$K = -2 \ln \left[(2b)^{3/2} + \hat{\xi} \right] \tag{3.78}$$

where $\hat{\xi} = -\zeta(3)\chi e^{-3\phi/2}/4$ with χ the Euler characteristic of the Calabi-Yau. The dilaton ϕ, and hence $\hat{\xi}$, is assumed to be fixed. Although we are considering IIA orientifolds (which are not directly related to IIB models, because of the flux), let us imagine the effects of a similar correction in our context. Our models have χ which is of $\mathcal{O}(1 - 100)$ – although the torus has vanishing χ, the fixed points of the orbifold group action introduce blow-up modes that generate non-trivial χ. However, in the regime where our classical analysis is trustworthy (and by choosing sufficiently large fluxes we can make it arbitrarily reliable [39]), we can neglect this effect. One could imagine that for more general Calabi-Yau's χ, and hence the $\hat{\xi}$ correction, is sometimes large.

It is known that such a term can indeed be important for inflation, see e.g. [68]. For $b^{3/2} \gg \hat{\xi}$, this correction is irrelevant and we return to the previous analysis. For $b^{3/2} \ll \hat{\xi}$, however, this changes the situation considerably. In this case, one finds that the kinetic term for b is modified to, roughly

$$T \sim -\frac{1}{\hat{\xi}\sqrt{b}} (\partial_\mu b)^2. \tag{3.79}$$

[22] Locality in the extra dimensions allows one to prove that any contribution to the Einstein-frame potential should fall at least as quickly as $1/r^6$ at large radius r for the extra dimensions (as described on e.g. page 12 of [65]). This puts a bound on the $|k_i|$, and explains why we have disallowed contributions which grow at large radius.

So the inverse of the metric on moduli space is $K^{i\bar{j}} \sim \hat{\xi}\sqrt{b}$. The inclusion of α' corrections into the Kähler potential also induces corrections to the potential V through the supergravity formula. However, let us again assume the form for V as given in eq. (3.77) for the purposes of illustration. We find that generically $\epsilon \sim \eta \sim k^2\hat{\xi}/b^{3/2}$, which is large in the assumed regime $b^{3/2} \ll \hat{\xi}$. Ref. [68] shows that even in the setups containing non-perturbative corrections, e.g., race-track etc, this α' correction usually makes achieving inflation harder or impossible. Of course this entire discussion should be viewed with caution: in the regime where the $\hat{\xi}$-correction has a significant effect, one would have to carefully justify any analysis which neglects additional α' corrections.

Approximate Kähler Potentials

One further comment on the form of the kinetic energy comes from "inflation in supergravity" treatments, e.g., [69]. There it is often assumed that the Kähler potential takes on the minimal form: $K = \phi^*\phi$ (giving $K_{i\bar{j}} = \delta_{i\bar{j}}$). This form does not literally occur in any string compactifications that we are aware of. It can appear as an approximate Kähler potential in models where one fixes the moduli of the compactification manifold and expands the Kähler potential for brane position moduli (or, sometimes, axions) in a Taylor expansion. So we believe that realizing inflation in these scenarios should be taken with a grain of salt, subject to justifying the appearance of the relevant K, for the relevant range of field space, in a model with fixed moduli.

3.4.3 The Challenge of Fluxes Scaling Out

Turning to the issue of scaling out fluxes, although this occurs in the DGKT model if one neglects blow-up modes and focuses on untwisted moduli, it did not occur in full in the other VZ and IW models. In the DGKT model, neglecting the blow-up modes, every member of the infinite set of vacua was identical from the point of view of the slow-roll conditions, and there was large, but not complete, degeneracy in the other models. Degeneracy is reduced in the presence of blow-up modes. In general though, the ability to exploit scaling and shift symmetries reduces the freedom allowed in dialing the shape of a potential in the landscape. Much like the relation whereby unbroken supersymmetry generically implies AdS vacua, or the simple geometric arguments which determine the logarithmic form of the Kähler potential, this all points to the idea that the landscape, although extremely large, has *structure*. On the other hand, the relatively simple form of flux potentials for toroidal moduli, which is behind the existence of some of these scaling symmetries, would not persist in generic Calabi-Yau models. Therefore, it is reasonable to postulate that the degeneracy we found may be an artifact of the particular simple models we have examined.

3.4.4 Outlook

Alternatively, it could be that the type of construction discussed in the introduction, where inflation is realized through a combination of ingredients, including non-perturbative corrections to the superpotential, is more promising. For example, let us make a comment on the N-flation idea [30, 31], which requires N massless axions at the perturbative level, whose mass is then generated by non-perturbative effects. We have seen one flat direction of the axions in the untwisted modes of the VZ model, and the papers [39, 30, 31] discuss how one can have $N \gg 1$ for more complicated compact spaces. However, various model building assumptions made in [30, 31] can certainly be questioned, and an explicit realization of this class of scenarios is important to unravel.

In summary, our work should be viewed as a starting point for a much more general study into inflation driven by computable flux potentials. One obvious next step would be to study a similar class of problems in more general Calabi-Yau manifolds, rather than orbifolds of the torus. The more complicated structure of the internal geometry should translate into richer flux potentials, which could solve some of the problems we found in the toroidal models. Another approach could be to develop statistical arguments along the lines of [70] to quantify how generically or non-generically one expects to find inflation in flux vacua.

3.5 Appendix: Potential Functions and Inflationary Predictions

3.5.1 DGKT Potential

In the DGKT setup, there are 4 complex pairs of moduli: a_1, $b_1, \ldots, a_4$, b_4. Using the Kähler potential of eq. (3.34) and the superpotential of eq. (3.38) we find the following 4-dimensional potential:

$$\begin{aligned} V = V_{\text{flux}}\big[& 2(a_1+a_2+a_3+2\sqrt{2}\,a_4)^2 - 4\,\delta\,a_1\,a_2\,a_3(a_1+a_2+a_3+2\sqrt{2}\,a_4) \\ & + 2\,a_1^2\,a_2^2\,a_3^2 + 2(b_1^2+b_2^2+b_3^2) + 4\,b_4^2 - 4\,\delta\,(a_2\,a_3\,b_1^2 + a_1\,a_3\,b_2^2 + a_1\,a_2\,b_3^2) \\ & + 2(a_2^2\,a_3^2\,b_1^2 + a_1^2\,a_3^2\,b_2^2 + a_1^2\,a_2^2\,b_3^2) + 2(a_3^2\,b_1^2\,b_2^2 + a_2^2\,b_1^2\,b_3^2 + a_1^2\,b_2^2\,b_3^2) \\ & + 2\,b_1^2\,b_2^2\,b_3^2 - 8\sqrt{2}\,b_1\,b_2\,b_3\,b_4\big]/(32\,b_1\,b_2\,b_3\,b_4^4), \end{aligned} \tag{3.80}$$

with V_{flux} given in (3.40) and $\delta = \hat{f}_0\hat{f}_{4,1}\hat{f}_{4,2}\hat{f}_{4,3} = \pm 1$. The slow-roll parameter ϵ is given by (3.27) with $n_1 = n_2 = n_3 = 1$, $n_4 = 4$.

3.5.2 VZ Potential

In the VZ setup, there are 7 complex pairs of moduli: a_1, $b_1, \ldots, a_7$, b_7. Let us focus on the symmetric case $a_1 = a_2 = a_3$, $b_1 = b_2 = b_3$, $a_5 = a_6 = a_7$, $b_5 = b_6 = b_7$. Using the Kähler potential of eq. (3.47) and the superpotential of eq. (3.53) in the symmetric case, we find the following 4-dimensional potential:

$$\begin{aligned} V = V_{\text{flux}}\ [\ & 36a_1^6 - 72sa_1^5 + 36s^2a_1^4 + 108b_1^2a_1^4 + 144s\hat{a}_4a_1^4 + 216a_1^4 - 144sb_1^2a_1^3 + 144sa_1^3 \\ & - 144s^2\hat{a}_4a_1^3 - 48\hat{a}_4a_1^3 + 108b_1^4a_1^2 - 360s^2a_1^2 + 144s^2\hat{a}_4^2a_1^2 + 48s^2b_1^2a_1^2 \\ & + 144s\hat{a}_4b_1^2a_1^2 + 216b_1^2a_1^2 + 144s^2b_4^2a_1^2 + 432s^2b_5^2a_1^2 + 480s\hat{a}_4a_1^2 + 324a_1^2 \\ & - 72sb_1^4a_1 - 96s\hat{a}_4^2a_1 - 144sb_1^2a_1 - 96s^2\hat{a}_4b_1^2a_1 - 96sb_4^2a_1 - 288sb_5^2a_1 \\ & + 1080sa_1 + 720s^2\hat{a}_4a_1 - 144\hat{a}_4a_1 + 36b_1^6 + 12s^2b_1^4 + 900s^2 + 16\hat{a}_4^2 + 48s^2\hat{a}_4^2b_1^2 \\ & + 144s\hat{a}_4b_1^2 + 108b_1^2 + 48s^2b_1^2b_4^2 + 16b_4^2 - 432s^2b_1^2b_5^2 + 48b_5^2 - 240s\hat{a}_4 \\ & + 48s^2b_1^3b_4 - 48b_1^3b_4 + 144s^2b_1^3b_5 - 144b_1^3b_5 - 576s^2b_1^2b_4b_5]/(b_1^3b_4b_5^3), \end{aligned} \tag{3.81}$$

with V_{flux} given in (3.54), $\hat{a}_4 = a_4 + 3a_5$, and $s = p_1/p_2$. The slow-roll parameter ϵ is given by (3.27) with $n_1 = \ldots = n_7 = 1$.

3.5.3 IW Potential

In the IW setup, there are 6 complex pairs of moduli: a_1, $b_1, \ldots, a_6$, b_6. Using the Kähler potential of eq. (3.61) and the superpotential of eq. (3.64), we find the following

4-dimensional potential:

$$
\begin{aligned}
V = V_{\text{flux}} \big[& \, a_3^2 a_4^4/2 + b_3^2 a_4^4/2 - 2a_1 a_2 a_3^2 a_4^2 - 2a_1 a_2 b_3^2 a_4^2 + 2b_1 b_2 b_3^2 a_4^2 + a_3^2 b_4^2 a_4^2 + b_3^2 b_4^2 a_4^2 \\
& + 2\delta a_2 a_3 a_4^2 + 2a_3^2 b_1 b_2 a_4^2 + 2\delta_{30} a_3^2 a_4^2 + 2\delta_{30} b_3^2 a_4^2 + 2\delta_{30} a_1 a_3 a_4^2 \\
& + 4\sqrt{2}\delta_{30} a_3 a_5 a_4^2 + 4\sqrt{2}\delta_{30} a_3 a_6 a_4^2 - 4a_2 b_1 b_3^2 b_4 a_4 - 4a_1 b_2 b_3^2 b_4 a_4 \\
& - 4a_2 a_3^2 b_1 b_4 a_4 - 4a_1 a_3^2 b_2 b_4 a_4 + 4\delta a_3 b_2 b_4 a_4 + 4\delta_{30} a_3 b_1 b_4 a_4 + a_3^2 b_4^4/2 \\
& + b_3^2 b_4^4/2 + 2a_1^2 + 2a_2^2 + 2a_1^2 a_2^2 a_3^2 + 2a_3^2 + 16a_5^2 + 16a_6^2 + 2a_2^2 a_3^2 b_1^2 + 2b_1^2 \\
& + 2a_1^2 a_3^2 b_2^2 + 2a_3^2 b_1^2 b_2^2 - 4\delta a_1 a_3 b_2^2 + 2b_2^2 + 2a_1^2 a_2^2 b_3^2 + 2a_2^2 b_1^2 b_3^2 + 2a_1^2 b_2^2 b_3^2 \\
& + 2b_1^2 b_2^2 b_3^2 + 2b_3^2 + 2a_1 a_2 a_3^2 b_4^2 + 2a_1 a_2 b_3^2 b_4^2 - 2b_1 b_2 b_3^2 b_4^2 - 2\delta a_2 a_3 b_4^2 \\
& - 2a_3^2 b_1 b_2 b_4^2 + 8b_5^2 + 8b_6^2 - 4\delta a_1 a_2^2 a_3 + 4a_1 a_3 + 8\sqrt{2} a_1 a_5 + 8\sqrt{2} a_3 a_5 \\
& + 8\sqrt{2} a_1 a_6 + 8\sqrt{2} a_3 a_6 + 32 a_5 a_6 + 4\sqrt{2} b_3 b_4^2 b_5 - 8\sqrt{2} b_1 b_2 b_3 b_5 + 4\sqrt{2} b_3 b_4^2 b_6 \\
& - 8\sqrt{2} b_1 b_2 b_3 b_6 + 2\delta_{12} b_4^2 + 4\delta_{12} a_1 a_2 + 4\delta_{12} a_2 a_3 + 8\sqrt{2} a_2 a_5 \delta_{12} + 8\sqrt{2}\delta_{12} a_2 a_6 \\
& - 4\delta_{30} a_1 a_2 a_3^2 - 4\delta_{30} a_2 a_3 b_1^2 - 4\delta_{30} a_1 a_2 b_3^2 - 2\delta_{30} a_1 a_3 b_4^2 - 4\delta_{30} a_1^2 a_2 a_3 \\
& - 8\sqrt{2}\delta_{30} a_1 a_2 a_3 a_5 - 8\sqrt{2}\delta_{30} a_1 a_2 a_3 a_6 + (2\delta_{30} a_3 a_4^3 + 4a_1 a_4 + 4a_3 a_4 \\
& + 8\sqrt{2} a_5 a_4 + 8\sqrt{2} a_6 a_4 + 4\delta_{12} a_2 a_4 + 2\delta_{30} a_3 b_4^2 a_4 - 4\delta_{30} a_1 a_2 a_3 a_4 \\
& + 4\delta_{30} a_3 b_1 b_2 a_4 + 4b_1 b_4 + 4\delta_{12} b_2 b_4 - 4\delta_{30} a_2 a_3 b_1 b_4 - 4\delta_{30} a_1 a_3 b_2 b_4) t \\
& + (2a_4^2 + b_4^2 + 2b_1 b_2) t^2 \big] / [2\,(b_1 b_2 - b_4^2/2) b_3 b_5^2 b_6^2], \qquad (3.82)
\end{aligned}
$$

with V_{flux} given in (3.67), t given in (3.65), $\delta = \hat{f}_0 \hat{f}_{4,1} \hat{f}_{4,2} \hat{f}_{4,3} = \pm 1$, $\delta_{12} = \hat{f}_{4,1} \hat{f}_{4,2} = \pm 1$, $\delta_{30} = \hat{f}_{4,3} \hat{f}_0 = \pm 1$. The slow-roll parameter ϵ is given by (3.69).

3.5.4 Cosmological Parameters from Slow-Roll Inflation

The mathematical prescription in this Section allows one to compute cosmological parameters corresponding to an arbitrary string potential without understanding the derivation or interpretation of the results.

Suppose from some string model we are given a potential energy function $\bar{V}$ of some complex scalar fields ψ^i in the Einstein frame (see eq. (3.15)) and a Kähler potential K. For example, $\bar{V}$ may be given by the supergravity formula in eq. (3.11) complemented by the rescaling eq. (3.14). We then compute the following slow-roll parameters:

$$
\epsilon = \frac{K^{i\bar{j}} \bar{V}_{,i} \bar{V}_{,\bar{j}}}{\bar{V}^2} \left(= \frac{g^{ab} \bar{V}_{,a} \bar{V}_{,b}}{2\bar{V}^2} \right), \qquad (3.83)
$$

$$
\eta = \text{min eigenvalue} \left\{ \frac{g^{ab} \left(\bar{V}_{,bc} - \Gamma^d_{bc} \bar{V}_{,d} \right)}{\bar{V}} \right\}, \qquad (3.84)
$$

where η (and ϵ) is written in terms of the metric g_{ab} governing real scalar fields ϕ^a: $K_{i\bar{j}} \partial_\mu \psi^i \partial^\mu \psi^{\bar{j}} = \frac{1}{2} g_{ab} \partial_\mu \phi^a \partial^\mu \phi^b$, with $\phi^{2i-1} = \text{Re}[\psi^i]$ and $\phi^{2i} = \text{Im}[\psi^i]$.

The universe inflates until a time t_e when the slow-roll conditions ($\epsilon < 1$, $|\eta| < 1$)

are no longer satisfied. The number of e-foldings from time t to t_e is defined by

$$N = \int_t^{t_e} dt\, H. \tag{3.85}$$

All the cosmological parameters defined below are a function of N. A good value to use is 55 (see [71]), with a reasonable range being $50 < N < 60$.

According to inflation, several cosmological parameters can be computed as follows:

$$\begin{aligned}
Q_s &= \sqrt{\frac{\bar{V}}{150\pi^2 \bar{m}_{\rm Pl}^4 \epsilon}}, && (3.86)\\
n_s &= 1 - \partial_N \ln Q_s^2 = 1 - 6\epsilon + 2\eta, && (3.87)\\
\alpha_s &= -\partial_N^2 \ln Q_s^2, && (3.88)\\
Q_t &= \sqrt{\frac{\bar{V}}{75\pi^2 \bar{m}_{\rm Pl}^4}}, \quad r \equiv \left(\frac{Q_t}{Q_s}\right)^2 = 16\epsilon, && (3.89)\\
n_t &= -\partial_N Q_t^2 = -2\epsilon, && (3.90)
\end{aligned}$$

which corresponds to the amplitude, spectral index, and running of spectral index of scalar fluctuations, and the amplitude and spectral index of tensor fluctuations, respectively. The expressions giving Q_t and n_t have general validity. In contrast, the expressions for Q_s, n_s and α_s are good approximations for the most studied cases of multi-field inflation in the literature, where the walls of the multi-dimensional gorge in which the inflaton slowly rolls are much steeper than the roll direction, but do not hold more generally. The expression for Q_s always provides a lower limit on the correct value.

The predictions for these cosmological parameters can be directly compared with with observation. The most recent constraints from combining WMAP (Wilkinson Microwave Anisotropy Probe) microwave background data with SDSS (Sloan Digital Sky Survey) galaxy clustering data are [2]

$$\begin{aligned}
Q_s &= 1.945^{+0.051}_{-0.053} \times 10^{-5}, && (3.91)\\
n_s &= 0.953^{+0.016}_{-0.016}, && (3.92)\\
\alpha_s &= -0.040^{+0.027}_{-0.027}, && (3.93)\\
r &< 0.30\,(95\%), && (3.94)\\
n_t + 1 &= 0.9861^{+0.0096}_{-0.0142}. && (3.95)
\end{aligned}$$

Bibliography

[1] D. N. Spergel et. al. "Wilkinson Microwave Anisotropy Probe (WMAP) Three Year Results: Implications for Cosmology," (2006), [arXiv:astro-ph/0603449v2].

[2] M. Tegmark, "Cosmological Constraints from the SDSS Luminous Red Galaxies," Phys. Rev. D74 123507 (2006) [arXiv:astro-ph/0608632v2].

[3] A.G. Sanchez et al., "Cosmological Parameters from Cosmic Microwave Background Measurements and the Final 2dF Galaxy Redshift Survey Power Spectrum," MNRAS **366** 189 (2006).

[4] M. Tegmark, "What Does Inflation Really Predict?," JCAP, (2005), [arXiv:astro-ph/0410281].

[5] E. Silverstein, "TASI/PiTP/ISS Lectures on Moduli and Microphysics," in Boulder 2003, Progress in string theory, J. Maldacena ed., [arXiv:hep-th/0405068].

[6] M. Grana, "Flux Compactifications in String Theory: A Comprehensive Review," Phys. Rept. **423** 91 (2006), [arXiv:hep-th/0509003].

[7] M. Douglas and S. Kachru, "Flux Compactification," Rev. Mod. Phys. **79** 733 (2007), [arXiv:hep-th/0610102].

[8] R. Blumenhagen, B. Körs, D. Lust and S. Stieberger, "Four-Dimensional String Compactifications with D-branes, Orientifolds and Fluxes," Phys. Rept. **445** 1 (2007), [arXiv:hep-th/0610327].

[9] J. Kumar, "A Review of Distributions on the String Landscape," Int. J. Mod. Phys. **A21** 3441 (2006), [hep-th/0601053].

[10] J. Conlon, "Moduli Stabilisation and Applications in IIB String Theory," Fortsch. Phys. **55** 287 (2007), [arXiv:hep-th/0611039].

[11] P. Binetruy and M. Gaillard, "Candidates for the Inflaton Field in Superstring Models," Phys. Rev. **D34** 3069 (1986).

[12] T. Banks, M. Berkooz, S. H. Shenker, G. Moore and P. J. Steinhardt, "Modular Cosmology," Phys. Rev. **D52** 3548 (1995), [arXiv:hep-th/9503114].

[13] G. Dvali and S.-H. Tye, "Brane Inflation," Phys. Lett. B450 (1999) 72-82, [arXiv:hep-ph/9812483v1].

[14] C.P. Burgess, P. Martineau, F. Quevedo, G. Rajesh and R.J. Zhang, "Brane-antibrane Inflation in Orbifold and Orientifold Models," JHEP **0203** 052 (2002), [arXiv:hep-th/0111025].

[15] B. Dutta, J. Kumar and L. Leblond, "An Inflationary Scenario in Intersecting Brane Models," JHEP **0707** 045 (2007), [arXiv:hep-th/0703278].

[16] A. Krause and E. Pajer, "Chasing Brane Inflation in String Theory," (2007) [arXiv:hep-th:0705.4682].

[17] S. Panda, M. Sami, S. Tsujikawa, "Prospects of Inflation in Delicate D-brane Cosmology," arXiv:0707.2848.

[18] R. Bean, S. E. Shandera, S. H. Tye, and J. Xu, "Comparing Brane Inflation to WMAP," JCAP 05 (2007) 004, arXiv:hep-th/0702107.

[19] S. Kachru, R. Kallosh, A. Linde, J. Maldacena, L. McAllister, and S. P. Trivedi, "Towards Inflation in String Theory," JCAP 0310 (2003) 013 [arXiv:hep-th/0308055v2].

[20] D. Baumann, A. Dymarsky, I. R. Klebanov, L. McAllister, and P. J. Steinhardt, "A Delicate Universe," (2007) [arXiv:hep-th/0705.3837v1];

[21] D. Baumann, A. Dymarsky, I. R. Klebanov, and L. McAllister, "Towards an Explicit Model of D-brane Inflation," (2007) [arXiv:hep-th/0706.0360v1].

[22] U. Guenther and A. Zhuk, "Stabilization of Internal Spaces in Multidimensional Cosmology," [arXiv:hep-ph/0002009].

[23] B. de Carlos, J. A. Casas, A. Guarino, J. M. Moreno, and O. Seto "Inflation in Uplifted Supergravities," arXiv:hep-th/0702103.

[24] C.P. Burgess, "Lectures on Cosmic Inflation and its Potential Stringy Realizations," [arXiv:hep-th/0708.2865].

[25] R. Kallosh, "On inflation in String Theory," (2007), [arXiv:hep-th/0702059]; J. Cline, "String Cosmology," (2006), [arXiv:hep-th/061219].

[26] S.H. Henry Tye, "Brane inflation: String Theory Viewed from the Cosmos," (2006), [arXiv:hep-th/0610221].

[27] L. McAllister and E. Silverstein, "String cosmology: A review," to appear.

[28] S. Kachru, R. Kallosh, A. Linde, and S. P. Trivedi, "de Sitter Vacua in String Theory," Phys. Rev. D68, 046005 (2003), [arXiv:hep-th/0301240].

[29] E. Silverstein, "(A)dS Backgrounds from Asymmetric Orientifolds," (2001), [arXiv:hep-th/0106209].

[30] S. Dimopoulos, S. Kachru, J. McGreevy, J. Wacker, "N-flation," (2005), [arXiv:hep-th/0507205v3].

[31] R. Easther and L. McAllister, "Random Matrices and the Spectrum of N-flation," JCAP **0605** 018 (2006), [arXiv:hep-th/0512102].

[32] J. Conlon and F. Quevedo, "Kähler Moduli Inflation," JHEP 0601 146 (2006), [arXiv:hep-th/0509012v2].

[33] J. R. Bond, L. Kofman, S. Prokushkin, P. Vaudrevange, "Roulette Inflation with Kähler Moduli and their Axions," Phys. Rev. **D75** 123511 (2007), [arXiv:hep-th/0612197].

[34] J. Simon, R. Jimenez, L. Verde, P. Berglund and V. Balasubramanian, "Using Cosmology to Constrain the Topology of Hidden Dimensions," (2006) [arXiv:astro-ph/0605371].

[35] J. J. Blanco-Pillado, C. P. Burgess, J. M. Cline, C. Escoda, M. Gómez-Reino, R. Kallosh, A. Linde, and F. Quevedo, "Inflating in a Better Racetrack," JHEP0609 002 (2006) [arXiv:hep-th/0603129v2].

[36] E. Silverstein and D. Tong, "Scalar Speed Limits and Cosmology: Acceleration from D-cceleration," Phys. Rev. **D70** 103505 (2004), [arXiv:hep-th/0310221].

[37] T. W. Grimm and J. Louis, "The Effective Action of Type IIA Calabi-Yau Orientifolds," (2005), [arxiv:hep-th/0412277].

[38] J. P. Derendinger, C. Kounnas, P. M. Petropoulos, and F. Zwirner, "Superpotentials in IIA Compactifications with General Fluxes," Nucl. Phys. B **715** 211 (2005) [arXiv:hep-th/0411276].

[39] O. DeWolfe, A. Giryavets, S. Kachru, and W. Taylor, "Type IIA Moduli Stabilization," J. High Energy Phys. 07 (2005) 066, [arXiv:hep-th/0505160].

[40] G. Villadoro and F. Zwirner, "$\mathcal{N} = 1$ Effective Potential from Type-IIA D6/O6 Orientifolds with General Fluxes," (2005), [arXiv:hep-th/0503169].

[41] M. Ihl and T. Wrase, "Towards a Realistic Type IIA $T^6/\mathbb{Z}_4$ Orientifold Model with Background Fluxes, Part 1: Moduli Stabilization," (2006), [arXiv:hep-th/0604087].

[42] F. Denef, M. Douglas and S. Kachru, "Physics of String Flux Compactifications," (2007), [arXiv: hep-th/0701050].

[43] J. Polchinski, "String Theory, Vol II, Superstring Theory and Beyond," Cambridge University Press, (1998).

[44] J. Polchinski, "Dirichlet-Branes and Ramond-Ramond Charges," Phys. Rev. Lett. 75 4724-4727 (1995), [arXiv:hep-th/9510017v3].

[45] L. Dixon, J. Harvey, C. Vafa and E. Witten, "Strings on Orbifolds," Nucl. Phys. **B261** 678 (1985).

[46] J. Dai, R.G. Leigh and J. Polchinski, "New Connections Between String Theories," Mod. Phys. Lett. **A4** 2073 (1989).

[47] G. Pradisi and A. Sagnotti, "Open String Orbifolds," Phys. Lett. **B216** (1989) 59.

[48] P. Horava, "Strings on Worldsheet Orbifolds," Nucl. Phys. **B327** 461 (1989).

[49] J. P. Conlon, "The QCD Axion and Moduli Stabilisation," JHEP 0605 078 (2006), [arXiv:hep-th/0602233].

[50] P. Svrcek and E. Witten, "Axions In String Theory," JHEP 0606 051 (2006), [arXiv:hep-th/0605206].

[51] E. Copeland, A. Liddle, D. Lyth, E. Stewart and D. Wands, "False Vacuum Inflation with Einstein Gravity," Phys. Rev. **D49** 6410 (1994), [arXiv:astro-ph/9401011].

[52] L. McAllister, "An Inflaton Mass Problem in String Inflation from Threshold Corrections to Volume Stabilization," JCAP **0602** 010 (2006), [arXiv:hep-th/0502001].

[53] G. Gibbons, "Aspects of Supergravity Theories," in GIFT seminar 0123 (1984).

[54] J. Maldacena and C. Nunez, "Supergravity Description of Field Theories on Curved Manifolds and a No Go Theorem," Int J. Mod. Phys. **A16** 822 (2001), [arXiv:hep-th/0007018].

[55] S. B. Giddings, S. Kachru, and J. Polchinski, "Hierarchies from Fluxes in String Compactifications," (2002), [arXiv:hep-th/0105097].

[56] R. Bousso and J. Polchinski, "Quantization of Four-form Fluxes and Dynamical Neutralization of the Cosmological Constant," JHEP 0006 006 (2000), [arXiv:hep-th/0004134v3].

[57] P. Breitenlohner and D. Z. Freedman, "Positive Energy in Anti-De Sitter Backgrounds and Gauged Extended Supergravity," Phys. Lett. B **115** 197 (1982).

[58] A. Linde, "Chaotic Inflation," Phys. Lett. **B129** 177 (1983).

[59] A. Linde, "Eternal Chaotic Inflation," Mod. Phys. Lett. A1, 81 (1986).

[60] J. P. Derendinger, C. Kounas, P. M. Petropoulos, and F. Zwirner, "Fluxes and Gaugings: $\mathcal{N} = 1$ Effective Superpotentials," (2005), [arXiv:hep-th/0503229].

[61] G. Aldazabal, P. G. Camara, A. Font, and L. E. Ibanez, "More Dual Fluxes and Moduli Fixing," (2006), [arXiv:hep-th/0602089].

[62] J. Shelton, W. Taylor, B. Wecht, "Nongeometric Flux Compactifications," (2005), [arxiv:hep-th/0508133];

[63] K. Becker, M. Becker, C. Vafa and J. Walcher, "Moduli Stabilization in Non-Geometric Backgrounds," Nucl. Phys. **b770** 1 (2007), [arXiv:hep-th/0611001].

[64] P. Candelas and X. de la Ossa, "Moduli Space of Calabi-Yau Manifolds," Nucl. Phys. **B355** 455 (1991).

[65] S. Kachru, J. McGreevy and P. Svrcek, "Bounds on Masses of Bulk Fields in String Compactifications," JHEP **0604** 023 (2006), [arXiv:hep-th/0601111].

[66] N. Itzhaki and E. Kovets, "Inflection Point Inflation and Time Dependent Potentials in String Theory," (2007) [arXiv:hep-th/0708.2798].

[67] K. Becker, M. Becker, M. Haack, and J. Louis, "Supersymmetry Breaking and Alpha–Prime Corrections to Flux Induced Potentials," JHEP 0206:060, (2002) [arXiv:hep-th/0204254].

[68] B. R. Greene and A. Weltman, "An Effect of α' Corrections on Racetrack Inflation," JHEP 0603 035 (2006), [arXiv:hep-th/0512135].

[69] M. Kawasaki, M. Yamaguchi, and T. Yanagida, "Natural Chaotic Inflation in Supergravity," Phys. Rev. Lett. 85, 3572 - 3575 (2000).

[70] F. Denef and M.R. Douglas, "Distributions of Flux Vacua," JHEP **0405** 072 (2004) [arXiv:hep-th/0404116].

[71] A. Liddle and A. Leach, "How Long Before the End of Inflation were Observable Perturbations Produced?" Phys. Rev. **D68** 103503 (2003), [arXiv:astro-ph/0305263].

Chapter 4

Inflationary Constraints on Type IIA String Theory

We prove that inflation is forbidden in the most well understood class of semi-realistic type IIA string compactifications: Calabi-Yau compactifications with only standard NS-NS 3-form flux, R-R fluxes, D6-branes and O6-planes at large volume and small string coupling. With these ingredients, the first slow-roll parameter satisfies $\epsilon \geq \frac{27}{13}$ whenever $V > 0$, ruling out both inflation (including brane/anti-brane inflation) and de Sitter vacua in this limit. Our proof is based on the dependence of the 4-dimensional potential on the volume and dilaton moduli in the presence of fluxes and branes. We also describe broader classes of IIA models which may include cosmologies with inflation and/or de Sitter vacua. The inclusion of extra ingredients, such as NS 5-branes and geometric or non-geometric NS-NS fluxes, evades the assumptions used in deriving the no-go theorem. We focus on NS 5-branes and outline how such ingredients may prove fruitful for cosmology, but we do not provide an explicit model. We contrast the results of our IIA analysis with the rather different situation in IIB.

4.1 Introduction

Our desire to understand the large-scale properties of our Universe is one of the motivations for studying fundamental microphysical theories such as string theory. Indeed, observing the early Universe may be our most promising path toward confronting string theory with data. The leading paradigm for explaining the large-scale isotropy, homogeneity, and flatness of the Universe, as well as its $\mathcal{O}(10^{-5})$ seed fluctuations, is cosmological inflation [1, 2, 3, 4]. Specifically, by assuming that there exist one or more scalar fields that undergo slow rolling in the early Universe in a potential energy function of just the right shape, one can explain these large-scale properties and predict the numerical values of as many as eight cosmological parameters, many of which have now been accurately measured [5, 6, 7]. The leading candidate for a fundamental microphysical theory is string theory, and so we would like to know how generically string theory can accommodate such potential energy functions.

It is rather well known that the conditions for inflation do not arise easily in

string theory, or in other words, that a *generic* point in field space may not be expected to satisfy the slow-roll conditions [8, 9, 10]. In part this is because of issues like the η problem (essentially, that the potential varies too quickly), which also complicate attempts to build inflationary models in quantum field theory and supergravity theories. There are, however, three reasons to suspect a priori that string theory can accommodate inflation: Firstly, the potential energy is typically a function of hundreds of fields, which means that there is a large field space to explore. Secondly, there are an exponentially large number of infinite families of potential energy functions, parameterized by typically hundreds of discrete fluxes in the compact space [11, 12, 13, 14, 15]. Thirdly, there are at least many millions of different topologies, such as Calabi-Yau manifolds, that in general give rise to qualitatively different physical theories in 4 dimensions. It is reasonable to suspect that occasionally in this vast space of possibilities, the conditions for inflation are satisfied.

For this reason, the past few years have seen intense investigation into the possibility of inflation driven by closed string moduli [16, 17, 18, 19, 20], axions [21, 22, 23, 24, 25], or brane positions [26, 27, 28, 29, 30, 31, 32, 33] in the extra dimensions. In the most intensely studied case of IIB string compactifications on Calabi-Yau orientifolds, the conclusion at this point is that one can probably build working models, at the cost of fine-tuning the relevant potentials. Some of these models could even have interesting observable signatures [34, 35]. Reviews of this general subject appear in [36, 37, 38, 39, 40].

One feature of the existing constructions is that they are implicit, relying at some point on either non-compact models of regions of the compactification space, or on the ability to perform tunes which (though seemingly possible based on detailed theoretical considerations) are not performed explicitly. One would ideally like to build simpler models, where all of the calculations are performed in a completely explicit and reliable way. Commonly, compactifications suffer from unstabilized moduli in the low energy description, or the calculations stabilizing moduli apply in a regime that, while apparently numerically controlled, is not under parametric control.[1]

In some limits of string theory, however, we now have explicit examples of stabilized models with parametric control of the moduli potential. The best understood case occurs in massive IIA string theory; namely IIA string theory with R-R 0-form flux, compactified on a Calabi-Yau orientifold. The 10-dimensional massive IIA supergravity action was suggested in [41]. The compactification of this theory on a Calabi-Yau orientifold was performed in a 4-dimensional supergravity formalism in [42] and the stabilization was obtained in [43].[2] The 10-dimensional description of these compactifications was further studied in [45]. Since these models carry at most $\mathcal{N} = 1$ supersymmetry in 4 dimensions, and gauge groups and chiral matter can be

[1]This in particular applies in any class of stabilized compactifications where the number of choices of fluxes, branes, etc., while perhaps very large, is *finite*. In such models, there is perforce a limit on how small g_s can be, namely the smallest g_s obtained in the finite list. Of course if the finite number is sufficiently large, the smallest attainable coupling may be quite small, so this may not be a serious limitation.

[2]With additional ingredients (geometric flux) stabilization was achieved in [44].

incorporated in this context, we consider them to be at least semi-realistic.

An investigation of cosmology in these IIA compactifications was initiated in Ref. [46] by considering some specific simple examples and showing that inflation could not occur in these examples. In the present work we extend that study. We use a simple scaling analysis of various terms which appear in the low-energy 4-dimensional potential to rule out inflation and de Sitter vacua at large volume and weak coupling in all IIA Calabi-Yau models with conventional fluxes, D-branes, and O-planes. This means that inflation imposes the constraint that our Universe is not in this portion of the landscape. We emphasize that the derivation of this no-go result is only valid in the large volume limit and can be evaded by various other structures including NS 5-branes as well as geometric and non-geometric NS-NS fluxes, indicating that IIA compactifications containing these ingredients may be a good place to look for string models with inflation and/or de Sitter vacua. Indeed, as we were completing this work we received a copy of [47], in which more explicit IIA de Sitter models are constructed by using geometric NS-NS fluxes, 5-branes, and various other ingredients.

The structure of this chapter is as follows: In Section 4.2 we summarize the IIA supergravity theory in 10 dimensions and outline the dimensional reduction to 4 dimensions. We explain the key step in the analysis of the chapter, which involves considering 2-dimensional slices in the full moduli space parameterized by the volume and dilaton moduli of the compactification. The behavior of the four (space-time) dimensional potential on these 2-dimensional slices of moduli space allows us to place a lower bound on the slow-roll parameter ϵ. In Section 4.3 we compute the scaling of the various terms appearing in the 4-dimensional potential energy function V in terms of the two model-independent moduli. We prove that both inflation and de Sitter vacua are forbidden at large volume and weak string coupling when standard fluxes, D6-branes, and O6-planes are included; the slow-roll parameter is bounded below in this case by $\epsilon \geq \frac{27}{13}$ whenever $V > 0$. In Section 4.4 we describe some additional ingredients such as NS 5-branes, geometric fluxes and non-geometric NS-NS fluxes which can be included in type IIA and which lead to terms in the 4-dimensional potential with scaling properties allowing us to evade the no-go theorem. In Section 4.5 we discuss the type IIB theory. We show how the structure of the IIB theory differs from the IIA theory from the point of view taken in this chapter and discuss the connection of our IIA results with previous work on inflation in IIB models. We discuss our results in Section 3.4. More details regarding the kinetic energy and potential energy are provided in Appendix 4.7.

4.2 Type IIA Compactifications

We investigate large volume and small string coupling compactifications where it is valid to perform computations using supergravity. We study the 10-dimensional type IIA supergravity theory, where we include conventional NS-NS and R-R field

strengths, as well as D6-branes and O6-planes:

$$S = \frac{1}{2\kappa_{10}^2}\int d^{10}x\sqrt{-g}\,e^{-2\phi}\Big(R + 4(\partial_\mu\phi)^2 - \frac{1}{2}|H_3|^2 - e^{2\phi}\sum_p |F_p|^2\Big)$$
$$-\mu_6\int_{\rm D6} d^7\xi\sqrt{-g}\,e^{-\phi} + 2\mu_6\int_{\rm O6} d^7\xi\sqrt{-g}\,e^{-\phi} \tag{4.1}$$

where R is the 10-dimensional Ricci scalar, ϕ is the scalar dilaton field, H_3 is the NS-NS 3-form field strength that is sourced by strings, F_p are the R-R p-form field strengths ($p = 0, 2, 4, 6$) that are sourced by branes, $\kappa_{10}^2 = 8\pi G_{10}$ is the gravitational strength in 10-dimensions, and μ_6 ($-2\mu_6$) is the D6-brane (O6-plane) charge and tension. We have set all fermions to zero as we are interested in solutions with maximal space-time symmetry. There are also Chern-Simons contributions to the action. These are essentially topological, and are independent of the dilaton as well as the overall scale of the metric (in string frame). We expect that as in [43] the contribution to the action from the Chern-Simons terms will vanish on-shell,[3] so that we need not consider it further here. Although there are some subtle questions regarding the definition of orientifolds in these massive IIA backgrounds and whether these backgrounds can be described in a weak coupling string expansion [48, 49], these compactifications seem to be described adequately in the 4-dimensional supergravity formalism of [42], corresponding to the 10-dimensional massive supergravity analysis when the sources are uniformly distributed in the compactification space.

4.2.1 Compactification

We now perform a Kaluza-Klein compactification of this theory from 10 dimensions to 4 dimensions. Let us first focus on the gravity sector. Assuming that we can neglect any dependence of the Ricci scalar on the compact space coordinates, we can integrate over the compact space, giving

$$\int d^{10}x\sqrt{-g_{10}}\,e^{-2\phi}R = \int d^4x\sqrt{-g_4}\,\mathrm{Vol}\,e^{-2\phi}R \tag{4.2}$$

where Vol is the 6-dimensional volume of the compact space. The volume, dilaton, and all the other fields that describe the size and shape of the compact space are scalar fields in the 4-dimensional description, known as moduli. In addition to kinetic energy terms, the remaining terms in the supergravity action, when reduced to 4 dimensions, describe a potential function V which depends on the moduli and fluxes in any given model.

The key observation of this chapter is that by studying the dependence of the potential energy function V on only two of the moduli, we can learn a great deal about the structure of the potential relevant for the possibility of inflation. We define

[3] More precisely, integrating out the 4-dimensional non-dynamical field dC_3 as a Lagrange multiplier gives an equation which must be satisfied by the axions, but the Chern-Simons terms do not otherwise affect the 4-dimensional potential.

the volume modulus of the compact space ρ and the dilaton modulus τ by[4]

$$\rho \equiv (\text{Vol})^{\frac{1}{3}}, \quad \tau \equiv e^{-\phi}\sqrt{\text{Vol.}} \tag{4.3}$$

While V depends on all moduli, we can explore the behavior of this function on the whole moduli space by considering 2-dimensional slices of the moduli space where all moduli other than τ and ρ are fixed. By showing that V has a large gradient in the τ-ρ plane on every slice wherever V is positive, we will be able to rule out inflation on the *entire* moduli space, regardless of which fields we would like to identify as the inflaton.

In order to bring the gravity sector into canonical form, we perform a conformal transformation on the metric to the so-called Einstein frame:

$$g^E_{\mu\nu} \equiv \frac{\tau^2}{\bar{m}^2_{\text{Pl}}\,\kappa^2_{10}} g^4_{\mu\nu}, \tag{4.4}$$

where $\bar{m}_{\text{Pl}} = 1/\sqrt{8\pi G} \approx 2 \times 10^{18}\,\text{GeV}$ is the (reduced) Planck mass and G is the 4-dimensional Newton constant. By re-expressing R in terms of a 4-dimensional Ricci scalar and performing the conformal transformation, one finds that the gravity sector is canonical and that the fields ρ and τ carry kinetic energy that is diagonal. Although ρ and τ do not have canonical kinetic energies, they are related to fields which do:

$$\hat{\rho} \equiv \sqrt{\frac{3}{2}}\,\bar{m}_{\text{Pl}} \ln \rho, \quad \hat{\tau} \equiv \sqrt{2}\,\bar{m}_{\text{Pl}} \ln \tau. \tag{4.5}$$

Altogether we obtain the effective Lagrangian in 4 dimensions in the Einstein frame

$$\mathcal{L} = \frac{1}{16\pi G} R_E - \left[\frac{1}{2}(\partial_\mu \hat{\rho})^2 + \frac{1}{2}(\partial_\mu \hat{\tau})^2 + \ldots\right] - V \tag{4.6}$$

where the Ricci scalar R_E and all derivatives are defined with respect to the Einstein metric. The dots indicate further kinetic energy terms from all the other fields of the theory associated with the compactification (ϕ_i): the so-called Kähler moduli, complex structure moduli, and axions[5]. The important point is that their contributions will always be *positive*. More details are provided in Appendix 4.7.1. All contributions from the field strengths and D-branes & O-planes from eq. (4.1) are described by some potential energy function V.

4.2.2 The Slow-Roll Condition

From this action we can derive the slow-roll conditions on the potential V for inflation. In order to write down the conditions in detail we would need to know the precise

[4]Note that we have defined the volume modulus in the string frame, since this definition relates to the Kähler moduli in the IIA theory. This differs from the conventional definition of Kähler moduli in analogous IIB orientifolds, where the Einstein frame metric is used in defining the chiral multiplets.

[5]The axions arise from zero-modes of the various gauge fields.

form of the kinetic energy with respect to all the moduli. This can be done cleanly in the 4-dimensional supergravity formalism, as mentioned in Appendix 4.7.1. The important point here is that the first slow-roll parameter ϵ involves partial derivatives of the potential with respect to each direction in field space, and that the contribution from $\hat{\rho}$, $\hat{\tau}$, ϕ_i is non-negative. In fact it is roughly the square of the gradient of $\ln V$. The contributions from $\hat{\rho}$ and $\hat{\tau}$ thus give the following lower bound:

$$\epsilon \geq \frac{\bar{m}_{\rm Pl}^2}{2}\left[\left(\frac{\partial \ln V}{\partial \hat{\rho}}\right)^2 + \left(\frac{\partial \ln V}{\partial \hat{\tau}}\right)^2\right]. \tag{4.7}$$

A *necessary* condition for inflation is $\epsilon \ll 1$ with $V > 0$. We will now prove that this condition is impossible to satisfy in this IIA framework, at large volume and weak coupling where our calculations apply.

4.3 No-Go Theorem

Having discussed the gravity and kinetic energy sector of the theory, let us now describe the form of the potential energy function V. By using the bound on ϵ we will then prove that inflation is forbidden for any Calabi-Yau compactification of type IIA string theory in the large volume and small string coupling limit, when only conventional NS-NS and R-R field strengths, D6-branes and O6-planes are included.

4.3.1 Potential Energy

The potential energy arises from the dimensional reduction of the terms in (4.1) associated with the various field strengths H_3 & F_p ($p = 0, 2, 4, 6$) and the D6-branes & O6-planes. Let us focus on some field strength F_p. Such a field can have a nonvanishing integral over any closed p-dimensional internal manifold (homology cycle) of the compact space, and must satisfy a generalized Dirac charge quantization condition:

$$\int_\Sigma F_p \propto f_\Sigma, \tag{4.8}$$

where f_Σ is an integer associated with number of flux quanta of F_p through each p-dimensional cycle Σ. By choosing different values for f_Σ over a basis set of p-cycles, one obtains a *landscape* of possible allowed potential energy functions V.

The energy arising from a p-form flux F_p comes from a term in (4.1) proportional to $|F_p|^2$; since the p-form transforms as a covariant p-tensor (*i.e.*, has p lower indices), we contract with p factors of $g_6^{\mu\nu}$, so that

$$|F_p|^2 \propto \rho^{-p} \tag{4.9}$$

in the string frame. By including the appropriate factors of the volume and dilaton from the compactification and performing the conformal transformation to the

Einstein frame, we have the following contributions to V:

$$\begin{aligned} V_3 &\propto \rho^{-3}\tau^{-2} \quad \text{for } H_3, \\ V_p &\propto \rho^{3-p}\tau^{-4} \quad \text{for } F_p. \end{aligned} \tag{4.10}$$

We also need the contribution from D6-branes and O6-planes. In the Einstein frame they scale as

$$\begin{aligned} V_{\text{D6}} &\propto \tau^{-3} \quad \text{for D6-branes}, \\ V_{\text{O6}} &\propto -\tau^{-3} \quad \text{for O6-planes}, \end{aligned} \tag{4.11}$$

where we have indicated that O6-planes provide a negative contribution, while all others are positive.

Altogether, we have the following expression for the scalar potential in 4-dimensions

$$\begin{aligned} V &= V_3 + \sum_p V_p + V_{\text{D6}} + V_{\text{O6}} \\ &= \frac{A_3(\phi_i)}{\rho^3\tau^2} + \sum_p \frac{A_p(\phi_i)}{\rho^{p-3}\tau^4} + \frac{A_{\text{D6}}(\phi_i)}{\tau^3} - \frac{A_{\text{O6}}(\phi_i)}{\tau^3}. \end{aligned} \tag{4.12}$$

Here we have written the various coefficients as A_j (≥ 0), which in general are complicated functions of all the other fields of the theory ϕ_i, namely the remaining set of Kähler moduli, complex structure moduli, and axions. The coefficients A_j also depend on the choice of flux integers f_Σ. This means that V is in general a function of hundreds of fields, for each of the exponentially large number of infinite families of possible flux combinations on each of the many available Calabi-Yau manifolds. We have simply described its dependence on two of the fields: ρ and τ. An alternative proof of eq. (4.12) from the perspective of the 4-dimensional supergravity formalism is given in Appendix 4.7.2. As discussed in Ref. [43] this potential ensures that there exist special points in field space which stabilize all the geometric moduli and many axions. We are interested, however, in exploring the full moduli space in search of inflation.

4.3.2 Proof of No-Go Theorem

We find that for any potential of the class we have constructed so far, inflation is impossible anywhere in the moduli space. The proof of this result is quite simple. The key is to observe that the potential in eq. (4.12) satisfies

$$\begin{aligned} -\rho\frac{\partial V}{\partial \rho} - 3\tau\frac{\partial V}{\partial \tau} &= 9V + \sum_p p\, V_p \\ &\geq 9V, \end{aligned} \tag{4.13}$$

where the inequality comes from the fact that all $V_p \geq 0$ ($A_p \geq 0$) and $p \geq 0$. Now, assuming we are in a region where $V > 0$, which is necessary for inflation, we can

divide both sides by V and rewrite this inequality in terms of $\hat{\rho}$ and $\hat{\tau}$ as

$$\bar{m}_{\rm Pl}\left|\sqrt{\frac{3}{2}}\left(\frac{\partial \ln V}{\partial \hat{\rho}}\right)+3\sqrt{2}\left(\frac{\partial \ln V}{\partial \hat{\tau}}\right)\right| \geq 9. \tag{4.14}$$

This implies that it is impossible for both terms in eq. (4.7) to be simultaneously small. Specifically, comparing this inequality with eq. (4.7), we see that $\sqrt{2\epsilon}/\bar{m}_{\rm Pl}$ is the distance to the origin on the plane spanned by $(\partial \ln V/\partial\hat{\rho}, \partial \ln V/\partial\hat{\tau})$, while a sloped band around the origin is forbidden, implying the existence of a lower bound on ϵ. By minimizing ϵ subject to the constraint (4.14), we find the bound on the slow-roll parameter ϵ to be

$$\epsilon \geq \frac{27}{13} \quad \text{whenever } V > 0. \tag{4.15}$$

Hence both inflation and de Sitter vacua are forbidden everywhere in field space.

Indeed in any vacuum $\partial V/\partial\rho = \partial V/\partial\tau = 0$, so eq. (4.13) implies $V = -(\sum pV_p)/9$. By assuming $V_p > 0$ for at least one of $p = 2$, 4, or 6, then Minkowski vacua are forbidden also.[6] This type of relation was used in [43] to show that vacua in a specific IIA compactification must be anti-de Sitter, and in [50] to rule out a simple F-term uplift of this model; here we have shown that this relation holds very generally in IIA compactifications, and furthermore rules out inflation anywhere on moduli space for potentials containing only terms of the form (4.12). For stabilized compactifications, the implication is that the field vector always undergoes fast rolling from a region with $V > 0$ towards an anti-de Sitter vacuum.

The no-go theorem we have derived here can be interpreted as defining a necessary condition for inflation in IIA models: *In order to inflate, a IIA compactification must contain some additional structure beyond that considered so far which gives a term in the potential V whose scaling leads to a term on the RHS of (4.13) with a coefficient less than 9 if positive or greater than 9 if negative.* In the next section we turn to a discussion of specific types of structure which can realize this necessary condition for inflation.

4.4 Evading the No-Go Theorem

In the analysis which led to the preceding no-go theorem for inflation in IIA compactifications, we allowed a specific set of ingredients in the IIA models considered. Following [43], we included NS-NS 3-form flux, R-R fluxes, D6-branes and O6-planes, which are sufficient to stabilize all geometric moduli. Because D6-branes and anti-D6-branes give terms to V which scale in the same way, this no-go theorem rules out brane-antibrane inflation as well as inflationary models using other moduli as

[6] We almost certainly need $V_p > 0$ for at least one of $p = 2$, 4, or 6 in order to be in the large volume and small string coupling limit, since the 3-form, 0-form, and number of D6/O6 planes are tightly constrained by a tadpole condition.

the inflaton. Corrections to 10-dimensional supergravity which arise at small compactification volume may evade our no-go result, so one approach to finding a IIA compactification with inflation is to include finite volume corrections to the Kähler potential. There are also other structures we can include in a compactification besides those already mentioned which can evade the no-go result. In this section we consider other possibilities which give rise to terms in the potential V which have scaling coefficients violating (4.13). This may give some guidance for where to look to find IIA compactifications with inflation and/or de Sitter vacua. Note that many of the structures suggested go beyond the range of compactifications which are so far well understood in string theory.

4.4.1 Various Ingredients

One obvious possibility is to include Dp-branes and Op-planes of dimensionality other than $p = 6$. The scaling of the resulting contributions to V are as follows

$$\begin{aligned} V_{\mathrm{Dp}} &\propto \rho^{\frac{p-6}{2}}\tau^{-3} \quad \text{for D}p\text{-branes,} \\ V_{\mathrm{Op}} &\propto -\rho^{\frac{p-6}{2}}\tau^{-3} \quad \text{for O}p\text{-planes.} \end{aligned} \tag{4.16}$$

In these cases, the right hand of eq. (4.13) is $(12 - p/2)V_{\mathrm{Dp/Op}}$, so the no-go theorem applies to Dp-branes with $p \leq 6$ and Op-planes with $p \geq 6$, but is evaded otherwise. Since the branes must extend in all non-compact directions of space-time[7] in IIA we can only consider $p = 4, 8$. Wrapping such a brane, however, would either require a compactification with non-trivial first homology class H_1 or finite π_1, or perhaps require the use of so-called coisotropic 8-branes [51, 52]. These branes carry charge and so would generate additional tadpoles which would need to be cancelled, as well as potentially breaking supersymmetry. Compactifications with branes of this type have not been studied extensively in the string theory literature, but it would be interesting to investigate this range of possibilities further.

Another possibility is to include more general NS-NS fluxes, such as geometric fluxes *a la* Scherk and Schwarz [53] or non-geometric fluxes [54]. Geometric fluxes parameterize a "twisting" away from Calabi-Yau topology, generalizing the notion of a twisted torus [53, 55]. These fluxes are associated with a metric with curvature on the compact space. Geometric fluxes arise under T-duality or mirror symmetry when an NS-NS 3-form H has a single index in a dualized direction [56]. Further T-dualities generate non-geometric fluxes, described in [57] through the sequence $T : H_{abc} \rightarrow f^a_{bc} \rightarrow Q^{ab}_c \rightarrow R^{abc}$, where f parameterize geometric fluxes, Q are locally geometric but globally non-geometric fluxes which can in some cases be realized in the language of "T-folds" [58, 59], and R parameterize fluxes associated with compactifications which are apparently not even locally geometric. These general NS-NS fluxes and the associated compactifications are still poorly understood. From the T-duality picture, however, it is straightforward to determine the scalings of these 3 types of fluxes.

[7] Otherwise they would describe localized excitations in an asymptotic vacuum not including these branes.

Each T-duality inverts the size of a dimension of the compactification, replacing a factor of ρ^{-1} in the scaling with ρ, so we have

$$\begin{aligned} V_{\rm f} &\propto \pm\rho^{-1}\tau^{-2} \quad \text{for geometric (f) flux,} \\ V_{\rm Q} &\propto \pm\rho\,\tau^{-2} \quad \text{for Q flux,} \\ V_{\rm R} &\propto \pm\rho^{3}\tau^{-2} \quad \text{for R flux.} \end{aligned} \tag{4.17}$$

Applying the linear operator of eq. (4.13) to each of these terms gives a right hand side of $7V_{\rm f}$, $5V_{\rm Q}$, and $3V_{\rm R}$, respectively. So although our no-go theorem applies when such terms are negative, it is evaded if any of these contributions are positive. Among these fluxes, the best understood are geometric fluxes, which are realized in many simple compactifications such as twisted tori [44, 53, 55]. Compactification on spaces with these fluxes (and other ingredients) is studied in the forthcoming paper by Silverstein [47], where it is shown that de Sitter vacua can indeed be realized in such backgrounds. This is a promising place to look for string inflation models. Note, however, that general NS-NS fluxes cannot in general be taken to the large volume limit. For example, fluxes of the Q type involve a T-duality inverting the radius of a circle in a fiber when a circle in the base is traversed. Thus, somewhere the size of the fiber must be sub-string scale. This makes solutions of the naive 4-dimensional supergravity theory associated with flux compactifications such as those found in [60] subject to corrections from winding modes and also to uncontrolled string theoretic corrections if curvatures become large.

Another possible ingredient which can be added to the IIA compactification models are NS 5-branes; these are the magnetic duals of the string. Such objects are non-perturbative and carry tension in 10 dimensions that scales as $g_s^{-2} = e^{-2\phi}$. While they backreact more significantly than D-branes and so are not as simple to describe at the supergravity level, their presence can be captured by adding a term of the form

$$-\mu_5 \int_{\rm NS5} d^6\xi \sqrt{-g}\, e^{-2\phi} \tag{4.18}$$

to the action of eq. (4.1). The compactification and conformal transformation yields a new term that scales as

$$V_{\rm NS5} \propto \rho^{-2}\tau^{-2}, \tag{4.19}$$

yielding a right hand side of eq. (4.13) of $8V_{\rm NS5}$, hence evading the no-go theorem. To satisfy tadpole cancellation, one would probably wish to add metastable pairs of separated NS 5-branes and anti-NS 5-branes, wrapping distinct isolated curves in the same homology class. Such configurations have been a focus of study in the dual IIB theory in recent works, starting with [61].

4.4.2 An Illustration

In this section we illustrate how the above ingredients may be useful from the point of view of building de Sitter vacua and inflation. We focus the discussion on NS 5-branes, which appear particularly promising. We will not attempt an explicit construction,

since that would take us beyond the scope of this work. Our goal is only to show that simple, available ingredients in the IIA theory have energy densities which scale with the volume and dilaton moduli in a way which suffices to overcome our no-go theorem, which was based purely on scalings of energy densities. This should act as a guide to model building, but should be taken in the heuristic spirit it is offered.

Using the potential of eq. (4.12) and adding to it a (necessarily) positive term from NS 5-branes wrapping 2-cycles, we obtain[8]

$$V = \frac{A_3(\phi_i)}{\rho^3\tau^2} + \sum_p \frac{A_p(\phi_i)}{\rho^{p-3}\tau^4} - \frac{A_{\rm O6}(\phi_i)}{\tau^3} + \frac{A_{\rm NS5}(\phi_i)}{\rho^2\tau^2}. \tag{4.20}$$

Let us now streamline V and focus on the most important features of this setup; we set $A_2 = A_6 = 0$, and expect the remaining coefficients to scale with fluxes and numbers of planes and branes as

$$A_3 \sim h_3^2,\ \ A_0 \sim f_0^2,\ \ A_4 \sim f_4^2,\ \ A_{\rm O6} \sim N_{\rm O6},\ \ A_{\rm NS5} \sim g(\omega) N_{\rm NS5}, \tag{4.21}$$

where we have introduced a function g of the modulus ω associated with the 2-cycle wrapped by each NS 5-brane.

There is a tadpole constraint that the charge on the O6-plane must be balanced by the fluxes from H_3 and F_0, i.e., $h_3 f_0 \sim -N_{\rm O6}$. We use this to eliminate h_3. Now since it has no effect on the kinetic energy, let us rescale our fields as: $\rho \to \rho\sqrt{|f_4/f_0|}$ and $\tau \to \tau\sqrt{|f_0 f_4^3|}/N_{\rm O6}$. Then we have

$$V = V_{\rm flux}\left[\frac{B_3(\phi_i)}{\rho^3\tau^2} + \sum_p \frac{B_p(\phi_i)}{\rho^{p-3}\tau^4} - \frac{B_{\rm O6}(\phi_i)}{\tau^3} + \frac{B_{\rm NS5}(\phi_i)}{\rho^2\tau^2}\right] \tag{4.22}$$

with $V_{\rm flux} \equiv N_{\rm O6}^4/\sqrt{|f_0^3 f_4^9|}$, and

$$B_3 \sim 1,\ \ B_0 \sim 1,\ \ B_4 \sim 1,\ \ B_{\rm O6} \sim 1,\ \ B_{\rm NS5} \sim c(\omega) \equiv g(\omega) N_{\rm NS5}\sqrt{|f_0^3 f_4|}/N_{\rm O6}^2 \tag{4.23}$$

So the shape of the potential is essentially controlled by the parameter c.

Let us make some comments on the value of c which determines the contribution of the NS 5-brane. In the large f_4 flux limit of [43], c is parametrically larger than all other contributions, including that from the 3-form flux to which we compare: The NS 5-brane contribution has the same scaling with τ, but scales more slowly to zero as $\rho \to \infty$. Therefore, one would expect that one needs *small* $N_{\rm NS5}$ to push V up without causing a runaway to infinite volume. As we will see, we need c to be fined tuned and $\mathcal{O}(1)$. An analogous situation occurs in IIB string theory with compactifications involving anti-D3 branes and non-perturbative volume stabilization. There the presence of strong warping allows one to construct states where the anti-D3 energy density is exponentially suppressed, naturally providing a small coefficient to the perturbation of the energy density [62]. This plays an important role in de Sitter

[8]We have set $A_{\rm D6} = 0$ as it adds little to the analysis.

constructions in that context [63], and one might expect an analogous mechanism (involving large warping or a very small cycle) could similarly dynamically explain a small $g(\omega)$ to compensate large f_4 in the IIA context. We could also imagine a compactification where N_{O6} is very large to achieve the same result.

In any case, by treating c as a continuous parameter and ignoring the dynamics of all other moduli, we can obtain a meta-stable de Sitter vacuum. We set ($\bar{m}_{\rm Pl} = 1$) $B_3 = 1/4$, $B_0 = 1/4$, $B_4 = 3/8$, $B_{O6} = 2$, and stabilize ρ by satisfying $\partial V/\partial\rho = 0$. In Figure 4-1 we plot $V = V(\hat{\tau})$ of eq. (4.22) for different choices of $B_{\rm NS5} = c$. We find that there are critical values of c: for a Minkowski vacuum $c_M \approx 2.205$ and for a point of inflection $c_I \approx 2.280$. For $c < c_M$ an anti-de Sitter vacuum exists, for $c_M < c < c_I$ a de Sitter vacuum exists, and for $c > c_I$ no vacuum exists. For c close to c_M^+ we expect the de Sitter vacuum lifetime to be long, as in the KKLT meta-stable vacuum of type IIB [63].

If one could realize such a construction explicitly in a controlled regime of the IIA theory, then we anticipate that the opportunities to realize inflation will be greatly enhanced. For example, the local maximum in Figure 4-1 may be useful since $\epsilon \to 0$ there. It will often suffer, however, from the so-called η problem; the second slow-roll parameter (which measures the second derivative of the potential) will typically be large and negative. It is possible that, for example, by moving in another transverse direction at the hill-top one could build some form of hybrid-inflation. We do note that by choosing $c = c_I$ we would immediately solve the η problem and have inflection-point-inflation as advocated in [64]. In this situation, however, inflation would finish with two difficulties; runaway moduli and little to no reheating. A different approach is to simply fix the volume and the dilaton at a de Sitter (or Minkowski) minimum with high mass and strictly use other lighter moduli to drive inflation in transverse directions. Scenarios such as N-flation may be possible here [22].

4.5 Type IIB Compactifications

In this section we discuss the relationship between the results we have derived here for IIA string theory and previous work on inflation in type IIB string theory. Although the type IIA and IIB string theories are related through T-duality, this duality acts in a complicated way on many of the ingredients used in constructing flux compactifications. The basic IIB flux compactifications of [65] (see also [66, 67, 68]) involve H-flux, F_3-flux, D3-branes and O3-planes. T-duality/mirror symmetry on such compactifications transforms the NS-NS H-flux into a complicated combination of H-flux, geometric flux and non-geometric flux which generically violates the restrictions needed in the no-go theorem we have proven here. Conversely, the backgrounds which we have proven here cannot include inflation are T-dual/mirror to complicated IIB backgrounds with geometric and non-geometric fluxes. Furthermore, the volume modulus used in our analysis is dual to a complex structure modulus in the IIB theory which is difficult to disentangle from the other moduli, so that proving the analogous no-go theorem in IIB, on the exotic class of backgrounds where it is relevant, would be quite difficult without recourse to duality.

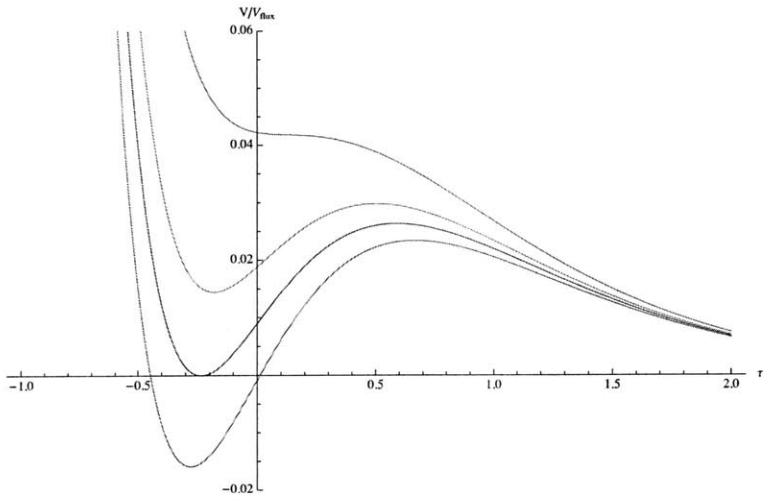

Figure 4-1: The potential $V(\hat{\tau})/V_{\text{flux}}$ with ($\bar{m}_{\text{Pl}} = 1$) $B_3 = 1/4$, $B_0 = 1/4$, $B_4 = 3/8$, $B_{\text{O6}} = 2$, and ρ satisfying $\partial V/\partial\rho = 0$. From bottom to top, the curves correspond to the following choices of c: 2.183 (anti-de Sitter), 2.205 (Minkowski), 2.227 (de Sitter), and 2.280 (inflection), respectively.

Despite these complications, we can easily explain why standard IIB flux vacua behave so differently with respect to potential constructions of de Sitter space and inflation as seen through the methods of this chapter. For IIB flux vacua, the basic ingredients of H-flux, F_3-flux, D3-branes and O3-planes give contributions to the 4-dimensional potential which scale as

$$\frac{e^{2\phi}}{\rho^6}, \quad \frac{e^{4\phi}}{\rho^6}, \quad \frac{e^{3\phi}}{\rho^6}, \quad -\frac{e^{3\phi}}{\rho^6} \tag{4.24}$$

respectively.[9] Thus, the scaling equation (4.13) is not the appropriate equation for gaining useful information about the cosmological structure. Instead, we simply have

$$-\rho\frac{\partial V(\phi,\rho)}{\partial\rho} = 6V(\phi,\rho)\,. \tag{4.25}$$

This shows immediately that any classical vacuum must have $V = 0$, as is well known from the tree-level no-scale structure of this class of models [65]. When the dilaton and complex structure moduli are chosen to fix $V = 0$ then there is a classical flat direction. When such moduli cannot be chosen then the (positive) potential causes a runaway to large volume. In such a simple setting, one can show that $\epsilon \geq 3$ whenever $V > 0$, but this *should not* be viewed as a serious obstacle to realizing inflation or de Sitter vacua. This is because the classical flat directions along which $V = \partial V/\partial\rho = 0$ can be lifted by including quantum contributions (or other fluxes etc) to the potential, and then the naive bound on ϵ is irrelevant. Typically, non-perturbative corrections to the superpotential are included to stabilize the Kähler moduli. This should be contrasted with the classical stabilization in IIA.[10]

In this general IIB setting, starting with the no-scale vacuum and including various corrections to achieve de Sitter, many inflationary models have been proposed. The basic strategy, starting with [29], has usually been to stabilize the dilaton and volume moduli at a high scale, and inflate at a lower energy scale. Then, the dilaton and volume contributions to ϵ (which were the focus of our no-go theorem in the IIA context) are simply absent. The most explicit models to date appear in [31], though one should consult the reviews [36, 37, 38, 39, 40] for a much more extensive list of approaches and references.

4.6 Discussion

In this chapter we have demonstrated that a large class of flux compactifications of type IIA string theory cannot give rise to inflation in the regime of moduli space where we have parametric control of the potential. This result applies to large-volume, weak coupling compactifications on arbitrary Calabi-Yau spaces with NS-NS 3-form

[9] This ρ scales as $\mathrm{Vol}^{1/3}$ defined in the string frame, as in Section II, and should not be confused with the ρ modulus of e.g. [65], which scales as $\mathrm{Vol}^{2/3}$ defined in the Einstein frame.

[10] Note that in the absence of fluxes, branes and orientifolds, mirror symmetry relates supersymmetric IIA and IIB compactifications. In this case $V = 0$ exactly.

flux, general R-R fluxes, D6-branes and O6-planes. These ingredients are arguably the most well understood in IIA compactifications. The no-go theorem of Section 4.3 applies in particular to the $T^6/\mathbb{Z}_3^2$ orientifold model of Ref. [43] and the $T^6/\mathbb{Z}_4$ orientifold model of Ref. [69], and explains the numerical results of Ref. [46] which suggested that inflation is impossible in these models. The no-go theorem we have derived here, however, applies to all other Calabi-Yau compactifications of this general type as well. So the simplest part of the IIA flux compactification landscape does not inflate. This implies the following constraint: the portions of the landscape that are possibly relevant to phenomenology will necessarily involve interplay of more diverse ingredients, as has also been found in the IIB theory.

We emphasize that while our derivation has only involved two moduli (the volume ρ and the dilaton τ) we are not assuming that either of those moduli necessarily play the role of the inflaton. Instead, inflation by *any* modulus (or brane/anti-brane) is always spoiled due to the fast-roll of ρ and/or τ. This follows because a necessary condition for slow-roll inflation is that the potential be flat in *every* direction in field space, as quantified by ϵ. In fact because the first slow-roll parameter is so large $\epsilon \geq \frac{27}{13}$, there can never be many e-foldings, even ruling out so-called *fast-roll inflation* [70]. We point out that this result is slightly non-trivial because it requires analyzing *both* ρ and τ, as in eq. (4.13), and cannot be proven by focussing on only one of them.

A simple corollary of our result is that no parametrically controlled de Sitter vacua exist in such models. We emphasize, though, that proving the non-existence of inflation is a much stronger statement than proving the non-existence of de Sitter vacua. In particular we can imagine a priori a scenario where, although the vacuum is anti-de Sitter (or Minkowski) or there is no vacuum at all, inflation is still realized somewhere in a region where $V > 0$. We find, however, that this does not occur. It is intriguing that the proof of the non-existence of inflation is so closely related to that of the non-existence of de Sitter. This *suggests* that there may be a close connection between building de Sitter vacua and realizing inflation.

Our result can be interpreted as giving a necessary condition for inflation in IIA models. To have inflation, some additional structure must be added which gives rise to a potential term in 4 dimensions with scaling such that $-\rho\partial V/\partial\rho - 3\tau\partial V/\partial\tau = \alpha V$ with a coefficient $\alpha < 9$ for a positive contribution and $\alpha > 9$ for a negative contribution. We described various ingredients which give rise to such terms; compactifications including these ingredients may be promising places to look for inflationary models. Some of these ingredients take us outside the range of string compactifications which are understood from a perturbative/supergravity point of view. Among the possibilities which evade the assumptions made in deriving the no-go theorem are other NS-NS fluxes, such as geometric and non-geometric fluxes. It is currently difficult to construct models with such generic fluxes in a regime that is under control, but progress in this direction has been made in [47], where IIA de Sitter vacua are found using a specific set of geometrical fluxes and other ingredients. Another promising direction which we have indicated here (also incorporated in [47]) is to include NS 5-branes and anti-NS 5-branes on 2-cycles in the Calabi-Yau. More work is needed to find explicit models where these branes are stabilized in a regime allowing de Sitter vacua and inflation, but this does not seem to be impossible or ruled out by any

obvious considerations. In addition to the mechanisms we have discussed, there are probably other structures (e.g., D-terms [71, 72]) which violate the conditions of the no-go theorem.

Including any of these ingredients does not guarantee that inflation will be realized. It may be the case that a slightly more general no-go theorem for inflation exists with certain combinations of additional ingredients. This may follow from studying other moduli, since any field can ruin inflation by fast-rolling. It may also be that with more work an elegant realization of inflation and de Sitter vacua can be found in type IIA string theory. This deserves further investigation.

4.7 Appendix: 4-Dimensional $\mathcal{N} = 1$ Supergravity

It was shown in [42] that with the ingredients used in Section 4.3, the dimensional reduction of massive IIA supergravity can be described in the language of a 4-dimensional $\mathcal{N} = 1$ supergravity theory in terms of a Kähler potential and superpotential.

4.7.1 Kinetic Energy from Kähler Potential

The authors of Ref. [42] showed that in the large volume limit the Kähler potential is given by $K = K^K + K^Q$ where

$$K^K = -\bar{m}_{\rm Pl}^2 \ln\left(\frac{4}{3}\kappa_{abc}v^a v^b v^c\right), \tag{4.26}$$

$$K^Q = -2\bar{m}_{\rm Pl}^2 \ln\Big(2\,{\rm Im}(CZ_\lambda){\rm Re}(Cg_\lambda) - 2\,{\rm Re}(CZ_k){\rm Im}(Cg_k)\Big) \tag{4.27}$$

Here v^a are the Kähler moduli, κ_{abc} are the triple-intersection form constants, the set of Z and g are the co-ordinates in some basis of a holomorphic 3-form that describes the complex structure moduli, and C is the "compensator" which incorporates the dilaton. If the set of complex moduli is denoted ψ^i, then the kinetic energy is given by

$$T = -K_{i\bar{j}}\partial_\mu\psi^i\partial^\mu\psi^{\bar{j}} \tag{4.28}$$

with corresponding first slow-roll parameter

$$\epsilon = \bar{m}_{\rm Pl}^2\frac{K^{i\bar{j}}V_iV_{\bar{j}}}{V^2}. \tag{4.29}$$

Let us focus on the Kähler contribution. We write the Kähler moduli as $\psi^a = a^a + i\,v^a$, so the kinetic energy is given by

$$T^K = -\frac{1}{4}\frac{\partial^2 K^K}{\partial v^a\partial v^b}\left(\partial_\mu v^a\partial^\mu v^b + \partial_\mu a^a\partial^\mu a^b\right). \tag{4.30}$$

Now we change coordinates from v^a to $\{\rho,\gamma^a\}$ as follows:

$$v^a = \rho\,\gamma^a, \quad \text{with } \kappa_{abc}\gamma^a\gamma^b\gamma^c = 6, \tag{4.31}$$

so ${\rm Vol} = \rho^3$. Then using $\partial_\mu(\kappa_{abc}\gamma^a\gamma^b\gamma^c) = 0$, we obtain:

$$\begin{aligned} T^K = -\,\bar{m}_{\rm Pl}^2\Bigg[&\frac{3(\partial_\mu\rho)^2}{4\,\rho^2} - \frac{1}{4}\kappa_{abc}\gamma^c\partial_\mu\gamma^a\partial^\mu\gamma^b \\ &+\frac{\kappa_{acd}\gamma^c\gamma^d\kappa_{bef}\gamma^e\gamma^f - 4\kappa_{abc}\gamma^c}{16\,\rho^2}\partial_\mu a^a\partial^\mu a^b\Bigg] \end{aligned} \tag{4.32}$$

By switching from ρ to $\hat{\rho}$, we see that the first term is precisely the kinetic energy

for $\hat{\rho}$. The remaining kinetic energy terms for γ^a and a^a are block diagonal (there are no cross terms involving $\partial_\mu\rho\,\partial^\mu\gamma^a$ etc), and this has an important consequence: We know that in the physical region the total kinetic energy must be positive, so *each* of the above 3 terms must be positive. Hence, $T^K = -(\partial_\mu\hat{\rho})^2/2 + \text{positive}$.

For the complex structure/dilaton sector the procedure is similar, although more subtle. In (4.27) the expression for K^Q is not a completely explicit function of the moduli; although $\mathrm{Re}(CZ_k)$ and $\mathrm{Re}(Cg_k)$ are explicitly half of the complex structure moduli, $\mathrm{Im}(CZ_k)$ and $\mathrm{Im}(Cg_k)$ are only *functions* of the remaining complex structure moduli. Nevertheless, the kinetic term is again block diagonal. To see this, note that the compensator C, and hence all moduli in this sector, are proportional to τ. Furthermore, the Z and g are constrained to the surface: $K^Q = -2\bar{m}_{\rm Pl}^2\ln(\tau^2)$. This is analogous to the Kähler sector. Without going through the details here, we find $T^Q = -(\partial\hat{\tau})^2/2 + \text{positive}$. In fact we know this *must* be true from the 10-dimensional point of view; the dilaton modulus is inherited directly from 10 dimensions, and so cannot possibly give rise to mixed kinetic terms with the complex structure moduli in 4 dimensions.

4.7.2 Potential Energy from Superpotential

From [42] the superpotential in the IIA theory is given by $W = W^K + W^Q$ where

$$W^K = f_6 + f_{4a}t^a + \frac{1}{2}f_{2a}\kappa_{abc}t^bt^c - \frac{f_0}{6}\kappa_{abc}t^at^bt^c, \tag{4.33}$$

$$W^Q = \left(\bar{h}_\lambda\xi^\lambda - h_k\xi^k\right) + 2i\left(\bar{h}_\lambda\mathrm{Re}(Cg_\lambda) - h_k\mathrm{Re}(CZ_k)\right). \tag{4.34}$$

We will not explain all the details of this here; the interested reader is pointed to Refs. [42, 43]. For our present purposes it suffices to note that $\mathrm{Im}(t^a) = v^a \propto \rho$ and $\mathrm{Im}(W^Q) \propto \tau$. Hence, the superpotential is cubic in ρ and linear in τ.

From the supergravity formula for the Einstein frame potential

$$V = e^{K/\bar{m}_{\rm Pl}^2}\left(D_iWK^{i\bar{j}}\overline{D_jW} - 3\frac{|W|^2}{\bar{m}_{\rm Pl}^2}\right), \tag{4.35}$$

we easily infer the dependence on ρ and τ. Firstly, since the constraints on γ^a and complex structure imply that $K = -\bar{m}_{\rm Pl}^2\ln(8\rho^3\tau^4)$, the pre-factor scales as

$$e^{K/\bar{m}_{\rm Pl}^2} \propto \rho^{-3}\tau^{-4}. \tag{4.36}$$

Also, the scaling contributions from the parenthesis in eq. (4.35), which is roughly $|W|^2$, can be easily determined. By analyticity, only terms of the form $\rho^p\tau^q$ can appear where $p+q$ is even. This leaves only the following 7 possible scalings: $\tau^2, \rho^6, \rho^4, \rho^2, 1, \rho^3\tau, \rho\tau$. By multiplying by the pre-factor, we see that the first 6 terms are precisely those that arise from H_3, F_0, F_2, F_4, F_6, and D6/O6, respectively. The 7th term ($\rho\tau$) is new, but cancels between the two terms inside the parenthesis of eq. (4.35). Hence we obtain the form of the potential given in eq. (4.12).

Bibliography

[1] A. Guth, "The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems," Phys. Rev. D, 23, 347 (1981).

[2] A. D. Linde, "A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems," Phys. Lett. B, 108, 389 (1982).

[3] A. Albrecht and P. J. Steinhardt, "Reheating an Inflationary Universe," Phys. Rev. Lett., 48, 1220 (1982)

[4] A. D. Linde, "Chaotic Inflation," Phys. Lett. B, 129, 177 (1983)

[5] M. Tegmark, "What does inflation really predict?," JCAP 0504 (2005) 001 [arXiv:astro-ph/0410281].

[6] D. N. Spergel et al., "Wilkinson Microwave Anisotropy Probe (WMAP) Three Year Results: Implications for Cosmology", Astroph. J. Supp., 170, 377 (2007) [arXiv:astro-ph/0603449].

[7] M. Tegmark et al., " Cosmological Constraints from the SDSS Luminous Red Galaxies," Phys. Rev. D74, 123507 (2006) [arXiv:astro-ph/0608632].

[8] S. Hellerman, N. Kaloper, and L. Susskind, "String Theory and Quintessence," (2001) [arXiv:hep-th/0104180].

[9] W. Fischler, A. Kashani-Poor, R. McNees, and S. Paban, "The Acceleration of the Universe, a Challenge for String Theory," (2001) [arXiv:hep-th/0104181].

[10] R. Brustein and S.P. de Alwis, "Inflationary cosmology in the central region of string / M theory moduli space," Phys. Rev. **D68** (2003) 023517 [arXiv:hep-th/0205042].

[11] E. Silverstein, "TASI/PiTP/ISS lectures on moduli and microphysics," [arXiv:hep-th/0405068].

[12] M. Grana, "Flux compactifications in string theory: A comprehensive review," Phys. Rept. **423** (2006) 91 [arXiv:hep-th/0509003].

[13] M. Douglas and S. Kachru, "Flux compactification," Rev. Mod. Phys. **79** (2007) 733 [arXiv:hep-th/0610102].

[14] R. Bousso, "Precision cosmology and the landscape," [arXiv:hep-th/0610211].

[15] R. Blumenhagen, B. Kors, D. Lust and S. Stieberger, "Four-dimensional string compactifications with D-branes, orientifolds and fluxes," Phys. Rept. **445** (2007) 1 [arXiv:hep-th/0610327].

[16] P. Binetruy and M.K. Gaillard, "Candidates for the Inflaton Field in Superstring Models," Phys. Rev. **D34** (1986) 3069.

[17] T. Banks, M. Berkooz, S. Shenker, G. Moore and P. Steinhardt, "Modular cosmology," Phys. Rev. **D52** (1995) 3548 [arXiv:hep-th/9503114].

[18] J. Conlon and F. Quevedo, "Kahler moduli inflation," JHEP **0601** (2006) 146 [arXiv:hep-th/0509012].

[19] J.R. Bond, L. Kofman, S. Prokushkin and P. Vaudrevange, "Roulette inflation with Kahler moduli and their axions," Phys. Rev. **D75** (2007) 123511 [arXiv:hep-th/0612197].

[20] J. Blanco-Pillado et al, "Racetrack inflation," JHEP **0411** (2004) 063 [arXiv:hep-th/0406230].

[21] J. Kim, H. Nilles and M. Peloso, "Completing natural inflation," JCAP **0501** (2005) 0501 [arXiv:hep-ph/0409138].

[22] S. Dimopoulos, S. Kachru, J. McGreevy and J. Wacker, "N-flation," [arXiv:hep-th/0507205].

[23] R. Easther and L. McAllister, "Random matrices and the spectrum of N-flation," JCAP **0605** (2006) 018 [arXiv:hep-th/0512102].

[24] R. Kallosh, N. Sivanandam and M. Soroush, "Axion inflation and gravity waves in string theory," [arXiv:0710.3249].

[25] T.W. Grimm, "Axion inflation in type II string theory," [arXiv:0710.3883].

[26] G. Dvali and S.H. Tye, "Brane inflation," Phys. Lett. **B450** (1999) 72 [arXiv:hep-ph/9812483].

[27] G. Shiu and S.H. Tye, "Some aspects of brane inflation," Phys. Lett. **B516** (2001) 421 [arXiv:hep-th/0106274].

[28] C. Burgess et al, "Brane/anti-brane inflation in orbifold and orientifold models," JHEP **0203** (2002) 052 [arXiv:hep-th/0111025].

[29] S. Kachru et al, "Towards inflation in string theory," JCAP **0310** (2003) 013 [arXiv:hep-th/0308055].

[30] E. Silverstein and D. Tong, "Scalar speed limits and cosmology: Acceleration from D-cceleration," Phys. Rev. **D70** (2004) 103505 [arXiv:hep-th/0310221].

[31] D. Baumann, A. Dymarsky, I. Klebanov and L. McAllister, "Towards an explicit model of D-brane inflation," [arXiv:0706.0360].

[32] K. Becker, M. Becker, and A. Krause, "M-Theory Inflation from Multi M5-Brane Dynamics," Nucl.Phys. B715 (2005) 349-371, [arXiv:hep-th/0501130v4].

[33] A. Krause and E. Pajer, "Chasing Brane Inflation in String-Theory," [arXiv:0705.4682v2 [hep-th]].

[34] J. Polchinski, "Introduction to cosmic F- and D-strings," [arXiv:hep-th/0412244].

[35] M. Alishahiha, E. Silverstein and D. Tong, "DBI in the sky," Phys. Rev. **D70** (2004) 123505 [arXiv:hep-th/0404084].

[36] S. H. Tye, "Brane inflation: String theory viewed from the cosmos," [arXiv:hep-th/0610221].

[37] J. Cline, "String Cosmology," [arXiv:hep-th/0612129].

[38] R. Kallosh, "On inflation in string theory," [arXiv:hep-th/0702059].

[39] C.P. Burgess, "Lectures on cosmic inflation and its potential stringy realizations," [arXiv:0708.2865].

[40] L. McAllister and E. Silverstein, "String cosmology: A review," [arXiv:0710.2951].

[41] L. J. Romans, "Massive $\mathcal{N} = 2A$ Supergravity in Ten-Dimensions," Phys. Lett. B169 (1986) 374.

[42] T. Grimm and J. Louis, "The Effective Action of Type IIA Calabi-Yau Orientifolds," Nucl.Phys. B718 (2005) 153-202 [arXiv:hep-th/0412277].

[43] O. DeWolfe, A. Giryavets, S. Kachru, W. Taylor, "Type IIA Moduli Stabilization," JHEP 0507, (2005) [arXiv:hep-th/0505160].

[44] G. Villadoro and F. Zwirner, "N = 1 Effective Potential from Dual Type-IIA D6/O6 Orientifolds with General Fluxes," JHEP 0506 (2005) 047 [arXiv:hep-th/0503169].

[45] B. S. Acharya, F. Benini, and R. Valandro, "Fixing Moduli in Exact Type IIA Flux Vacua," JHEP 0702 018 (2007), [arXiv:hep-th/0607223v2].

[46] M. P. Hertzberg, M. Tegmark, S. Kachru, J. Shelton, O. Özcan, "Searching for Inflation in Simple String Theory Models," Phys. Rev. D 76, 103521 (2007) [arXiv:0709.0002 [astro-ph]].

[47] E. Silverstein, "Simple de Sitter Solutions," to appear.

[48] G. Moore, private communication.

[49] T. Banks and K. van den Broek, "Massive IIA flux compactifications and U-dualities," JHEP **0703** 068 (2007) [arXiv:hep-th/0611185].

[50] R. Kallosh and M. Soroush, "Issues in IIA uplifting," JHEP **0706** 041 (2007) [arXiv:hep-th/0612057].

[51] A. Font, L.E. Ibanez and F. Marchesano, "Coisotropic D8-branes and model-building," JHEP **0609** 080 (2006) [arXiv:hep-th/0607219].

[52] A. Kapustin and T. Orlov, "Remarks on A branes, mirror symmetry, and the Fukaya category," J. Geom. Phys. **48** (2003) 84 [arXiv:hep-th/0109098].

[53] J. Scherk and J. H. Schwarz, "How To Get Masses From Extra Dimensions," Nucl. Phys. B **153** 61 (1979).

[54] B. Wecht, "Lectures on Nongeometric Flux Compactifications," [arXiv:0708.3984 [hep-th]].

[55] N. Kaloper and R. C. Myers, "The O(dd) story of massive supergravity," JHEP **9905** 010 (1999) [arXiv:hep-th/9901045].

[56] S. Kachru, M. Schulz, P. Tripathy and S. Trivedi, "New supersymmetric string compactifications," JHEP **0303** (2003) 061 [arXiv:hep-th/0211182].

[57] J. Shelton, W. Taylor, and B. Wecht, "Nongeometric Flux Compactifications," JHEP 0510 085 (2005), [arXiv:hep-th/0508133v4].

[58] A. Dabholkar and C. Hull, "Duality twists, orbifolds, and fluxes," JHEP **0309** 054 (2003) [arXiv:hep-th/0210209].

[59] C. M. Hull, "A geometry for non-geometric string backgrounds," JHEP **0510** 065 (2005) [arXiv:hep-th/0406102].

[60] J. Shelton, W. Taylor and B. Wecht, "Generalized flux vacua," JHEP **0702** 095 (2007) [arXiv:hep-th/0607015].

[61] M. Aganagic, C. Beem, J. Seo and C. Vafa, "Geometrically induced metastability and holography," [arXiv:hep-th/0610249].

[62] S. Kachru, J. Pearson and H. Verlinde, "Brane/flux annihilation and the string dual of a nonsupersymmetric field theory," JHEP **0206** (2002) 021 [arXiv:hep-th/0112197].

[63] S. Kachru, R. Kallosh, A. Linde, and S. P. Trivedi, "de Sitter Vacua in String Theory," Phys. Rev. D68 (2003) 046005, [arXiv:hep-th/0301240].

[64] N. Itzhaki and E. D. Kovetz, "Inflection Point Inflation and Time Dependent Potentials in String Theory," JHEP 10 (2007) 054, [arXiv:0708.2798].

[65] S. Giddings, S. Kachru and J. Polchinski, "Hierarchies from Fluxes in String Compactifications," Phys. Rev. **D66** (2002) 106006 [arXiv:hep-th/0105097].

[66] K. Becker and M. Becker, "M theory on eight manifolds," Nucl. Phys. **B477** (1996) 155 [arXiv:hep-th/9605053].

[67] S. Gukov, C. Vafa and E. Witten, "CFTs from Calabi-Yau fourfolds," Nucl. Phys. **B584** (2000) 69 [arXiv:hep-th/9906070].

[68] K. Dasgupta, G. Rajesh and S. Sethi, "M theory, orientifolds, and G-flux," JHEP **9908** (1999) 023 [arXiv:hep-th/9908088].

[69] M. Ihl and T. Wrase, "Towards a Realistic Type IIA $T^6/\mathbb{Z}_4$ Orientifold Model with Background Fluxes, Part 1: Moduli Stabilization," JHEP 0607 027 (2006), [arXiv:hep-th/0604087].

[70] A. Linde, "Fast-Roll Inflation," JHEP11 (2001) 052.

[71] M. Ihl, D. Robbins, and T. Wrase, "Toroidal Orientifolds in IIA with General NS-NS Fluxes," JHEP 08 043 (2007), [arXiv:0705.3410 [hep-th]].

[72] D. Robbins and T. Wrase, "D-Terms from Generalized NS-NS Fluxes in Type II," [arXiv:0709.2186 [hep-th]].

Chapter 5

On Inflation with Non-minimal Coupling

A simple realization of inflation consists of adding the following operators to the Einstein-Hilbert action: $(\partial\phi)^2$, $\lambda\phi^4$, and $\xi\phi^2\mathcal{R}$, with ξ a large non-minimal coupling. Recently there has been much discussion as to whether such theories make sense quantum mechanically and if the inflaton ϕ can also be the Standard Model Higgs. In this chapter we answer these questions. Firstly, for a single scalar ϕ, we show that the quantum field theory is well behaved in the pure gravity and kinetic sectors, since the quantum generated corrections are small. However, the theory likely breaks down due to scattering provided by the self-interacting potential. Secondly, we show that the theory changes for multiple scalars $\vec{\phi}$ with non-minimal coupling $\xi\vec{\phi}\cdot\vec{\phi}\mathcal{R}$, since this introduces qualitatively new interactions which manifestly generate large quantum corrections even in the gravity and kinetic sectors, spoiling the theory for energies $\gtrsim \bar{m}_{\rm Pl}/\xi$. Since the Higgs doublet of the Standard Model includes the Higgs boson and 3 Goldstone bosons, it falls into the latter category and therefore its validity is manifestly spoiled. We show that these conclusions hold in both the Jordan and Einstein frames and describe an intuitive analogy in the form of the pion Lagrangian. We also examine the recent claim that curvature-squared inflation models fail quantum mechanically. Our work appears to go beyond the recent discussions.

5.1 Introduction

Cosmological inflation is our leading theory of the very early universe [1, 2, 3, 4], although its underlying microphysics is still unknown. Slow-roll inflation occurs in many models constructed over the years, with a scattering of model-dependent predictions. Models of inflation are quite UV sensitive since it may have occurred at extremely high energy scales, far higher than that which we can probe at colliders, and since some models involve super-Planckian excursions in field space. This suggests that top-down approaches may be required to make progress, although progress in that direction has not been easy (e.g., see [5]). On the other hand, it is interesting to explore simple models to see what we might learn and if we can connect inflation

to low energy physics.

One approach is to focus on dimension 4 Lagrangians, which allows the inclusion of the operators $(\partial\phi)^2$, $\lambda\phi^4$, and $\xi\phi^2\mathcal{R}$ in addition to the Einstein Hilbert term $\bar{m}_{\rm Pl}^2\mathcal{R}$ (by "dimension 4" we mean that each operator has a dimensionless coefficient, although gravity modifies the power counting since every term is an infinite tower of operators if we expand around flat space). The $\xi\phi^2\mathcal{R}$ term is in fact required to exist for an interacting scalar field in curved space. By taking the non-minimal coupling ξ to be large, a phase of inflation takes place, as originally discussed in [6] with constraints discussed in [7, 8, 9, 10].

A large dimensionless coupling is unusual from the perspective of particle physics, leading to much recent debate as to whether such models make sense quantum mechanically [6, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28]. In this work we show that for a single scalar field ϕ the model is well behaved quantum mechanically in the pure gravity and kinetic sectors, since these do not generate large quantum corrections. However, the self-interacting potential does likely cause the theory to fail. For a vector of fields $\vec{\phi}$ carrying an $O(N)$ symmetry the theory manifestly generates large quantum corrections in the gravity and kinetic sectors and breaks down at high scales relevant to inflation. As far as we aware, this difference between the single field and multi-field case has not been fully appreciated in the literature. So multi-field models of non-minimal inflation are manifestly ruled out (including the Standard Model Higgs), while single field models of non-minimal inflation are also problematic, in the sense that the Einstein frame potential is non-polynomial and therefore these models likely fail due to high energy scattering processes. This presents a challenge to them having a UV completion. We also examine curvature-squared models and show that their scalar-tensor formulation is perturbatively well behaved in the gravity, kinetic, and potential sectors; this contradicts the claims of Ref. [17].

Our chapter is organized as follows: in Section 5.2 we briefly review the class of non-minimal models, in Section 5.3 we discuss the quantum corrections, in Section 5.4 we discuss the pion analogy, in Section 5.5 we discuss curvature-squared models, and finally we conclude in Section 5.6.

5.2 Non-minimally Coupled Models

Consider a vector of scalar fields $\vec{\phi}$ with an $O(N)$ symmetry, non-minimally coupled to gravity. It is permissible to introduce the following operators into the action: $\partial\vec{\phi}\cdot\partial\vec{\phi}$, $\lambda(\vec{\phi}\cdot\vec{\phi})^2$, and $\xi\vec{\phi}\cdot\vec{\phi}\mathcal{R}$, where the last term represents a non-minimal coupling of the field to gravity. Adding these to the usual Einstein-Hilbert term, we have the action

$$S=\int d^4x\sqrt{-g}\left[\frac{1}{2}\bar{m}_{\rm Pl}^2 f(\vec{\phi})\mathcal{R}-\frac{1}{2}\partial\vec{\phi}\cdot\partial\vec{\phi}-V(\vec{\phi})\right], \tag{5.1}$$

where $f(\vec{\phi})=1+\xi\vec{\phi}\cdot\vec{\phi}/\bar{m}_{\rm Pl}^2$ and $V(\vec{\phi})=\frac{\lambda}{4}(\vec{\phi}\cdot\vec{\phi})^2$. Since we have a new parameter ξ, which is essentially unconstrained by observation, this allows the self-coupling

parameter to be non-negligible, such as $\lambda = \mathcal{O}(10^{-1})$, and yet we can achieve the correct amplitude of density fluctuations by making ξ very large: $\xi = \mathcal{O}(10^4)$. This was the original motivation of Ref. [6] for introducing such models, since it appeared that embedding the inflaton into the matter sector of the theory would be easier. By contrast compare to minimal $\lambda\phi^4$ inflation, with $\lambda = \mathcal{O}(10^{-12})$, which is evidently radiatively unstable if ϕ carries appreciable couplings to other fields.

One may either study the theory in this original Jordan frame, or switch to the Einstein frame by defining a new metric $g^E_{\mu\nu} = f(\vec{\phi})g_{\mu\nu}$. The corresponding Einstein frame potential is

$$V_E(\vec{\phi}) = \frac{V(\vec{\phi})}{f(\vec{\phi})^2} = \frac{\frac{\lambda}{4}(\vec{\phi}\cdot\vec{\phi})^2}{\left(1+\frac{\xi\vec{\phi}\cdot\vec{\phi}}{\bar{m}_{\rm Pl}^2}\right)^2}. \tag{5.2}$$

In the Einstein frame, although the gravity sector is simple, the kinetic sector for $\vec{\phi}$ is not. Ignoring total derivatives, we find the Einstein frame action to be

$$S = \int d^4x\sqrt{-g_E}\left[\frac{1}{2}\bar{m}_{\rm Pl}^2\mathcal{R}_E - \frac{1}{2}\frac{1}{f(\vec{\phi})}\partial\vec{\phi}\cdot\partial\vec{\phi} - \frac{3\,\xi^2}{\bar{m}_{\rm Pl}^2 f(\vec{\phi})^2}(\vec{\phi}\cdot\partial\vec{\phi})^2 - V_E(\vec{\phi})\right]. \tag{5.3}$$

The first term of the kinetic sector comes from the simple conformal rescaling and the second term comes from transforming the Ricci scalar.

In the classical dynamics we can assume that only a single component of the $\vec{\phi}$ vector is rolling during inflation since all transverse modes will be frozen due to Hubble friction. Assuming only one component is active (call it ϕ), the kinetic energy sector can be made canonical. Let us make this explicit. We introduce a new field σ, defined through the integral

$$\sigma \equiv \mathrm{sign}(\phi)\int_0^{\phi} d\bar{\phi}\sqrt{\frac{1}{f(\bar{\phi})} + \frac{6\,\xi^2\bar{\phi}^2}{\bar{m}_{\rm Pl}^2 f(\bar{\phi})^2}}\,. \tag{5.4}$$

In terms of σ both the gravity and kinetic sectors are minimal, but the potential $V_E(\sigma)$ is somewhat complicated, but monotonic and well behaved.

It is simple to see from eq. (5.2) that when $\phi \gg \bar{m}_{\rm Pl}/\sqrt{\xi}$ (corresponding to $\sigma \gg \bar{m}_{\rm Pl}$) the potential V_E approaches a constant; this is the regime in which the effective Planck mass runs in the original Jordan frame. In this regime the flatness of the potential ensures that a phase of slow-roll inflation takes place. Inflation ends when $\phi \sim \phi_e \equiv \bar{m}_{\rm Pl}/\sqrt{\xi}$. For details on the inflationary predictions we point the reader to Refs. [6, 11].

5.3 Quantum Corrections

Various attempts have been made at addressing whether quantum corrections spoil the validity of these non-minimal models of inflation due to the $\xi\phi^2\mathcal{R}$ term; see [6, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28]. In each of these

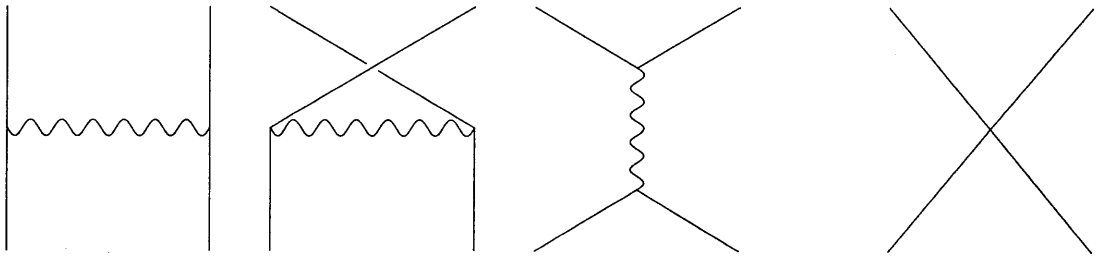

Figure 5-1: Tree-level scattering $2\phi \to 2\phi$. Diagrams 1–3: In Jordan frame it is due to graviton exchange through t, u, and s-channels. Diagram 4: In Einstein frame it is due to a single 4-point vertex.

works, the conclusion (either positive or negative) came independent of the number of scalars ϕ^i. Here we will show that the answer is in fact two-fold: for single field models the the gravity and kinetic sectors do not generate large quantum corrections, instead it is the potential $V(\phi)$ that causes the theory to enter a regime in which standard perturbation theory likely breaks down. For multiple field models, we show that the theory manifestly generates large quantum corrections in the gravity and kinetic sectors and breaks down in the regime relevant for inflation.

5.3.1 Single Field

Most descriptions of non-minimal models have focussed on just a single field ϕ, even for cases where ϕ represents the Higgs. For example, in [18] it is explicitly assumed that all quantum issues require only a single field analysis (and it is concluded in [18] that the quantum field theory is highly unnatural and breaks down). In this case it is subtle as to whether the effective field theory makes sense. Let us go through the analysis first in the Jordan frame and then the Einstein frame. We begin by examining the gravity and kinetic sectors (as was the focus of Ref. [17]) before addressing the potential V in the next section.

In the Jordan frame we must examine the consequences of the $\xi\phi^2\mathcal{R}$ operator. Let us expand around flat space. We decompose the metric as

$$g_{\mu\nu} = \eta_{\mu\nu} + \frac{h_{\mu\nu}}{\bar{m}_{\rm Pl}}, \tag{5.5}$$

here $h_{\mu\nu}$ are metric perturbations with mass dimension 1. The Ricci scalar has an expansion around flat space that goes as $\mathcal{R} \sim \Box h/\bar{m}_{\rm Pl} + \ldots$. To leading order in the Planck mass, this gives the following dimension 5 operator in the Jordan frame action

$$\frac{\xi}{\bar{m}_{\rm Pl}}\phi^2\Box h. \tag{5.6}$$

It is tempting to declare that this implies the theory has a cutoff at $\Lambda = \bar{m}_{\rm Pl}/\xi$, but is this correct? To answer this, let us consider the scattering process: $2\phi \to 2\phi$.

At tree-level this proceeds via a single exchange of a graviton, giving the following leading order contribution to the matrix element (we assume massless ϕ particles)

$$\mathcal{M}_{\rm c}(2\phi \to 2\phi) \sim \frac{\xi^2 E^2}{\bar{m}_{\rm Pl}^2}. \tag{5.7}$$

This appears to confirm that the cutoff is indeed $\Lambda = \bar{m}_{\rm Pl}/\xi$, but this conclusion is premature. Let us explain why. The index c on $\mathcal{M}$ stands for "channel"; there are s, t, and u-channels, which all scale similarly (Fig. 5-1 diagrams 1–3). When we sum over all 3 channels to get $\mathcal{M}_{\rm tot}$ and put the external particles on-shell, an amusing thing happens: they cancel. So the leading term in powers of ξ vanishes [29, 30]. The first non-zero piece is

$$\mathcal{M}_{\rm tot}(2\phi \to 2\phi) \sim \frac{E^2}{\bar{m}_{\rm Pl}^2}, \tag{5.8}$$

which shows that the true cutoff is $\Lambda = \bar{m}_{\rm Pl}$. Hence, according to $2 \to 2$ tree-level scattering the theory only fails at Planckian energies.

What about in the Einstein frame? To address this question we need to focus on the kinetic sector, which we presented earlier in eq. (5.3). By expanding for small ϕ, we have the following contributions

$$(\partial\phi)^2 + \frac{6\,\xi^2}{\bar{m}_{\rm Pl}^2}\phi^2(\partial\phi)^2 + \dots \tag{5.9}$$

The second term appears to represent a dimension 6 interaction term with cutoff $\Lambda = \bar{m}_{\rm Pl}/\xi$. We can compute $2\phi \to 2\phi$ scattering at tree-level using this single 4-point vertex (Fig. 5-1 diagram 4). We find the scaling $\mathcal{M} \sim \xi^2 E^2/\bar{m}_{\rm Pl}^2$, but when the external particles are put on-shell we find that the result is zero. This is also true at arbitrary loop order (see Fig. 5-2 for the 1-loop diagrams) and for *any* scattering process (despite the power counting estimates of Ref. [17]). The reason for this was alluded to earlier: in this single field case, we can perform a field redefinition to σ (eq. (5.4)) which carries canonical kinetic energy. Hence the kinetic sector is that of a free theory, modulo its minimal coupling to gravity, giving the correct cutoff $\Lambda = \bar{m}_{\rm Pl}$. Hence scattering amplitudes generated from purely the gravity and kinetic sectors allow the theory to be well defined up to Planckian energies. This shows that naive power counting can be misleading.

5.3.2 The Potential

Next we examine the potential V. This potential seems particularly strange in the Einstein frame, where it takes the following form

$$V_E(\sigma) = \frac{\frac{\lambda}{4}\phi(\sigma)^4}{\left(1 + \frac{\xi\phi(\sigma)^2}{\bar{m}_{\rm Pl}^2}\right)^2} = \frac{\lambda}{4}\sigma^4 - \frac{\lambda\,\xi^2}{\bar{m}_{\rm Pl}^2}\sigma^6 + \dots \tag{5.10}$$

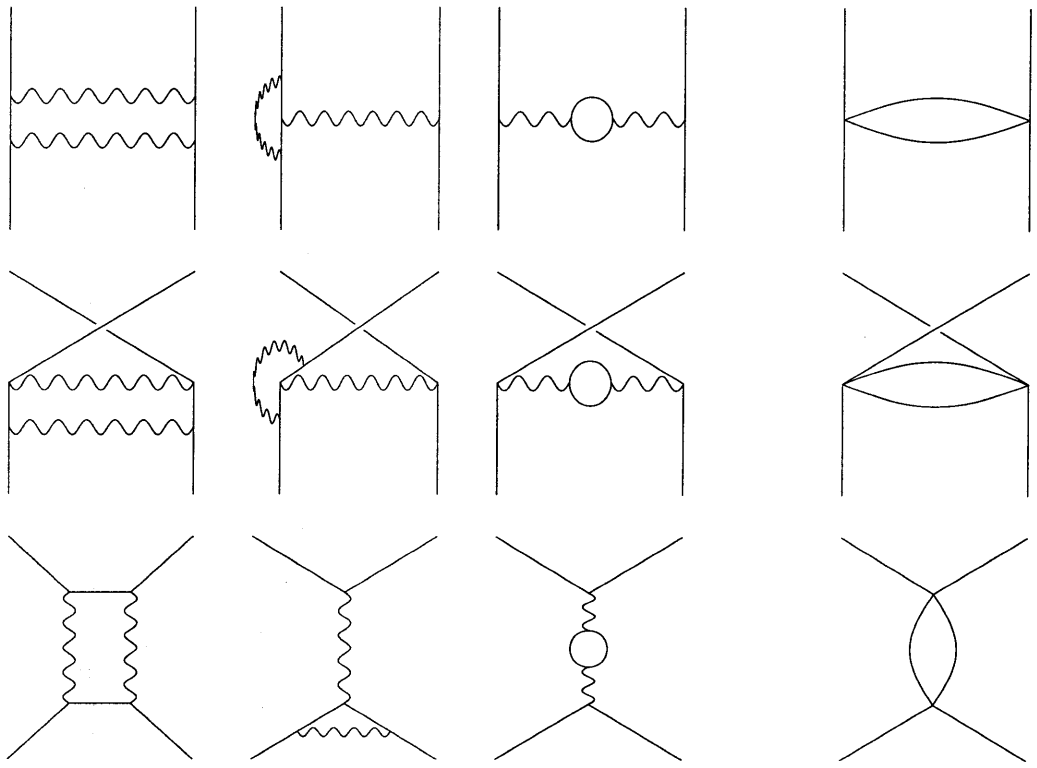

Figure 5-2: One-loop scattering $2\phi \to 2\phi$. Diagrams 1–9 (left block): In Jordan frame it is due to graviton exchange. Diagrams 10–12 (right block): In Einstein frame it is due to a 4-point vertex. Top row: t-channel, middle row: u-channel, bottom row: s-channel.

In this case it would seem that the theory breaks down for $\Lambda \sim \bar{m}_{\rm Pl}/\xi$. It is important to compute the quantum corrections. Recall the general form for the one-loop quantum corrected potential [31]

$$V_{1\text{ loop}}(\sigma) = V_{\rm cl}(\sigma) + \frac{1}{64\pi^2}\left(V''_{\rm cl}(\sigma)\right)^2 \ln(V''_{\rm cl}(\sigma)) + \dots . \tag{5.11}$$

Since the full potential is slowly varying it is simple to check that the corrections are always small, especially for small λ. Since the one-loop quantum corrections are small, there is some naive hope that this theory may make sense; this appears to be the philosophy advocated in recent studies of Higgs-inflation. On the other hand, one might take the point of view that by simply expanding around small σ, the theory fails when the dimension 6 term is comparable to the dimension 4 term [18]. The breakdown of this theory would appear in many-particle hard scattering processes.

The behavior of a scalar field theory with a slowly varying non-polynomial potential is considered an open problem in field theory. However, a perturbative analysis suggests that the theory fails at $\Lambda \sim \bar{m}_{\rm Pl}/\xi$. Note that this conclusion arises only if one studies scattering amplitudes involving vertices provided by the potential V. In particular, this necessarily requires scattering amplitudes to involve powers of λ. Such factors of λ were not included in Ref. [17], who focussed exclusively on the gravity and kinetic sectors. Hence, the scattering estimates provided in Ref. [17] do not apply in the case of a singlet scalar, since they arise exclusively from the gravity and kinetic sectors alone (as we analyzed in the previous section). However, by including corrections from the potential, the theory likely fails at $\sim \bar{m}_{\rm Pl}/\xi$.

To say it differently, the Einstein frame Lagrangian acquires a shift symmetry $\sigma \to \sigma + \sigma_0$ in the $\lambda \to 0$ limit, which protects scattering amplitudes from becoming large. This also occurs in the Jordan frame, though the symmetry is nonlinearly realized. However, finite λ breaks the symmetry and likely compromises the theory.

5.3.3 Multiple Fields

In the case of N-fields, the story is rather different. Here we can just focus on the gravity and kinetic sectors to understand that the effective theory breaks down. Let us begin the discussion again in the Jordan frame. Expanding the metric around flat space as before, we have the set of dimension 5 interactions

$$\frac{\xi}{\bar{m}_{\rm Pl}}\vec{\phi}\cdot\vec{\phi}\Box h \tag{5.12}$$

Consider particles ϕ_1 and ϕ_2. Lets compute the tree-level scattering process $\phi_1+\phi_2 \to \phi_1+\phi_2$ due to the exchange of one graviton. The scattering matrix scales as earlier

$$\mathcal{M}_c(\phi_1+\phi_2 \to \phi_1+\phi_2) \sim \frac{\xi^2 E^2}{\bar{m}^2_{\rm Pl}}. \tag{5.13}$$

But unlike the case of identical particles, this process only occurs through a single channel – the t-channel (Fig. 5-1 diagram 1). Hence there is no cancellation. This

means that the theory *does* become strongly coupled at $\Lambda = \bar{m}_{\rm Pl}/\xi$. Furthermore, the probability for a pair of different particles to scatter off each other grows without bound, violating unitarity (as noted in Refs. [17, 32]).

In the Einstein frame the kinetic sector takes the following form for small $\vec{\phi}$

$$\partial\vec{\phi}\cdot\partial\vec{\phi} + \frac{6\,\xi^2}{\bar{m}_{\rm Pl}^2}(\vec{\phi}\cdot\partial\vec{\phi})^2 + \ldots \tag{5.14}$$

It is impossible to perform a field re-definition to bring this into canonical form. The second term introduces cross terms between the different fields: $\xi^2\phi_1\partial\phi_1\phi_2\partial\phi_2/\bar{m}_{\rm Pl}^2$. As in the Jordan frame, this allows for $\phi_1 + \phi_2 \to \phi_1 + \phi_2$ scattering to take place at tree-level through a simplified vertex with $\mathcal{M} \sim \xi^2 E^2/\bar{m}_{\rm Pl}^2$. This confirms that the theory is strongly interacting at $\Lambda = \bar{m}_{\rm Pl}/\xi$.

But what if we focus purely on the scattering of identical particles $2\phi_1 \to 2\phi_1$? Then at tree-level we seem fine, due to the cancellation among the 3 channels. However, the strong coupling between ϕ_1 and ϕ_2 ensures that quantum corrections from ϕ_2 particles running in the loop are large and do not cancel among channels. In Fig. 5-2 we have included all 1PI one-loop diagrams in the Jordan frame (9 diagrams in the left block) and Einstein frame (3 diagrams in the right block). If only a single field ϕ_1 is present then cancellation occurs from summing the 9 diagrams in Jordan frame, or summing the 3 diagrams in Einstein frame. When ϕ_2 is included it only appears in the loop in the 3rd column of diagrams in the Jordan frame, and in all 3 diagrams in the Einstein frame but with a truncated vertex rule. This prevents cancellation among the diagrams.

Note that the breakdown of the theory at $\Lambda = \bar{m}_{\rm Pl}/\xi$ is due to scalar fields running in a loop, not gravitons. Hence this represents a breakdown of the matter sector, not a quantum-gravity effect per se. In order to eliminate the ensuing divergences, we require the introduction of new operators into the classical Lagrangian:

$$\frac{c_1}{\Lambda^4}(\partial\vec{\phi}\cdot\partial\vec{\phi})^2\ , \qquad \frac{c_2}{\Lambda^4}(\partial_\mu\vec{\phi}\cdot\partial_\nu\vec{\phi})(\partial^\mu\vec{\phi}\cdot\partial^\nu\vec{\phi}), \tag{5.15}$$

with $\Lambda = \bar{m}_{\rm Pl}/\xi$ and the c_i cannot be much smaller than $\mathcal{O}(4\pi)^{-2}$ without fine tuning. In addition, the quantum field theory will also generate corrections to the potential, such as $\bar{c}_1\lambda^2(\vec{\phi}\cdot\vec{\phi})^3/\Lambda^2$ and $\bar{c}_2\lambda^2(\vec{\phi}\cdot\vec{\phi})^4/\Lambda^4$. Note the factor of λ^2 here, which arises because the potential is protected by a shift symmetry in the $\lambda \to 0$ limit.

How do such operators affect the inflationary model? Lets focus on the end of inflation where $\phi \sim \bar{m}_{\rm Pl}/\sqrt{\xi}$. Here the kinetic and potential terms in the classical Lagrangian are comparable, with value $V \sim (\partial\phi)^2 \sim \lambda\,\bar{m}_{\rm Pl}^4/\xi^2$. The new dimension 8 operators would be $\sim c\,\lambda^2\bar{m}_{\rm Pl}^4$; parametrically larger than the included operators by a factor $c\,\lambda\,\xi^2$, destroying the theory for $\xi \gg 1$. This is relevant to the Standard Model Higgs which is comprised of multiple scalars.

5.4 An Analogy - the Pion Lagrangian

If this distinction between the behavior of a single scalar field theory compared to a multiple scalar field theory seems strange, let us recall a familiar case where the same is true.

Consider the Lagrangian of the σ-model

$$\mathcal{L} = -\frac{1}{2}\partial\phi_n\partial\phi_n + \frac{\mu^2}{2}\phi_n\phi_n - \frac{\lambda}{4}(\phi_n\phi_n)^2 \tag{5.16}$$

where n is summed over $n = 1, 2, \ldots, N, N+1$, with the first N components forming a vector $\vec{\phi}$ and the last component ϕ_{N+1} a scalar. After spontaneous symmetry breaking, it is useful to use fields that make it manifest that the potential only depends on the length of ϕ_n, which we call $\sigma(x) \equiv \sqrt{\sum_n \phi_n(x)^2}$, and the Goldstone bosons, which we call $\vec{\zeta}(x)$. Following Ref. [33] the original Lagrangian can be recast as

$$\mathcal{L} = -\frac{1}{2}(\partial\sigma)^2 - 2\sigma^2\frac{\partial\vec{\zeta}\cdot\partial\vec{\zeta}}{(1+\vec{\zeta}\cdot\vec{\zeta})^2} + \frac{\mu^2}{2}\sigma^2 - \frac{\lambda}{4}\sigma^4. \tag{5.17}$$

At low energies, the massive field σ will relax to its minimum at $\langle\sigma\rangle = \mu/\sqrt{\lambda}$. By defining $F \equiv 2\langle\sigma\rangle$ and introducing properly normalized Goldstone bosons $\vec{\pi}(x) \equiv F\vec{\zeta}(x)$, the low energy Lagrangian for the Goldstone bosons is

$$\mathcal{L} = -\frac{1}{2}\frac{\partial\vec{\pi}\cdot\partial\vec{\pi}}{(1+\vec{\pi}\cdot\vec{\pi}/F^2)^2}. \tag{5.18}$$

For 3 pions it is well known that this theory breaks down for energies much larger than F, which can be confirmed by computing scattering amplitudes. A rough estimate for the breakdown of the quantum theory is $\Lambda = 4\pi F$, where the 4π comes from loop integrals.

The $2 \to 2$ tree-level scattering process has the matrix element

$$\begin{aligned}&\mathcal{M}(\pi_a + \pi_b \to \pi_c + \pi_d)\\ &= 4F^{-2}\left[\delta_{ab}\delta_{cd}(-p_ap_b - p_cp_d) + \delta_{ac}\delta_{bd}(p_ap_c + p_bp_d) + \delta_{ad}\delta_{bc}(p_ap_d + p_bp_c)\right]\end{aligned} \tag{5.19}$$

which is non-zero whenever some of the a, b, c, d are different (scaling as $\sim E^2/F^2$), but is zero if $a = b = c = d$ and the π's are put on-shell ($p^2 = 0$). The latter is the case for a single field. The same is true for loop corrections. As earlier, what is special about the $N = 1$ case is that the Lagrangian (5.18) is really a free field theory in disguise (which can be made manifest through a field redefinition), but it is unambiguously an interacting field theory for $N > 1$.

On the other hand, even for $N = 1$, if we explicitly break the global symmetry, then the (pseudo)-Goldstone bosons acquire a mass and a potential V. This potential is non-polynomial and likely causes the theory to fail for $E \gg F$.

5.5 Curvature-Squared Models

Let us now turn to another type of non-standard model for inflation, which comes from the inclusion of curvature-squared terms in the action [34]

$$S = \int d^4x\sqrt{-g}\left[\frac{1}{2}\bar{m}_{\rm Pl}^2\mathcal{R} + \zeta\,\mathcal{R}^2\right], \tag{5.20}$$

with $\zeta = \mathcal{O}(10^8)$ to achieve the correct amplitude of density fluctuations. Although we consider this model to be ad hoc and not necessarily "natural", it is still worthwhile to discuss whether this theory makes sensible inflationary predictions or not.

Since $\zeta \gg 1$ we again should ask if scattering amplitudes becoming large at energies well below $\bar{m}_{\rm Pl}$. In particular, by expanding around flat space, the curvature-squared term introduces the following dimension 6 and 7 operators into the action

$$\frac{\zeta}{\bar{m}_{\rm Pl}^2}(\Box h)^2, \qquad \frac{\zeta}{\bar{m}_{\rm Pl}^3}(\Box h)^2 h, \tag{5.21}$$

respectively. The former modifies the graviton propagator at high energies. The latter introduces a 3-point vertex, which can be used to construct tree-level graviton-graviton scattering processes. In particular, each of the t, u, and s-channels scale to leading order as

$$\mathcal{M}_c(2\,h_{\mu\nu} \to 2\,h_{\mu\nu}) \sim \frac{\zeta^2 E^6}{\bar{m}_{\rm Pl}^6}, \tag{5.22}$$

suggesting that the theory violates unitarity above the cutoff $\Lambda = \bar{m}_{\rm Pl}/\zeta^{1/3}$, as claimed recently in [17] and criticized on that basis. However, as was the case in Section 5.3.1, if we sum over all diagrams, then this piece vanishes. Instead, the leading non-zero piece scales as $\mathcal{M}_{\rm tot} \sim (E/\bar{m}_{\rm Pl})^2$. (Note that there is also a 4-point graviton vertex and that the correction to the graviton propagator is important here). So according to the tree-level analysis of $2 \to 2$ scattering, the correct cutoff is $\Lambda = \bar{m}_{\rm Pl}$.

It is not obvious how to compute all quantum corrections to this theory in the Jordan frame, since it is a higher derivative theory due to the $\mathcal{R}^2$ term. It has been argued in the literature that it is best to define the quantum theory in the frame in which the degrees of freedom are manifest, which is the Einstein frame. Ref. [17] claimed that this theory fails equally in any frame, including the Einstein frame. So lets re-write the action as a scalar-tensor theory. This involves a canonical gravity sector, a canonical kinetic sector for some scalar σ, and a potential $V_E(\sigma)$ which exhibits $\mathcal{O}(1)$ variations on $\Delta\sigma \sim \bar{m}_{\rm Pl}$. It can be shown that this gives

$$S = \int d^4x\sqrt{-g_E}\left[\frac{1}{2}\bar{m}_{\rm Pl}^2\mathcal{R}_E - \frac{1}{2}(\partial\sigma)^2 - \frac{\bar{m}_{\rm Pl}^4}{16\,\zeta}\left(1 - \exp(-\sqrt{2/3}\,\sigma/\bar{m}_{\rm Pl})\right)^2\right]. \tag{5.23}$$

This action seems rather innocuous, but according to Ref. [17] this theory still has a cutoff at $\Lambda = \bar{m}_{\rm Pl}/\zeta^{1/3}$. We begin by explaining how one could arrive at this conclusion by a naive power counting estimate. We then explain why this is incorrect and why the correct answer is $\Lambda = \bar{m}_{\rm Pl}$.

Let us expand the potential about its minimum at $\sigma = 0$: $V_E(\sigma) = \frac{\bar{m}_{\rm Pl}^2}{24\zeta}\sigma^2 + \ldots$ We see that the field σ has a mass-squared of $m_\sigma^2 = \bar{m}_{\rm Pl}^2/(12\,\zeta)$. This non-zero mass will obviously play an important role in computing low energy scattering amplitudes. For instance, consider the 1-loop scattering process involving $2 \to 2$ graviton scattering due to σ running in a loop. We use 4 insertions of the following 3-point interaction term from the kinetic sector

$$\sim \frac{h}{\bar{m}_{\rm Pl}}(\partial\sigma)^2. \tag{5.24}$$

This gives rise to the following contribution to the matrix element

$$\mathcal{M}(2\,h_{\mu\nu} \to 2\,h_{\mu\nu}) \sim \frac{1}{\bar{m}_{\rm Pl}^4}\int d^4q \frac{q_1\cdot q_2\;\; q_2\cdot q_3\;\; q_3\cdot q_4\;\; q_4\cdot q_1}{(q_1^2-m_\sigma^2)(q_2^2-m_\sigma^2)(q_3^2-m_\sigma^2)(q_4^2-m_\sigma^2)}. \tag{5.25}$$

This loop integral can be estimated using, for instance, dimensional regularization. Its scaling depends critically on the energy scale. For input energy $E \ll m_\sigma$, the denominator can be approximated as m_σ^8. If one were to then approximate the numerator with the high energy behavior, the finite part would be estimated as

$$\mathcal{M}(2\,h_{\mu\nu} \to 2\,h_{\mu\nu}) \sim \frac{E^{12}}{\bar{m}_{\rm Pl}^4 m_\sigma^8} \sim \frac{\gamma^4 E^{12}}{\bar{m}_{\rm Pl}^{12}}. \tag{5.26}$$

If this scaling were correct up to arbitrarily high energies, then indeed the theory would have a cutoff at $\Lambda = \bar{m}_{\rm Pl}/\zeta^{1/3}$. However, this scaling is invalid at high energies. Instead, for $E \gg m_\sigma$, the matrix element scales as

$$\mathcal{M}(2\,h_{\mu\nu} \to 2\,h_{\mu\nu}) \sim \frac{E^4}{\bar{m}_{\rm Pl}^4}, \tag{5.27}$$

indicating that the correct cutoff is $\Lambda = \bar{m}_{\rm Pl}$. A similar argument for the cutoff goes through for *any* scattering process, including tree-level or arbitrary loop level.

Hence, if low energy scattering amplitudes are taken at face value, they naively suggest that the theory has a cutoff at $\bar{m}_{\rm Pl}/\zeta^{1/3}$. However, a full calculation at the appropriate energy scale reveals that the correct cutoff of the scalar-tensor theory defined by (5.23) is $\bar{m}_{\rm Pl}$. This means that the effective field theory upon which the inflationary predictions are based makes sense.

5.6 Conclusions

Treated classically, a singlet scalar ϕ with non-minimal coupling to gravity $\xi\phi^2\mathcal{R}$ and $\xi \gg 1$ can drive a phase of slow-roll inflation. By examining the gravity and kinetic sectors alone, as in Ref. [17], it is tempting to declare that such theories become strongly interacting at $\bar{m}_{\rm Pl}/\xi$, which would spoil the model. Such a conclusion arises from power counting estimates of scattering processes in the Jordan frame without summing all diagrams to check for any possible cancellations. Here we have shown that cancellations do occur in the single field case, giving a Planckian cutoff. However,

the potential sector is problematic: the Einstein frame potential is non-polynomial, varying on scales $\Delta\sigma \sim \bar{m}_{\rm Pl}/\xi$. Although its Coleman-Weinberg quantum corrections [31] are small and the behavior of such a theory is considered an open problem in field theory, it is doubtful if high energy scattering amplitudes could be well behaved and if there is any sensible UV completion of the theory. This suggests that the theory likely fails.

On the other hand, these power counting estimates in the gravity and kinetic sectors *do* apply in the multi-field case: a vector of fields $\vec{\phi}$ is problematic as $\phi_a + \phi_b \to \phi_c + \phi_d$ scattering becomes strong at energies $E \gtrsim \bar{m}_{\rm Pl}/\xi$ relevant to inflation (as noted in [17, 32]) and large quantum corrections are generated. This requires new physics to intervene or that something peculiar happens in the RG flow that rescues the theory, such as flow to a UV fixed point. But the latter scenario seems rather unlikely. Our conclusions were shown to be true in both the Jordan and Einstein frames.

So if the Standard Model Higgs were a gauge singlet, then Higgs-inflation would be well behaved in the pure gravity and kinetic sectors, requiring a detailed analysis of the potential sector to draw conclusions about the quantum theory. However, since it is comprised of 4 real scalars in a complex doublet, it falls into the multi-field category and fails even at the level of tree-level scattering due to graviton exchange; breaking down due to Goldstone bosons. Explicitly identifying this difference between the single field case and the multi-field case was previously missed by us [15] and others. Similarly, this may spoil other attempts to embed inflation into particle physics models with multiple scalar fields [35], unless the underlying UV theory carries an appropriate structure.

We also examined curvature-squared models $\zeta \mathcal{R}^2$ [34] with $\zeta \gg 1$ (although we consider such models to be ad hoc). They were recently criticized in Ref. [17] for failing to even make sense at scales $E \gtrsim \bar{m}_{\rm Pl}/\zeta^{1/3}$. Again there exists cancellation among tree-level diagrams, suggesting a Planckian cutoff. Furthermore, by formulating the quantum theory as a scalar-tensor theory in the Einstein frame (which makes the degrees of freedom manifest) we showed that the inflationary theory is well behaved. In this case, all scattering processes, whether they arise from the gravity or kinetic or potential sectors, satisfy unitarity bounds and generate very small quantum corrections for energies below $\bar{m}_{\rm Pl}$.

Bibliography

[1] A. Guth, "The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems", Phys. Rev. D, 23, 347 (1981).

[2] A. D. Linde, "A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems", Phys. Lett. B, 108, 389 (1982).

[3] A. Albrecht and P. J. Steinhardt, "Reheating an Inflationary Universe", Phys. Rev. Lett., 48, 1220 (1982)

[4] A. D. Linde, "Chaotic Inflation", Phys. Lett. B, 129, 177 (1983)

[5] D. Baumann, A. Dymarsky, I. R. Klebanov, L. McAllister and P. J. Steinhardt, "A Delicate Universe," Phys. Rev. Lett. **99** (2007) 141601 [arXiv:0705.3837 [hep-th]].

[6] D. S. Salopek, J. R. Bond and J. M. Bardeen, "Designing Density Fluctuation Spectra in Inflation," Phys. Rev. D **40**, 1753 (1989).

[7] R. Fakir and W. G. Unruh, "Improvement on cosmological chaotic inflation through nonminimal coupling," Phys. Rev. D **41**, 1783 (1990).

[8] D. I. Kaiser, "Primordial spectral indices from generalized Einstein theories," Phys. Rev. D **52** (1995) 4295 [arXiv:astro-ph/9408044].

[9] E. Komatsu and T. Futamase, "Complete constraints on a nonminimally coupled chaotic inflationary scenario from the cosmic microwave background," Phys. Rev. D **59** (1999) 064029 [arXiv:astro-ph/9901127].

[10] K. Nozari and S. D. Sadatian, "Non-Minimal Inflation after WMAP3," Mod. Phys. Lett. A **23** (2008) 2933 [arXiv:0710.0058 [astro-ph]].

[11] F. L. Bezrukov and M. Shaposhnikov, "The Standard Model Higgs boson as the inflaton," Phys. Lett. B **659**, 703 (2008) [arXiv:0710.3755 [hep-th]].

[12] A. O. Barvinsky, A. Y. Kamenshchik and A. A. Starobinsky, "Inflation scenario via the Standard Model Higgs boson and LHC," JCAP **0811** (2008) 021 [arXiv:0809.2104 [hep-ph]].

[13] F. Bezrukov, D. Gorbunov and M. Shaposhnikov, "On initial conditions for the Hot Big Bang," arXiv:0812.3622 [hep-ph].

[14] J. Garcia-Bellido, D. G. Figueroa and J. Rubio, "Preheating in the Standard Model with the Higgs-Inflaton coupled to gravity," arXiv:0812.4624 [hep-ph].

[15] A. De Simone, M. P. Hertzberg and F. Wilczek, Phys. Lett. B **678** (2009) 1 [arXiv:0812.4946 [hep-ph]].

[16] F. L. Bezrukov, A. Magnin and M. Shaposhnikov, "Standard Model Higgs boson mass from inflation," arXiv:0812.4950 [hep-ph].

[17] C. P. Burgess, H. M. Lee and M. Trott, "Power-counting and the Validity of the Classical Approximation During Inflation," arXiv:0902.4465 [hep-ph].

[18] J. L. F. Barbon and J. R. Espinosa, "On the Naturalness of Higgs Inflation," arXiv:0903.0355 [hep-ph].

[19] F. Bezrukov and M. Shaposhnikov, "Standard Model Higgs boson mass from inflation: two loop analysis," JHEP **0907** (2009) 089 [arXiv:0904.1537 [hep-ph]].

[20] A. O. Barvinsky, A. Y. Kamenshchik, C. Kiefer, A. A. Starobinsky and C. Steinwachs, "Asymptotic freedom in inflationary cosmology with a non-minimally coupled Higgs field," JCAP **0912** (2009) 003 [arXiv:0904.1698 [hep-ph]].

[21] T. E. Clark, B. Liu, S. T. Love and T. ter Veldhuis, "The Standard Model Higgs Boson-Inflaton and Dark Matter," Phys. Rev. D **80** (2009) 075019 [arXiv:0906.5595 [hep-ph]].

[22] R. N. Lerner and J. McDonald, "Gauge singlet scalar as inflaton and thermal relic dark matter," Phys. Rev. D **80**, 123507 (2009) [arXiv:0909.0520 [hep-ph]].

[23] A. O. Barvinsky, A. Y. Kamenshchik, C. Kiefer, A. A. Starobinsky and C. F. Steinwachs, "Higgs boson, renormalization group, and cosmology," arXiv:0910.1041 [hep-ph].

[24] D. G. Figueroa, "Preheating the Universe from the Standard Model Higgs," arXiv:0911.1465 [hep-ph].

[25] N. Okada, M. U. Rehman and Q. Shafi, "Running Standard Model Inflation And Type I Seesaw," arXiv:0911.5073 [hep-ph].

[26] M. B. Einhorn and D. R. T. Jones, "Inflation with Non-minimal Gravitational Couplings and Supergravity," arXiv:0912.2718 [hep-ph].

[27] R. N. Lerner and J. McDonald, "Higgs Inflation and Naturalness," arXiv:0912.5463 [hep-ph].

[28] A. Mazumdar and J. Rocher, "Particle physics models of inflation and curvaton scenarios," arXiv:1001.0993 [hep-ph].

[29] S. R. Huggins and D. J. Toms "One-graviton exchange interaction of non-minimally coupled scalar fields," Class. Quantum Grav. 4 (1987) 1509-1513.

[30] S. R. Huggins, "Cross sections from tree-level gravitational scattering from a non-minimally coupled scalar field," Class. Quantum Grav. 4 (1987) 1515-1523.

[31] S. Coleman "Aspects of Symmetry", Selected Erice Lectures, Cambridge University Press (1985).

[32] M. Atkins and X. Calmet, "On the unitarity of linearized General Relativity coupled to matter," arXiv:1002.0003 [hep-th].

[33] S. Weinberg, "The Quantum Theory of Fields", Volume 2 Modern Applications, Cambridge University Press (1996).

[34] A. A. Starobinsky, "A new type of isotropic cosmological models without singularity," Phys. Lett. B **91** (1980) 99.

[35] M. P. Hertzberg, S. Kachru, W. Taylor and M. Tegmark, "Inflationary Constraints on Type IIA String Theory," JHEP **0712** (2007) 095 [arXiv:0711.2512 [hep-th]].

Chapter 6

Quantum Radiation of Oscillons

Many classical scalar field theories possess remarkable solutions: coherently oscillating, localized clumps, known as oscillons. In many cases, the decay rate of classical small amplitude oscillons is known to be exponentially suppressed and so they are extremely long lived. In this work we compute the decay rate of quantized oscillons. We find it to be a power law in the amplitude and couplings of the theory. Therefore, the quantum decay rate is very different to the classical decay rate and is often dominant. We show that essentially all oscillons eventually decay by producing outgoing radiation. In single field theories the outgoing radiation has typically linear growth, while if the oscillon is coupled to other bosons the outgoing radiation can have exponential growth. The latter is a form of parametric resonance: explosive energy transfer from a localized clump into daughter fields. This may lead to interesting phenomenology in the early universe. Our results are obtained from a perturbative analysis, a non-perturbative Floquet analysis, and numerics.

6.1 Introduction

Although there is no direct evidence yet, there are good reasons to think that scalar fields are plentiful in nature, such as the Higgs boson, axion, inflaton, p/reheating fields, moduli, squarks, sleptons, etc. In recent years, there has been increasing interest in certain kinds of long lived structures that scalar fields can support. In particular, under fairly broad conditions, if a massive scalar field ϕ possesses a non-linear self-interacting potential, such as $\phi^3 + \ldots$ or $-\phi^4 + \ldots$, then it can support coherently oscillating localized clumps, known as oscillons. Due to their oscillations in time and localization in space, they are time-dependent solitons.

Remarkably, oscillons have a long lifetime, i.e., they may live for very many oscillations, despite the absence of any internal conserved charges. Related objects are Q-balls [1], which do carry a conserved charge. If the field ϕ is promoted to a complex scalar carrying a U(1) global symmetry $\phi \to e^{i\alpha}\phi$, then oscillons correspond to oscillatory radial motion in the complex ϕ plane, while Q-balls correspond to circular motion in the complex ϕ-plane. Since oscillons do not require the U(1) symmetry for their existence (in fact they are typically comprised of only a real scalar, as we will

assume), then oscillons appear much more generically than Q-balls.

As classical solutions of nonlinear equations of motion, oscillons (related to "quasi-breathers") possess an asymptotic expansion which is exactly periodic in time and localized in space, characterized by a small parameter ϵ. In a seminal paper Segur and Kruskal [2] showed that the asymptotic expansion is not an exact solution of the equations of motion and in fact misses an exponentially small radiating tail, which they computed. They showed that oscillons decay into outgoing radiation at an exponentially suppressed rate $\sim \exp(-b/\epsilon)$, where $b = \mathcal{O}(1)$ is model dependent. Although their work was in 1 spatial dimension, similar results have been obtained in 2 and 3 spatial dimensions [3]. Furthermore, various long lived oscillons have been found in different contexts, including the standard model [4, 5, 6], abelian-Higgs models [7], axion models [8], in an expanding universe [9, 10], during phase transitions [11, 12, 13, 14], domain walls [15], gravitational systems [16, 17, 18] (called "oscillatons"), large $+\phi^6$ models [19], in other dimensions [20], etc, showing that oscillons are generic and robust.

In almost all cases, investigations into oscillons have thus far been at the classical level. There is a good reason for this approximation: the mass of an oscillon $M_{\rm osc}$ is typically much greater than the mass of the individual particles m_ϕ. This means that the spectrum is almost continuous and classical. But does this imply that every property of the oscillon, including its lifetime, is adequately described by the classical theory? To focus the discussion, suppose a scalar field ϕ in $d+1$ dimensional Minkowski space-time has a potential of the form

$$V(\phi) = \frac{1}{2}m^2\phi^2 + \frac{\lambda_3}{3!}m^{\frac{5-d}{2}}\phi^3 + \frac{\lambda_4}{4!}m^{3-d}\phi^4 + \dots \tag{6.1}$$

Since we will keep explicit track of $\hbar$ in this chapter, it has to be understood that m here has units T^{-1}; the quanta have mass $m_\phi = \hbar m$, and $\{\lambda_3^2, \lambda_4\}$ have units $\hbar^{-1}$. It is known that for $\lambda \equiv \frac{5}{3}\lambda_3^2 - \lambda_4 > 0$ such a theory possesses classical oscillons that are characterized by a small dimensionless parameter ϵ and have a mass given by $M_{\rm osc} \sim \frac{m}{\lambda} l(\epsilon) = \frac{m_\phi}{\lambda\hbar} l(\epsilon)$, where $l(\epsilon) = \epsilon^{2-d}$ in the standard expansion that we will describe in Section 6.2, but can scale differently in other models (e.g., see [18]). So if $\lambda\hbar \ll l(\epsilon)$, then $M_{\rm osc} \gg m_\phi$.

In such a regime we expect that various properties of the oscillon are well described classically, such as its size and shape. In this work we examine whether the same is true for the oscillon lifetime. We find that although the oscillon lifetime is exponentially long lived classically, it has a power law lifetime in the quantum theory controlled by the "effective $\hbar$" – in the above case this is $\lambda_3^2\hbar$ or $\lambda_4\hbar$.

In this chapter we treat the oscillon as a classical space-time dependent background, as defined by the ϵ expansion, and quantize field/s in this background. We go to leading order in $\hbar$ in the quantum theory. We find that oscillon's decay through the emission of radiation with wavenumbers $k = \mathcal{O}(m)$. We show that the decay is a power law in ϵ and the couplings, and we explain why this is exponentially suppressed classically. Our analysis is done for both single field theories, where the emitted radiation typically grows at a linear rate corresponding to $3\phi \to 2\phi$ or $4\phi \to 2\phi$

annihilation processes, and for multi-field theories, where the emitted radiation often grows at an exponential rate corresponding to $\phi \to 2\chi$ decay or $2\phi \to 2\chi$ annihilation processess. We calculate the quantum decay rates in several models, which is supported by numerical investigations, but our work is also qualitative and of general validity. We also comment on collapse instabilities for $k = \mathcal{O}(\epsilon m)$ modes, whose existence is model dependent.

The outline of this chapter is as follows: In Section 6.2 we start with a review of classical oscillons and describe their exponentially suppressed decay in Section 6.3. In Section 6.4 we outline the semi-classical quantization of oscillons and derive the decay rate of oscillons in Section 6.5. Having started with single field models, we move on to examine the effects of coupling to other fields in Section 6.6. In Section 6.7 we discuss when the decay products grow linearly in time and when it is exponential. Here we demonstrate that coupled fields can achieve (depending on parameters) explosive energy transfer, which may have some cosmological relevance. We comment on collapse instabilities in Section 6.8 and conclude in Section 6.9.

6.2 Classical Oscillons

In this section we review the asymptotic expansion for a single field oscillon, valid in 1, 2, and 3 dimensions. We then explain why the oscillon slowly radiates at an exponentially suppressed rate.

Consider a single scalar field ϕ in $d+1$ dimensional Minkowski space-time (signature $+---\ldots$) with classical action

$$S = \int d^{d+1}x \left[\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - V_I(\phi)\right], \tag{6.2}$$

where $V_I(\phi)$ is a nonlinear interaction potential. Here m has units T^{-1}. For simplicity we measure time in units of $1/m$, so without loss of generality, we set $m = 1$ from now on in the chapter, unless otherwise stated. The classical equation of motion is

$$\ddot{\phi} - \nabla^2\phi + \phi + V_I'(\phi) = 0. \tag{6.3}$$

In this chapter we will focus on the following types of interaction potentials

$$V_I(\phi) \;=\; \frac{\lambda_3}{3!}\phi^3 + \frac{\lambda_4}{4!}\phi^4 + \ldots, \tag{6.4}$$

where $\lambda \equiv \frac{5}{3}\lambda_3^2 - \lambda_4 > 0$ is assumed. This includes the cases (i) $\lambda_3 > 0$ and $\lambda_4 > 0$, which will occur for a generic symmetry breaking potential, and (ii) $\lambda_3 = 0$ and $\lambda_4 < 0$, which is relevant to examples such as the axion. In case (ii) higher order terms in the potential, such as $+\phi^6$ are needed for stabilization. Note that in both cases the interaction term causes the total potential to be reduced from the pure ϕ^2 parabola. This occurs for $\phi < 0$ in the ϕ^3 case, and for $|\phi| > 0$ in the ϕ^4 case. It is straightforward to check that a ball rolling in such a potential will oscillate at a

frequency *lower* than $m = 1$. If the field ϕ has spatial structure, then one can imagine a situation in which the gradient term in eq. (6.3) balances the nonlinear terms, and a localized structure oscillates at such a low frequency – this is the oscillon.

Naively, this low frequency oscillation cannot couple to normal dispersive modes in the system and should be stable. Higher harmonics are generated by the nonlinearity, but resonances can be cancelled order by order in a small amplitude expansion. Let us briefly explain how this works – more detailed descriptions can be found in Refs. [21, 22].

In order to find periodic and spatially localized solutions, it is useful to rescale time t and lengths $x = |\mathbf{x}|$ to

$$\tau = t\sqrt{1-\epsilon^2}, \qquad \rho = x\,\epsilon, \tag{6.5}$$

with $0 < \epsilon \ll 1$ a small dimensionless parameter. Here we search for spherically symmetric solutions. The equation of motion (6.3) becomes

$$(1-\epsilon^2)\partial_{\tau\tau}\phi - \epsilon^2\left(\partial_{\rho\rho}\phi + \frac{d-1}{\rho}\partial_\rho\phi\right) + \phi + V_I'(\phi) = 0 \tag{6.6}$$

To obtain an oscillon solution, the field ϕ is expanded as an asymptotic series in powers of ϵ as

$$\phi_{\rm osc}(\rho,\tau) = \sum_{n=1}^{\infty} \epsilon^n\,\phi_n(\rho,\tau). \tag{6.7}$$

The set of n depends on $V_I(\phi)$. For $V_I \sim \lambda_3\,\phi^3$; $n = 1, 2, 3, \ldots$, and for $V_I \sim -|\lambda_4|\,\phi^4$; $n = 1, 3, 5, \ldots$. Upon substitution of the series into eq. (6.6), the leading term must satisfy

$$\partial_{\tau\tau}\phi_1 + \phi_1 = 0, \tag{6.8}$$

with solution $\phi_1 = f(\rho)\cos\tau$, where $f(\rho)$ is some spatial profile. Since $\tau = t\sqrt{1-\epsilon^2}$ the fundamental frequency of oscillation is evidently $\omega = \sqrt{1-\epsilon^2} < 1$.

The next order terms in the expansion must not be resonantly driven by ϕ_1, or the solution would not be periodic. By writing down the equations for ϕ_2 and ϕ_3 and demanding that the driving terms are non-resonant, we establish an ODE for $f(\rho)$. Extracting the λ dependence by defining $f(\rho) = 4\tilde{f}(\rho)/\sqrt{\lambda}$, the ODE is found to be

$$\partial_{\rho\rho}\tilde{f} + \frac{d-1}{\rho}\partial_\rho\tilde{f} - \tilde{f} + 2\tilde{f}^3 = 0. \tag{6.9}$$

This ODE possesses a localized solution for $d = 1, 2, 3$. In $d = 1$ the solution is known analytically $\tilde{f}(\rho) = \operatorname{sech}\rho$, but is only known numerically in $d = 2$ and $d = 3$. (For $d = 2, 3$ there are infinitely many solutions, of which we take the fundamental solution). We plot $\tilde{f}(\rho)$ in Figure 6-1. Altogether this gives the leading order term

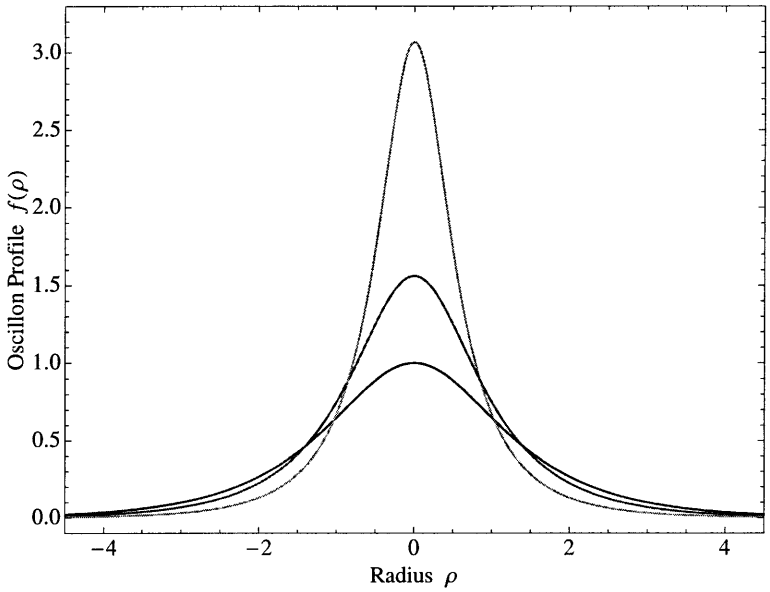

Figure 6-1: Leading order oscillon profile $\tilde{f}(\rho)$ in $d = 1$ (blue), $d = 2$ (red), and $d = 3$ (green) dimensions. We allow ρ to take on both positive and negative values here, i.e., a 1-d slice through the origin.

for the oscillon

$$\phi_{\rm osc}(\rho,\tau) = \frac{4\,\epsilon}{\sqrt{\lambda}}\tilde{f}(\rho)\cos\tau + \sum_{n>1}^{\infty}\epsilon^n\,\phi_n(\rho,\tau), \tag{6.10}$$

where higher order terms can be obtained in a similar manner. At any finite order in ϵ this provides a periodic and spatially localized oscillon. By integrating the oscillon's energy density over all space, the total mass of the oscillon is found to scale to leading order in ϵ as $M_{\rm osc} \sim m\,\epsilon^{2-d}/\lambda$.

6.3 Classical Radiation

In 1987 Segur and Kruskal [2] found that the above asymptotic expansion is not exact, as it misses an exponentially small radiating tail. They computed this using matched expansions between the inner core of the oscillon and infinity, involving some detailed analysis. Here we describe the physical origin of this classical radiation in simple terms.

The existence of outgoing radiation is ultimately tied to the fact that the oscillon expansion in not an exact solution of the equations of motion, it is only an asymptotic expansion which is correct order by order in ϵ, but not beyond all orders. Consider the oscillon expansion, truncated to order N, i.e.,

$$\phi_{\rm osc}(x,t) = \sum_{n=1}^{N}\epsilon^n\,\phi_n(x,t). \tag{6.11}$$

Lets substitute this back into the equation of motion. This will involve many harmonics. By construction the fundamental mode $\cos\tau = \cos(\omega\,t)$ will cancel, since the spatial structure was organized in order to avoid such a resonance. However, we do obtain remainders including higher harmonics. This leads to the following remainder

$$J(x,t) = j(x)\cos(\bar{n}\,\omega\,t) + \ldots, \tag{6.12}$$

where for $d = 1$ we have

$$j(x) = C_N\epsilon^{N+2}\mathrm{sech}^{N+2}(x\epsilon) + \ldots, \tag{6.13}$$

where we have included only the coefficient of the next harmonic, with some coefficient C_N. $\bar{n} = 2$ for asymmetric potentials and $\bar{n} = 3$ for symmetric potentials.

Let us now decompose a solution of the equations of motion $\phi_{\rm sol}$ into an oscillon piece and a correction δ

$$\phi_{\rm sol}(x,t) = \phi_{\rm osc}(x,t) + \delta(x,t) \tag{6.14}$$

with δ taking us from the oscillon expansion to the nearest solution. We find that

the remainder $J(x,t)$ acts as a source for the correction δ

$$\ddot{\delta} - \nabla^2\delta + \delta = -J(x,t) \tag{6.15}$$

where we have ignored non-linear terms (δ^2 etc) and a parametric driving term $V''(\phi_{\rm osc})\delta$ (although such a term will be important in the quantum theory). The solution in the far distance regime is obtained by standard means

$$\delta(x,t) = -\int \frac{d^d k\, d\omega}{(2\pi)^{d+1}} \frac{J(k,\omega)e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}}{-\omega^2 + k^2 + 1 \pm i.0^+} \tag{6.16}$$

$$\sim \frac{\cos(k_{\rm rad}x + \gamma)}{x^{(d-1)/2}} \cos(\omega_{\rm rad}t) j(k_{\rm rad}), \tag{6.17}$$

as $x \to \infty$ (γ is a phase), suggesting that there is a radiation tail with an amplitude determined by the Fourier transform of the spatial structure of the source, evaluated at $k_{\rm rad}$. The kinematics of (6.16) says that the radiation tail has frequency and wavenumber:

$$w_{\rm rad} = \bar{n}\,\omega, \quad k_{\rm rad} = \sqrt{\bar{n}^2\omega^2 - 1}, \tag{6.18}$$

where $\omega = \sqrt{1-\epsilon^2}$ is the fundamental frequency of the oscillon. This is the so-called "quasi-breather" which is used to match onto a radiating oscillon [3].

Let us now evaluate $j(k)$. In $d = 1$ we find

$$j(k) = C_N k^{N+1} \mathrm{sech}\left(\frac{\pi\, k}{2\,\epsilon}\right) + \ldots \tag{6.19}$$

So for $k = \mathcal{O}(\epsilon)$ the remainder is a power law in ϵ, but for $k = \mathcal{O}(1)$, the prefactor is $\mathcal{O}(1)$ with a sech function evaluated in its tail. For $k = \mathcal{O}(1)$, which is the case for $k = k_{\rm rad}$, we note that the source is comparable to the Fourier transform of the oscillon itself, i.e.,

$$j(k) \sim \phi_{\rm osc}(k) \tag{6.20}$$

Lets now compute the spatial Fourier transform of the oscillon. Taking only the leading order piece from eq. (6.10) at $t = 0$ we have

$$\phi_{\rm osc}(k) = \frac{4\,\epsilon}{\sqrt{\lambda}} \int d^d x\, \tilde{f}(x\epsilon) e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{6.21}$$

For $d = 1$ we have $\tilde{f}(\rho) = \mathrm{sech}\,\rho$, allowing us to compute the Fourier transform analytically:

$$\phi_{\rm osc}(k) = \frac{4\pi}{\sqrt{\lambda}} \mathrm{sech}\left(\frac{\pi\, k}{2\,\epsilon}\right). \tag{6.22}$$

For $k \gg \epsilon$ this is exponentially small. The same is true in $d = 2$ and $d = 3$ which can be computed numerically. We can summarize the behavior for $d = 1, 2, 3$ for $k \gg \epsilon$

by the following scaling

$$\phi_{\rm osc}(k) \sim \frac{1}{\sqrt{\lambda}\,(\epsilon\,k)^{(d-1)/2}} \exp\left(-\frac{c_d\,k}{\epsilon}\right), \tag{6.23}$$

where the constant in the argument of the exponential is dimensional dependent: $c_1 = \pi/2$, $c_2 \approx 1.1$, and $c_3 \approx 0.6$.[1] If we evaluate this at the relevant mode $k = k_{\rm rad} \approx \sqrt{\bar{n}^2 - 1}$, the Fourier amplitude is exponentially small as $\epsilon \to 0$.

Hence the longevity of the oscillon is due to the fact that the spatial structure is dominated by small $k = \mathcal{O}(\epsilon)$ wavenumbers, while any outgoing radiation occurs through $k = \mathcal{O}(1)$ wavenumbers, whose amplitude is suppressed. The rate at which energy is lost scales as

$$\left|\frac{dE_{\rm osc}}{dt}\right| \sim |\phi_{\rm osc}(k_{\rm rad})|^2 \sim \frac{1}{\lambda\,\epsilon^{d-1}} \exp\left(-\frac{b}{\epsilon}\right), \tag{6.24}$$

where $b \equiv 2\,c_d\,\sqrt{\bar{n}^2 - 1}$ is an $\mathcal{O}(1)$ number.

Now the energy of an oscillon scales as $E_{\rm osc} \sim 1/(\lambda\,\epsilon^{d-2})$, so the decay rate $\Gamma_d = |E_{\rm osc}^{-1}\,dE_{\rm osc}/dt|$ scales as $\Gamma_d \sim \epsilon^{-1}\exp(-b/\epsilon)$. The constant of proportionality in Γ_d is non-trivial to obtain. The reason is the following: at any finite order in the ϵ expansion, the corresponding terms in the equations of motion are made to vanish due to the spatial structure of the solution. However, there is always a residual piece left over which is $\mathcal{O}(\epsilon^0)$ in k-space, multiplied by $\phi_{\rm osc}(k)$. The limit of this residual piece as we go to higher and higher order encodes the constant, but we do not pursue that here. Alternate methods for obtaining this constant can be found in the literature, such as [2, 3, 23, 24, 25], (see also [26, 27]). A special case is Sine-Gordon in 1-d where the remainder J vanishes as $N \to \infty$, so the constant vanishes in this case.

What we have obtained in (6.24) is the leading ϵ scaling: it is from $\bar{n} = 2$ for asymmetric potentials and from $\bar{n} = 3$ for symmetric potentials. The form of the radiation in eq. (6.24) agrees with the scaling found by Segur and Kruskal [2] in $d = 1$ and generalized to other d. Our methodology of computing a spatial Fourier transform and evaluating it a wavenumber determined by kinematics is general and should apply to various other oscillon models, such as multi-field.

6.4 Quantization

The preceding discussion explains why a classical oscillon can live for an exceptionally long time. This does, however, require the oscillon to be placed in the right initial conditions for this to occur. Some investigation has gone into the stability/instability of classical oscillons under arbitrary initial conditions, e.g., [28, 29]. Here our focus is on a sharply defined question, free from the ambiguity of initial conditions: *If a single*

[1] The constant c_d is in fact the first simple pole of $\tilde{f}(\rho)$ along the imaginary ρ-axis; see Ref. [3] for comparison.

oscillon, as defined by the ϵ expansion, is present – what is its lifetime? There is a sharp answer in the classical theory – exponential, and now we address the question in the quantum theory.

A semi-classical (leading $\hbar$) description involves treating the oscillon $\phi_{\rm osc} = \phi_{\rm osc}(t, r)$ as a classical background and quantizing fields in this background. In this section, only the field ϕ itself is present to be quantized, but in Section 6.6 we will introduce a second field χ to quantize. Let's write

$$\phi(\mathbf{x}, t) = \phi_{\rm osc}(x, t) + \hat{\phi}(\mathbf{x}, t), \tag{6.25}$$

where $\hat{\phi}$ is a quantum field satisfying canonical commutation relations. At any finite order in the ϵ expansion, $\phi_{\rm osc}$ is an exact periodic solution of the equations of motion, as discussed in Section 6.2. Perturbing around a solution allows us to write down the following equation of motion for $\hat{\phi}$ in the Heisenberg picture

$$\ddot{\hat{\phi}} - \nabla^2\hat{\phi} + \hat{\phi} + \Phi(\phi_{\rm osc})\hat{\phi} = 0, \tag{6.26}$$

where $\Phi(\phi_{\rm osc}) \equiv V_I''(\phi_{\rm osc})$. We have neglected higher order terms in $\hat{\phi}$, since we are only interested in a leading order $\hbar$ analysis. This is the theory of a free quantum scalar field with a space-time dependent mass. The ground state of this theory is given by an infinite sum of 1-loop vacuum diagrams, coming from $N_\Phi = 0, 1, 2, \ldots$ insertions of the external field Φ. The $N_\Phi = 0$ diagram corresponds to the ordinary zero point energy of a free field, while the $N_\Phi \geq 1$ diagrams correspond to production of $\hat{\phi}$ quanta from the background source.

For small ϵ, the oscillon is wide, with k-modes concentrated around $k = \mathcal{O}(\epsilon) \ll 1$, which suggests it is more convenient to perform the analysis in k-space. So let's take the Fourier transform:

$$\ddot{\hat{\phi}}_k + \omega_k^2\,\hat{\phi}_k + \int \frac{d^dk'}{(2\pi)^d}\Phi(\mathbf{k}-\mathbf{k}')\hat{\phi}_{k'} = 0, \tag{6.27}$$

where $\omega_k^2 \equiv k^2+1$. Here we used the convolution theorem on the final term. Now if the background was homogeneous, then each $\hat{\phi}_k$ would decouple. This makes the solution rather straightforward, as is the situation during cosmological inflation, for instance. In that case each $\hat{\phi}_k$ is proportional to a single time independent annihilation operator $\hat{a}_k$ times a mode function $v_k(t)$ that satisfies the classical equation of motion, plus hermitian conjugate. But due to the inhomogeneity in $\phi_{\rm osc}$, the k-modes are coupled, and the solution in the background of an oscillon is non-trivial.

Nevertheless a formal solution of the Heisenberg equations of motion can be obtained. The key is to integrate over all annihilation operators:

$$\hat{\phi}_k(t) = \sqrt{\hbar}\int \frac{d^dq}{(2\pi)^d}\,\hat{a}_q\,v_{qk}(t) + h.c. \tag{6.28}$$

Upon substitution, each mode function $v_{qk}(t)$ must satisfy the classical equation of

motion

$$\ddot{v}_{qk} + \omega_k^2 v_{qk} + \int \frac{d^d k'}{(2\pi)^d} \Phi(k-k') v_{qk'} = 0. \tag{6.29}$$

We see that we have a matrix of time dependent mode functions $v_{qk}(t)$ to solve for. We choose initial conditions such that $\hat{\phi}$ is initially in its unperturbed vacuum state. This requires the following initial values of the mode functions:

$$v_{qk}(0) = \frac{1}{\sqrt{2\omega_k}} (2\pi)^d \delta^d(\mathbf{q}-\mathbf{k}), \tag{6.30}$$

$$\dot{v}_{qk}(0) = -i\,\omega_k\, v_{qk}(0). \tag{6.31}$$

The local energy density u and total energy E in $\hat{\phi}$ at time t can be defined by the unperturbed Hamiltonian, giving

$$u(x,t) = \frac{\hbar}{2} \int \frac{d^d q}{(2\pi)^d} \frac{d^d k}{(2\pi)^d} \frac{d^d k'}{(2\pi)^d} e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{x}} \left[\dot{v}_{qk} \dot{v}^*_{qk'} + \omega^2_{kk'} v_{qk} v^*_{qk'} \right], \tag{6.32}$$

$$E(t) = \frac{\hbar}{2} \int \frac{d^d q}{(2\pi)^d} \frac{d^d k}{(2\pi)^d} \left[|\dot{v}_{qk}|^2 + \omega_k^2 |v_{qk}|^2 \right] \tag{6.33}$$

($\omega^2_{kk'} \equiv \mathbf{k}\cdot\mathbf{k}' + 1$), where the second equation is easily obtained from the first by integrating over $\mathbf{x}$. Initially the total energy is $E(0) = \int d^d k \frac{1}{2}\hbar\omega_k\, \delta^d(0)$; the usual (infinite) zero point energy of the vacuum. The energy corresponding to $\hat{\phi}$ production is contained in the time evolution of $E(t)$. As we will explain, radiation comes from specific wavenumbers that are $\mathcal{O}(1)$, and are connected to tree-level processes of the underlying microphysical theory. This makes it straightforward to extract the correct finite result for the produced radiation energy, despite the zero point energy being UV divergent.

To solve the system numerically we operate in a box of volume $V = L^d$ and discretize the system as follows

$$\int \frac{d^d k'}{(2\pi)^d} \rightarrow \frac{1}{V} \sum_{\mathbf{k}} \quad \left(\mathbf{k} = \frac{2\pi \mathbf{n}}{L},\, n_i \in \mathbb{Z} \right), \tag{6.34}$$

$$\delta^d(\mathbf{q}-\mathbf{k}) \rightarrow \frac{V}{(2\pi)^d} \delta_{qk}. \tag{6.35}$$

The discretized equations of motion represent an infinite set of coupled oscillators. They have a periodic mass as driven by the background oscillon and as such are amenable to a generalized Floquet analysis of coupled oscillators. We use the Floquet theory to solve for the late time behavior, as discussed in Section 6.7.1.

It is interesting to compare this to a classical stability analysis. Here one should explore all initial conditions that span the complete space of perturbations. To do this, we can write $\delta\phi_k \rightarrow v_{qk}$ where q is merely an index that specifies the choice of initial condition. To span all initial conditions, choose $v_{qk}(0) \propto \delta(\mathbf{q}-\mathbf{k})$, as in the

standard Floquet theory. But this is precisely what we have done to solve the quantum problem. Hence a classical stability analysis over all initial conditions involves the same computation as solving the quantum problem for a fixed initial condition – the ground state.

6.5 Quantum Radiation

Perhaps the most interesting oscillons are those that are stable against instabilities that would appear in a classical simulation; such instabilities typically arise from $k = \mathcal{O}(\epsilon)$ modes of the oscillon, and will be discussed in Section 6.8. Classically, these oscillons appear to be *extremely* stable. In this Section we calculate and explain why the oscillon lifetime is shortened in the quantum theory due to $k = \mathcal{O}(1)$ modes, depending on the size of the effective $\hbar$.

In order to make progress, we will solve the mode functions perturbatively. Although our methodology is general, we will demonstrate this with a model that makes the computation the easiest. Lets consider the potential

$$V_I(\phi) = -\frac{\lambda}{4!}\phi^4 + \frac{\lambda_5}{5!}\phi^5 + \dots \tag{6.36}$$

The $-\lambda\,\phi^4$ term ensures that classical oscillons exist in the form given earlier (eq. (6.10)). The $\lambda_5\,\phi^5$ will be quite important in the quantum decay, but it does not effect the leading order term in the oscillon expansion. It can be verified that for this potential the oscillon has the form

$$\phi_{\rm osc}(x,t) = \phi_\epsilon(x)\cos(\omega t) + \mathcal{O}(\epsilon^3) \tag{6.37}$$

where $\phi_\epsilon(x) \equiv 4\,\epsilon\,\tilde{f}(x\epsilon)/\sqrt{\lambda}$ (note $\phi_\epsilon \sim \epsilon$). λ_5 only enters the expansion at $\mathcal{O}(\epsilon^4)$ (generating even harmonics).

Lets perform an expansion in powers of λ_5, i.e.,

$$v_q = v_q^{(0)} + \lambda_5\,v_q^{(1)} + \lambda_5^2\,v_q^{(2)} + \dots \tag{6.38}$$

with each term $v_q^{(i)}$ implicitly an expansion in powers of ϵ, with ϵ assumed small. We have suppressed the second index on v_q, which would be k in momentum space or x in position space, since both representations will be informative. In position space at zeroth order in λ_5, we have

$$\ddot{v}_{qx}^{(0)} - \nabla^2 v_{qx}^{(0)} + v_{qx}^{(0)} = \left[\frac{\lambda}{2}\phi_\epsilon^2(x)\cos^2(\omega t) + \mathcal{O}(\epsilon^4)\right] v_{qx}^{(0)} \tag{6.39}$$

The first term on the RHS is sometimes responsible for collapse instabilities, as we shall discuss in Section 6.8, but only for wavenumbers $k = \mathcal{O}(\epsilon)$. The existence of such instabilities are highly model dependent and are not the focus of this section. Instead here we focus on $k = \mathcal{O}(1)$ which is relevant for outgoing radiation. The second term on the RHS, which is $\mathcal{O}(\epsilon^4)$, is indeed relevant to the production of radiation, but it

will be superseded by radiation at $\mathcal{O}(\epsilon^3)$ that will enter when we examine $v_q^{(1)}$. For now, we only need to conclude from eq. (6.39) that $v_q^{(0)}$ is equal to the solution of the free-theory plus $\mathcal{O}(\epsilon^2)$ corrections. In k-space this means we take the unperturbed mode functions

$$v_{qk}^{(0)}(t) = \frac{e^{-i\omega_k t}}{\sqrt{2\omega_k}}(2\pi)^d \delta^d(\mathbf{q}-\mathbf{k}) + \mathcal{O}(\epsilon^2), \tag{6.40}$$

which match the initial conditions mentioned earlier in eqs. (6.30, 6.31). Note that an integral over q recovers the standard mode functions that occur in a free theory.

At next order in λ_5 we have

$$\ddot{v}_{qx}^{(1)} - \nabla^2 v_{qx}^{(1)} + v_{qx}^{(1)} = -\frac{1}{3!}\phi_\epsilon^3(x)\cos^3(\omega t) v_{qx}^{(0)} \tag{6.41}$$

Fourier transforming to k-space and inserting $v_{qk}^{(0)}$ gives

$$\ddot{v}_{qk}^{(1)} + \omega_k^2 v_{qk}^{(1)} = -\frac{\psi_\epsilon(\mathbf{q}-\mathbf{k})}{3!2^3\sqrt{2\omega_k}}\left[e^{i(\tilde{\omega}-\omega_q)t} + e^{-i(\tilde{\omega}+\omega_q)t}\right] \tag{6.42}$$

where $\psi_\epsilon(k)$ is the Fourier transform of $\phi_\epsilon^3(x)$ and $\tilde{\omega} \equiv 3\omega$. Here we have ignored terms on the RHS with frequency $\omega \pm \omega_q$ which cannot generate a resonance. The first term included does generate a resonance at $\omega_q \approx \omega_k \approx \tilde{\omega}/2$. This is the equation of a forced oscillator whose solution can be readily obtained. By imposing the initial conditions $v_{qk}^{(1)}(0) = \dot{v}_{qk}^{(1)}(0) = 0$ we find

$$v_{qk}^{(1)}(t) = -\frac{\psi_\epsilon(\mathbf{q}-\mathbf{k})}{3!2^3\sqrt{2\omega_k}} S(\tilde{\omega}, \omega_q, \omega_k) \tag{6.43}$$

where

$$\begin{aligned} S(\tilde{\omega}, \omega_q, \omega_k) \equiv & \frac{e^{-i(\tilde{\omega}+\omega_q)t}}{-(\omega_q+\tilde{\omega})^2+\omega_k^2} + \frac{e^{i(\tilde{\omega}-\omega_q)t}}{-(\omega_q-\tilde{\omega})^2+\omega_k^2} \\ & - \frac{\left(1-\frac{\omega_q}{\omega_k}\right)e^{-i\omega_k t}}{-\tilde{\omega}^2+(\omega_k-\omega_q)^2} - \frac{\left(1+\frac{\omega_q}{\omega_k}\right)e^{i\omega_k t}}{-\tilde{\omega}^2+(\omega_k+\omega_q)^2} \end{aligned} \tag{6.44}$$

Lets now insert our solution for v_{qk} into the expression for the energy density $u(x,t)$ (eq. (6.32)). This includes terms scaling as $v_{qk}^{(0)} v_{qk'}^{(0)}$ which is the usual zero point energy. Another term scales as $\lambda_5\, v_{qk}^{(0)} v_{qk'}^{(1)}$ which is non-resonant. Then there are two important terms: $\lambda_5^2\, v_{qk}^{(1)} v_{qk'}^{(1)}$ and $\lambda_5^2\, v_{qk}^{(0)} v_{qk'}^{(2)}$. It can be shown that the second term here provides the same contribution as the first term. (In the case of a homogeneous pump field the mode functions are often written in terms of Bogoliubov coefficients, which make this fact manifest. A similar argument goes through in our more complicated inhomogeneous case.) Given this, we will not write out the explicit solution for $v_{qk}^{(2)}$ here. Finally, we shall use the fact that only terms near resonance contribute significantly to the integral, and further we shall make simplifications using $\omega_q \approx$

$\omega_k \approx \tilde{\omega}/2$ whenever we can; however, this must not be done in singular denominators or in the arguments of oscillating functions.

Altogether at leading order in λ_5 and ϵ we find the following expression for the change in the energy density from the zero point:

$$\delta u(x,t) = \frac{\lambda_5^2 \hbar}{(3!)^2 2^7 \tilde{\omega}} \int \frac{d^d q}{(2\pi)^d} \frac{d^d k}{(2\pi)^d} \frac{d^d k'}{(2\pi)^d} \cos((\mathbf{k}' - \mathbf{k}) \cdot \mathbf{x}) \, \psi_\epsilon(\mathbf{q} - \mathbf{k}) \psi_\epsilon^*(\mathbf{q} - \mathbf{k}') \times \left[\frac{1 + \cos(t(\omega_{k'} - \omega_k)) - \cos(t(\omega_k + \omega_q - \tilde{\omega})) - \cos(t(\omega_{k'} + \omega_q - \tilde{\omega}))}{(\omega_k + \omega_q - \tilde{\omega})(\omega_{k'} + \omega_q - \tilde{\omega})} \right] \quad (6.45)$$

For the 1-dimensional case we have evaluated this numerically; the results are plotted in Fig. (6-2). As the figure shows the energy density initially grows in the core of the oscillon before reaching a maximum (around $\delta u \sim 10^{-7}$ in the figure for the parameters chosen). This occurs at each point in space and so the net effect is of continual growth in δu by spreading out. This has a simple interpretation: ϕ-particles are being produced from the core of the oscillon and moving outwards. We will show using scattering theory that this is connected to the annihilation process $3\,\phi \longrightarrow 2\,\phi$.

Let us turn now to the total energy output (after subtracting the zero point) $\delta E(t)$. This comes from integrating eq. (6.45) over $\mathbf{x}$, giving

$$\delta E(t) = \frac{\lambda_5^2 \hbar}{(3!)^2 2^6 \tilde{\omega}} \int \frac{d^d q}{(2\pi)^d} \frac{d^d k}{(2\pi)^d} |\psi_\epsilon(\mathbf{q} - \mathbf{k})|^2 \left[\frac{1 - \cos(t(\omega_k + \omega_q - \tilde{\omega}))}{(\omega_k + \omega_q - \tilde{\omega})^2} \right]. \quad (6.46)$$

This may be further simplified by recognizing that $|\psi_\epsilon(\mathbf{q} - \mathbf{k})|^2$ is non-negative and sharply spiked around $\mathbf{q} - \mathbf{k} \approx 0$ for small ϵ (as we discussed in Section 6.3) and so it acts like a δ-function, i.e.,

$$|\psi_\epsilon(\mathbf{q} - \mathbf{k})|^2 \approx \left[\int \frac{d^d k}{(2\pi)^d} |\psi_\epsilon(\mathbf{k})|^2 \right] (2\pi)^d \delta^d(\mathbf{q} - \mathbf{k}), \quad (6.47)$$

where we have intoduced the integral prefactor to ensure that the integration of $|\psi_\epsilon(\mathbf{q} - \mathbf{k})|^2$ over all $\mathbf{k}$ gives the correct value when we replace it by the δ-function. This quantity has a nice interpretation when re-written as an integral over position space. Since $\psi_\epsilon(k)$ is the Fourier transform of $\phi_\epsilon^3(x)$, we can write

$$\int \frac{d^d k}{(2\pi)^d} |\psi_\epsilon(k)|^2 \approx 2^3 \int d^d x \, u_\epsilon^3(x), \quad (6.48)$$

where $u_\epsilon(x) \approx \phi_\epsilon^2(x)/2$ is the oscillon's local energy density. Inserting this into (6.46) gives

$$\delta E(t) = \frac{\lambda_5^2 \hbar \int d^d x \, u_\epsilon^3(x)}{(3!)^2 2^3 \tilde{\omega}} \int \frac{d^d k}{(2\pi)^d} \frac{1 - \cos(t(2\omega_k - \tilde{\omega}))}{(2\omega_k - \tilde{\omega})^2} \quad (6.49)$$

This integral can be done explicitly for times $t \gg 1$. To do this we write $d^d k = d\Omega \, dk \, k^{d-1}$. The angular integration is trivial $\int d\Omega = 2\,\pi^{d/2}/\Gamma(d/2)$. Now write $dk =$

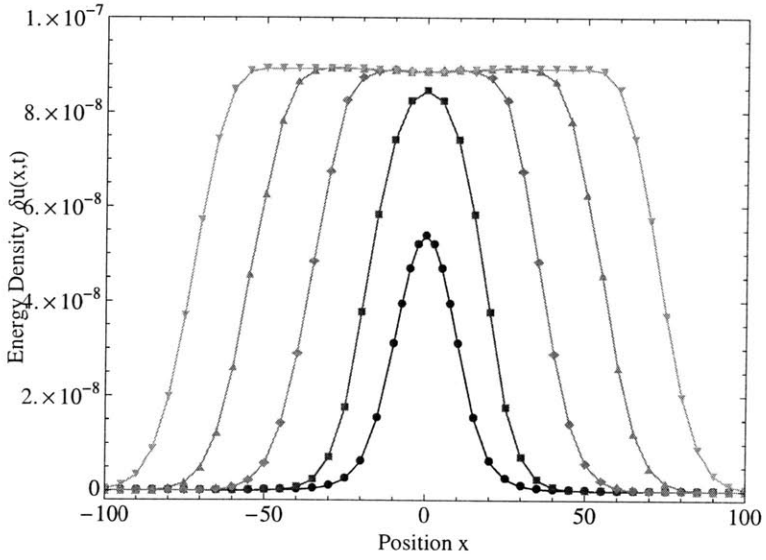

Figure 6-2: Energy density $\delta u(x,t)$ (in units of $\hbar$). Each curve is at a different time interval: blue is $t = 10$, red is $t = 25$, green is $t = 50$, orange is $t = 75$, and cyan is $t = 100$. Here $d = 1$, $\epsilon = 0.05$, and $\lambda_5^2/\lambda^3 = 1$.

$d\omega_k\omega_k/k$. The integral over $d\omega_k$ is in the domain $m = 1 < \omega_k < \infty$, but for late times t we can extend it to the whole domain $-\infty < \omega_k < \infty$, since only the region $\omega_k \approx \tilde{\omega}/2 = 3\,\omega/2 \approx 3m/2 = 3/2$ contributes significantly. In the integrand, we also replace k by its resonant value corresponding to radiation $k_{\rm rad}$ which satisfies $\omega_k = \sqrt{k_{\rm rad}^2 + m^2} = 3\,\omega/2$. The remaining integral to perform is

$$\int_{-\infty}^{\infty} d\omega_k \frac{1-\cos(t(2\omega_k - \tilde{\omega}))}{(2\omega_k - \tilde{\omega})^2} = \frac{\pi\, t}{2}. \tag{6.50}$$

Hence the energy output

$$\delta E(t) = \frac{\lambda_5^2 \hbar}{(3!)^2 2^4} \frac{\pi^{d/2+1} k_{\rm rad}^{d-2}}{\Gamma(\frac{d}{2})(2\pi)^d} \int d^d x\, u_\epsilon^3(x)\, t \tag{6.51}$$

increases linearly with time. So the oscillon must lose energy at this rate. The decay rate is $\Gamma_d = |E_{\rm osc}^{-1}\, dE_{\rm osc}/dt|$ and using $E_{\rm osc} = \int d^d x\, u_\epsilon(x)$ we obtain

$$\Gamma_d(3\,\phi \to 2\,\phi) = \frac{\lambda_5^2 \hbar}{(3!)^2 2^4} \frac{\pi^{d/2+1} k_{\rm rad}^{d-2}}{\Gamma(\frac{d}{2})(2\pi)^d} \frac{\int d^d x\, u_\epsilon^3(x)}{\int d^d x\, u_\epsilon(x)} \tag{6.52}$$

as our final result for the decay rate. We have labelled this "$3\,\phi \to 2\,\phi$" annihilation for a reason we now explain.

Its useful to connect this result to ordinary perturbation theory for a gas of incoherent particles. The differential transition rate for a single annihilation of $N_i\ \phi \to 2\,\phi$ in a box of volume $V_{\rm box}$, evaluated on threshold for non-relativistic initial particles of mass m, is given by (e.g., see [30])

$$d\Gamma_1 = \frac{V_{\rm box}^{1-N_i}}{N_i!(2m)^{N_i}} |\mathcal{M}|^2 (2\pi)^D \delta^D(p_i - p_f) \prod_f \frac{d^d p_f}{(2\pi)^d} \frac{1}{2E_f} \tag{6.53}$$

($D = d+1$). To connect to the previous calculation we choose $N_i = 3$ since the relevant interaction is provided by $3\,\phi \to 2\,\phi$ annihilation at tree-level due to the interaction term $\Delta V_I = \frac{1}{5!}\lambda_5\,\phi^5$. This gives $|\mathcal{M}|^2 = \lambda_5^2$. Performing the integration over phase space (including division by 2 since the final 2 particles are indistinguishable) gives the result

$$\Gamma_1(3\,\phi \to 2\,\phi) = \frac{\lambda_5^2 \hbar}{3! 2^4} \frac{\pi^{d/2+1} k_{\rm rad}^{d-2}}{\Gamma(\frac{d}{2})(2\pi)^d} \frac{V_{\rm box}^{-2}}{2E_f} \tag{6.54}$$

The inverse of this is the time taken for a given triplet of ϕ's to annihilate. For a box of N_ϕ particles the total rate for an annihilation to occur is Γ_1 multiplied by the number of indistinguishable ways we can choose a triplet, i.e.,

$$\Gamma_{\rm tot}(3\,\phi \to 2\,\phi) = \binom{N_\phi}{3} \Gamma_1(3\,\phi \to 2\,\phi) \tag{6.55}$$

$$\approx \frac{N_\phi^3}{3!} \Gamma_1(3\,\phi \to 2\,\phi) \tag{6.56}$$

for $N_\phi \gg 1$ (the semi-classical regime). Since each process produces a pair of particles

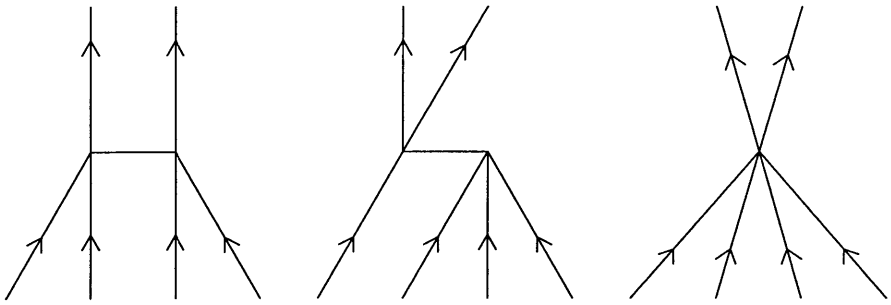

Figure 6-3: Feynman diagrams for the process $4\,\phi \to 2\,\phi$. Evaluated on threshold, the first diagram is $i\lambda_4^2/4$ ($\times$ 6 crossing symmetries), the second diagram $-i\lambda_4^2/8$ ($\times$ 4 crossing symmetries), and the third diagram is $-i\lambda_6$.

of total energy $2E_f$, the total energy output is

$$\delta E(t) = 2E_f \Gamma_{\rm tot}(3\,\phi \to 2\,\phi)\, t. \tag{6.57}$$

Together with (6.54) this recovers the result in eq. (6.51) precisely, as long as we make the replacement

$$V_{\rm box}^{-2} N_{\rm box}^3 \to \int d^d x\, u_\epsilon^3(x). \tag{6.58}$$

Hence to leading order in λ_5 and ϵ we find that decay rates computed using tree-level scattering theory recover the result computed from solving the mode functions for a coherent background oscillon. The important connection is provided by eq. (6.58) to account for the spatial structure of the oscillon.

Given this connection, the generalization of our result to arbitrary interactions is relatively straightforward. Consider the more general interaction potential

$$V_I(\phi) = \frac{\lambda_3}{3!}\phi^3 + \frac{\lambda_4}{4!}\phi^4 + \frac{\lambda_5}{5!}\phi^5 + \frac{\lambda_6}{6!}\phi^6 + \ldots \tag{6.59}$$

As mentioned earlier, a requirement for the existence of small amplitude oscillons is $\frac{5}{3}\lambda_3^2 - \lambda_4 > 0$; this in fact is the *only* requirement on the couplings [25]. If either λ_3 or λ_5 are non-zero, then the leading order radiation arises from $3\,\phi \to 2\,\phi$ annihilation, as we computed previously for the $\lambda_3 = 0$ case. This again gives the result of eq. (6.52), with the generalization of the λ_3^2 prefactor being replaced by the square of the matrix element for all such tree-level scattering diagrams, i.e.,

$$\lambda_5^2 \to |\mathcal{M}(3\,\phi \to 2\,\phi)|^2 \tag{6.60}$$

This includes the important case in $d = 3$ with $\lambda_3 > 0$, $\lambda_4 > 0$ and $\lambda_5 = \lambda_6 = \ldots = 0$, which is a renormalizable field theory. For brevity, we have not drawn the many relevant diagrams here, but in Fig. 6-3 we draw the diagrams that are relevant to the next example.

Now consider a purely symmetric potential with $\lambda_3 = \lambda_5 = 0$ and $\lambda_4 < 0$ and

$\lambda_6 > 0$. In this case, the leading order annihilation process is $4\,\phi \to 2\,\phi$, so we must choose $N_i = 4$ in eq. (6.53). By summing the diagrams of Fig. 6-3 we find $|\mathcal{M}(4\,\phi \to 2\,\phi)|^2 = (\lambda_4^2 - \lambda_6)^2$. Altogether we obtain the following result

$$\Gamma_d(4\,\phi \to 2\,\phi) = \frac{(\lambda_4^2 - \lambda_6)^2 \hbar}{(4!)^2 2^5} \frac{\pi^{d/2+1} k_{\rm rad}^{d-2}}{\Gamma(\frac{d}{2})(2\pi)^d} \frac{\int d^d x\, u_\epsilon^4(x)}{\int d^d x\, u_\epsilon(x)} \tag{6.61}$$

In this case the kinematical requirement on $k_{\rm rad}$ for resonance is $\omega_k = \sqrt{k_{\rm rad}^2 + m^2} = 2\,\omega$. Note this rate vanishes if and only if $\lambda_6 = \lambda_4^2$, which is true for the Sine-Gordon potential $V_{\rm SG}(\phi) = (1 - \cos(\sqrt{\lambda}\,\phi))/\lambda$. In fact it can be proven that in (1+1)-dimensions every ϕ number changing process vanishes on threshold for the Sine-Gordon potential.

Lets discuss the scaling of our results (6.52) and (6.61). Recall that $\phi_{\rm osc} \sim \epsilon/\sqrt{\lambda}$, so $u_\epsilon \sim \epsilon^2/\lambda$. Thus (depending on the relative size of couplings) the decay rates scale as

$$\Gamma_d(3\,\phi \to 2\,\phi) \sim \epsilon^4 \lambda_5^2 \hbar/\lambda^2, \quad \text{or} \quad \epsilon^4 \lambda_3^6 \hbar/\lambda^2 \tag{6.62}$$

$$\Gamma_d(4\,\phi \to 2\,\phi) \sim \epsilon^6 \lambda_4 \hbar, \quad \text{or} \quad \epsilon^6 \lambda_6^2 \hbar/\lambda^3. \tag{6.63}$$

(We also find that in the large $+\lambda_6\,\phi^6$ model of Ref. [19] the scaling is $\Gamma_d \sim \lambda_4^3 \hbar/\lambda_6$.) Hence the quantum mechanical decay rate is a power law in the parameters ϵ, λ_i. Such quantum decay rates will only be smaller than the corresponding classical decay rate (6.24) if the "effective $\hbar$" (such as $\lambda\,\hbar$) is *extremely* small, as we are comparing it to an exponentially small quantity in $\exp(-b/\epsilon)$. Unlike collapse instabilities that we will mention in Section 6.8, it is almost impossible to avoid this radiation by changing dimensionality, parameter space, or field theory (the only known exception is the Sine-Gordon model in $d = 1$).

One may wonder why the classical analysis failed to describe this decay rate accurately given that the oscillon can be full of many quanta, say $N_\phi \approx M_{\rm osc}/m_\phi$, with an almost continuous spectrum. Let us explain the resolution. For $N_\phi \gg 1$ the quantum corrections to the oscillon's bulk properties are small. For example, the quantum correction to the oscillon's width, amplitude, total mass, etc, should be small, since the classical values are large. In k-space we can say that these properties are governed by $k = \mathcal{O}(\epsilon)$ modes, which carry a large amplitude. However, the radiation is quite different; it is governed by $k = \mathcal{O}(1)$ modes, which are exponentially small in the classical oscillon, but are not exponentially small in the quantum oscillon due to zero point fluctuations. Another way to phrase this is to consider the commutation relation

$$\hat{\phi}(\mathbf{k})\hat{\pi}(\mathbf{k}') = \hat{\pi}(\mathbf{k}')\hat{\phi}(\mathbf{k}) + i\hbar\,(2\pi)^d \delta^d(\mathbf{k} - \mathbf{k}'). \tag{6.64}$$

For $k = \mathcal{O}(\epsilon)$ the classical value of the LHS and its counterpart on the RHS are both large, so the $\hbar$ correction is negligible. But for $k = \mathcal{O}(1)$ the classical values are small, so the quantum corrections are important.

An interesting issue is the behavior of the growth at late times. In the case of a homogeneous background pump field (for example, as is relevant during p/reheating

at the end of inflation, e.g., see [31, 32, 33]) it is known that the linear growth is really the initial phase of exponential growth. So is the same true for spatially localized oscillons? Here we claim that for sufficiently small amplitude oscillons, the answer is *no*, the linear growth rate is correct even at late times, although this can change for sufficiently large amplitude oscillons where our perturbative analysis breaks down. We have confirmed the linear growth for small amplitude oscillons in 2 ways: (i) by expanding v_{qk} to higher order in the coupling and (ii) by solving the full mode function equations (6.29) numerically. The reason for this result is subtle and will be discussed in detail in Section 6.7; we will demonstrate that whether the growth is in the linear regime or exponential regime depends critically on the oscillon's amplitude, width, and couplings.

6.6 Coupling to Other Fields

Most fields in nature interact considerably with others. It is important to know what is the fate of a ϕ-oscillon that is coupled to other fields. Let's couple ϕ to another scalar χ and consider the following Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\,\phi^2 - V_I(\phi) + \frac{1}{2}(\partial\chi)^2 - \frac{1}{2}m_\chi^2\,\chi^2 - \frac{1}{2}g_1\,m\,\phi\,\chi^2 - \frac{1}{2}\,g_2\,\phi^2\chi^2 \quad (6.65)$$

The last two terms represent interactions between the 2 fields, with g_i coupling parameters. The interaction term $g_1\,\phi\,\chi^2$ allows the following tree-level decay process to occur in vacuo: $\phi \to \chi + \chi$, if the following mass condition is met: $m > 2\,m_\chi$. Assuming this condition is met, we are led to ask: Will the ϕ-oscillon decay into χ? If so, will the growth in χ be linear or exponential? Otherwise, if $m < 2\,m_\chi$, or if $g_1 = 0$, $g_2 \neq 0$, we can focus on annihilations: $\phi + \phi \to \chi + \chi$, etc.

One could approach the issue of multiple fields by evolving the full ϕ, χ system under the classical equations of motion. In fact by scanning the mass ratio of ϕ and χ one can find interesting oscillons involving an interplay of both fields – one finds that the 2:1 mass ratio is of particular importance (as was the case for the SU(2) oscillon [4]). Although this is an interesting topic, here we would like to focus on the effects on the ϕ-oscillon due to the introduction of χ, initially in its vacuum state. At the classical level χ will remain zero forever, so this is trivial. We will return to the issue of the classical evolution for non-trivial initial conditions for χ in the next Section. For now we focus on placing χ in its quantum vacuum state with a classical background $\phi_{\rm osc}$.

We use the same formalism as we developed in Section 6.4. We make the replacements $\hat{\phi} \to \hat{\chi}$, $m = 1 \to m_\chi$ in eq. (6.27), giving the following Heisenberg equation of motion for $\hat{\chi}$ in k-space

$$\ddot{\hat{\chi}}_k + \omega_k^2\,\hat{\chi}_k + \int \frac{d^dk'}{(2\pi)^d}\Phi(\mathbf{k}-\mathbf{k}')\hat{\chi}_{k'} = 0, \quad (6.66)$$

where $\omega_k^2 \equiv k^2 + m_\chi^2$ and $\Phi \equiv g_1\,\phi_{\rm osc}(\mathbf{x},t) + g_2\,\phi_{\rm osc}^2(\mathbf{x},t)$. We write $\hat{\chi}$ in terms of its

mode functions v_{qk} as before

$$\hat{\chi}_k(t) = \sqrt{\hbar} \int \frac{d^d q}{(2\pi)^d} \hat{a}_q \, v_{qk}(t) + h.c. \tag{6.67}$$

For brevity, lets focus on leading order in ϵ behavior, coming from $\Phi \approx g_1 \, \phi_\epsilon(x) \cos(\omega t)$ (although later we will also mention the important case of $g_1 = 0$ with $\Phi = g_2 \, \phi_\epsilon^2(x) \cos^2(\omega t)$). This gives the following mode function equations

$$\ddot{v}_{qk} + \omega_k^2 v_{qk} + g_1 \cos(\omega t) \int \frac{d^d k'}{(2\pi)^d} \, \phi_\epsilon(\mathbf{k} - \mathbf{k}') \, v_{qk'} = 0. \tag{6.68}$$

For small coupling g_1 we expect the solutions of the mode equations to be small deformation of planes waves. To capture this, lets expand the mode functions in powers of g_1 (analogously to the earlier expansion in Section 6.5)

$$v_{qk} = v_{qk}^{(0)} + g_1 \, v_{qk}^{(1)} + g_1^2 v_{qk}^{(2)} + \ldots \tag{6.69}$$

At leading order $\mathcal{O}(g_1^0)$, we have $\ddot{v}_{qk}^{(0)} + \omega_k^2 v_{qk}^{(0)} = 0$, whose desired solution is the unperturbed mode functions

$$v_{qk}^{(0)}(t) = \frac{e^{-i\omega_k t}}{\sqrt{2\omega_k}} (2\pi)^d \delta^d(\mathbf{q} - \mathbf{k}), \tag{6.70}$$

as earlier. At next order we have the following forced oscillator equation

$$\ddot{v}_{qk}^{(1)} + \omega_k^2 v_{qk}^{(1)} = -\frac{\phi_\epsilon(\mathbf{q} - \mathbf{k})}{2\sqrt{2\omega_k}} \left[e^{-i(\omega + \omega_q)t} + e^{i(\omega - \omega_q)t} \right], \tag{6.71}$$

The solution with boundary conditions $v_{qk}^{(1)}(0) = \dot{v}_{qk}^{(1)}(0) = 0$ is

$$v_{qk}^{(1)}(t) = -\frac{\phi_\epsilon(\mathbf{q} - \mathbf{k})}{\sqrt{2\omega_k}} S(\omega, \omega_q, \omega_k) \tag{6.72}$$

where S was defined in eq. (6.44). Hence we obtain the same expressions for the energy density δu and δE as in eqs. (6.45, 6.46) upon making the replacements

$$\frac{\lambda_5}{3! 2^3} \psi_\epsilon(\mathbf{q} - \mathbf{k}) \to \frac{g_1}{2} \phi_\epsilon(\mathbf{q} - \mathbf{k}), \qquad \tilde{\omega} \to \omega \tag{6.73}$$

Evaluating this we find results that are qualitatively similar to before: the χ field is produced in the core of the oscillon and spreads out, and the energy grows linearly in time.

Following through a similar calculation to before in the small ϵ limit, by identifying

$|\phi_\epsilon(\mathbf{q}-\mathbf{k})|^2$ as proportional to a δ-function, i.e.,

$$|\phi_\epsilon(\mathbf{q}-\mathbf{k})|^2 \approx 2\left[\int d^dx\, u_\epsilon(x)\right](2\pi)^d\delta^d(\mathbf{q}-\mathbf{k}), \tag{6.74}$$

allows us to evaluate the decay rate at leading order in g_1, which is connected to the decay process $\phi \to \chi+\chi$. We also carry though the calculation at leading order in g_2 (including next order in g_1), which is connected to the annihilation process $\phi+\phi \to \chi+\chi$. We find the results

$$\Gamma_d(1\,\phi \to 2\,\chi) = \frac{g_1^2\hbar}{2^2}\frac{\pi^{d/2+1}k_{\rm rad}^{d-2}}{\Gamma(\frac{d}{2})(2\pi)^d}, \tag{6.75}$$

$$\Gamma_d(2\,\phi \to 2\,\chi) = \frac{(g_1^2-g_2)^2\hbar}{2^3}\frac{\pi^{d/2+1}k_{\rm rad}^{d-2}}{\Gamma(\frac{d}{2})(2\pi)^d}\frac{\int d^dx\, u_\epsilon^2(x)}{\int d^dx\, u_\epsilon(x)} \tag{6.76}$$

In (6.75) the kinematical requirement on $k_{\rm rad}$ is $\omega_k = k_{\rm rad}^2 + m_\chi^2 = (\omega/2)^2$ and in (6.76) the requirement is $\omega_k^2 = k_{\rm rad}^2 + m_\chi^2 = \omega^2$. So these decays only occur for sufficiently light χ. Here $\Gamma_d(1\,\phi \to 2\,\chi)$ coincides with the perturbative decay rate of ϕ and $\Gamma_d(2\,\phi \to 2\,\chi)$ is a generalization of the annihilation rate $\Gamma = n\,\langle\sigma v\rangle$ applied to a gas of non-relativistic particles with variable density. As an application, we expect this result to be relevant to the bosonic SU(2) oscillon [4]. Since it exists at the 2:1 mass ratio ($m_{\rm H} = 2m_{\rm W}$), it prevents Higgs decaying into W-bosons, but it should allow the quantum mechanical annihilation of Higgs into relativistic W-bosons.

6.7 Exponential vs Linear Growth

So far we have worked to leading order in the couplings and ϵ, this has resulted in a constant decay rate of the oscillon into quanta of ϕ in the single field case or quanta of χ in the two field case. These quanta have an energy that grows linearly in time, see eq. (6.51). However, one may question whether this result applies at late times. In the case of a homogeneous background pump field, it is always the case that the growth is exponential at late times if the daughter field is bosonic. This is due to a build up in the occupancy number in certain k-modes, leading to rapid growth for fields satisfying Bose-Einstein statistics.

6.7.1 Floquet Analysis

Such exponential growth is obtained by fully solving the mode functions non-perturbatively [31]. Since the background is periodic, the mode functions satisfy a form of Hill's equation, albeit with infinitely many coupled oscillators due to the spatial structure of the oscillon. There is a large literature on resonance from homogeneous backgrounds, particularly relevant to inflation, but rarely are inhomogeneous backgrounds studied as we do here.

The late time behavior is controlled by Floquet exponents μ. To make this precise,

consider the discrete (matrix) version of the mode function equations

$$\dot{w}_{qk} = z_{qk} \tag{6.77}$$

$$\dot{z}_{qk} = \sum_{k'} Q_{kk'}(t) w_{qk'} \tag{6.78}$$

where

$$Q_{kk'}(t) \equiv -\omega_k^2 \delta_{kk'} - \frac{1}{V}\Phi(\mathbf{k}-\mathbf{k}', t) \tag{6.79}$$

is a matrix in k-space, with period $T = 2\pi/\omega$. Here we have labelled the mode functions w instead of v, since we will impose slightly different initial conditions on w. In particular consider the following pair of (matrix) initial conditions

$$(i) \qquad w_{qk}(0) = \delta_{qk}, \quad z_{qk}(0) = 0, \tag{6.80}$$

$$(ii) \qquad w_{qk}(0) = 0, \quad z_{qk}(0) = \delta_{qk}. \tag{6.81}$$

Let N be the number of q, k values in our discretization. Lets organize this information into a $2N \times 2N$ matrix $M(t)$, whose upper left quadrant is w_{qk} with IC (i), upper right quadrant is z_{qk} with IC (i), lower left quadrant is w_{qk} with IC (ii), and lower right quadrant is z_{qk} with IC (ii). So initially we have $M(0) = \mathbf{1}_{2N,2N}$. Numerically, we evolve this through one period, giving the matrix $M(T)$. After n oscillations, we have $M(nT) = M(T)^n$. Hence the matrix $M(T)$ controls the behavior of the system. To obtain the result for the initial conditions of (6.30, 6.31) we multiply $M(nT)$ onto the following diagonal matrix

$$\begin{pmatrix} \frac{1}{\sqrt{2\omega_k}} V \delta_{kq} & 0 \\ 0 & \frac{-i\omega_{\tilde{k}}}{\sqrt{2\omega_{\tilde{k}}}} V \delta_{\tilde{k}\tilde{q}} \end{pmatrix}. \tag{6.82}$$

The existence of exponential growth is governed by the eigenvalues of $M(T)$, with some corresponding eigenvector $\{w_k, z_k\}$. Following the standard Floquet theory, we write the eigenvalues as $\exp(\mu T)$, where μ are the Floquet exponents, which are in general complex. Note that although we have phrased this in the context of solving the quantum problem, it is also a classical stability analysis, as we mentioned at the end of Section 6.4.

6.7.2 Results

We have carried out the numerical analysis for different models, but would like to report on the results for the $\frac{1}{2}g_1\phi\chi^2$ theory with $\Phi \approx g_1\,\phi_\epsilon(x)\cos(\omega t)$ in $d = 1$. The numerical results for the maximum value of the real part of $\mu_{\max}$ as a function of g_1 is given in Fig. 6-4 (top panel, red curve). As the figure reveals, there is a critical value of the coupling $g_1^* \approx 0.2\sqrt{\lambda}$ governing the existence of exponential growth. For $g_1 > g_1^*$ exponential growth occurs, but for $g_1 < g_1^*$ it does not; in the latter regime all Floquet exponents are imaginary. This is reflected in the evolution of the energy $\delta E(t)$, which we have plotted in Fig. 6-4 (lower panel). The lower (orange) curve is

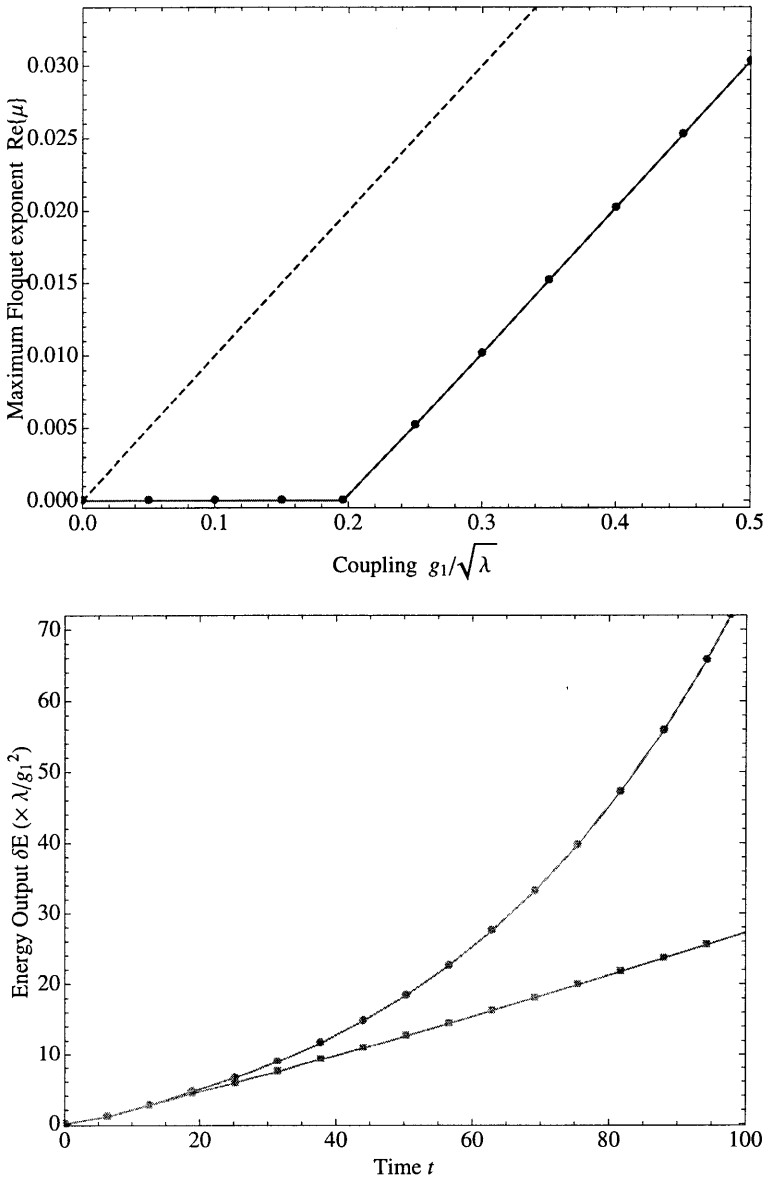

Figure 6-4: Top panel: the maximum value of the real part of the Floquet exponent $\mu_{\max}$ as a function of coupling $g_1/\sqrt{\lambda}$ for $\epsilon = 0.05$. Lower (red) curve is for an oscillon pump $\mu_{\max}$ and higher (blue-dashed) curve is for a homogeneous pump $\mu_{\mathrm{h,max}}$. Lower panel: total energy output $\delta E\,(\times\lambda/g_1^2)$ as a function of time for $g_1/\sqrt{\lambda} = 0.1$ (orange) and $g_1/\sqrt{\lambda} = 0.3$ (green).

for $g_1 = g_1^*/2$ and the upper (green) curve is for $g_1 = 3g_1^*/2$. Hence for sufficiently small couplings the perturbative analysis is correct – the growth is indeed linear at late times, but for moderate to large couplings the perturbative analysis breaks down – the growth is exponential at late times.

How can we understand this behavior? The answer resides in examining the structure of the corresponding eigenvector. For $g_1 > g_1^*$, we have numerically computed the eigenvector $\{w_k, z_k\}_{\max}$ corresponding to the maximum Floquet exponent. It is useful to represent this vector in position space, where it is some wave-packet. The real part of the (unnormalized) eigenvector is plotted in blue in Fig. 6-5 (top panel). In red we have also indicated the shape of the oscillon. We find that the shape of the wave-packet which carries $\mu_{\max}$ is approximately described by the function

$$\chi_{\max}(x) \sim \phi_\epsilon(x) \cos(k_{\rm rad}\, x) \tag{6.83}$$

where $k_{\rm rad}$ is the wavenumber we identified in the perturbative analysis ($\omega_k^2 = k^2 + m_\chi^2 = \omega/2$). Notice that the shape of this is independent of the coupling g_1. Such a result cannot make sense at arbitrarily small values of g_1. At sufficiently small g_1 *all* eigenvectors of the Floquet matrix $M(T)$ should be small deformations of plane waves, since we are then almost solving the Klein-Gordon equation. In particular, this means there should not be any localized wave-packet eigenvectors of the Floquet matrix. If the eigenvectors are spatially delocalized, they cannot grow exponentially, since there is nothing available to pump the wave at large distances from the oscillon. This explains why all μ are imaginary at sufficiently small g_1. Conversely, at sufficiently large g_1, some solutions can exist that are $\mathcal{O}(1)$ deviations from plane-waves, namely the wave-packet of Fig. 6-5. Clearly then it is inapplicable to treat this as a small perturbation from a plane wave. This explains why exponential growth can occur at sufficiently large coupling.

With this understanding, let's postdict the critical value of the coupling in this model. If we ignore the spatial structure and treat the background oscillon as homogeneous with amplitude $\phi_\epsilon(0)$, then χ's mode functions satisfy a Mathieu equation, whose properties are well known (let's call their Floquet exponents $\mu_{\rm h}$). In the regime of narrow resonance, the first instability band (connected to $\phi \to \chi + \chi$ decay) has a maximum growth rate $\mu_{\rm h,max} \approx g_1 \phi_\epsilon(0)/2$ (plotted as the blue-dashed curve in Fig. 6-4). On the other hand, the spatial structure of the oscillon means that modes that are produced in the core of the oscillon will try to "escape" at a rate set by the inverse of the oscillon's width. Let's define an escape rate as $\mu_{\rm esp} = 1/(2R_{\rm osc})$, where $R_{\rm osc}$ is the oscillon's radius. The critical g_1 can be estimated by the condition

$$\mu^*_{\rm h,max} \approx \mu_{\rm esc}. \tag{6.84}$$

To achieve exponential growth, we require $\mu_{\rm h,max} \gtrsim \mu_{\rm esc}$ in order for there to be sufficient time for growth to occur in the core of the oscillon before escaping, allowing Bose-Einstein statistics to be effective. Using $\phi_\epsilon(0) = 4\,\epsilon/\sqrt{\lambda}$ and $1/R_{\rm osc} \approx \epsilon$, gives $g_1^* \approx \sqrt{\lambda}/4$, in good agreement with the full numerical result. This reasoning can be extended to other scenarios. For $g_1 = 0$, we can focus on annihilation driven

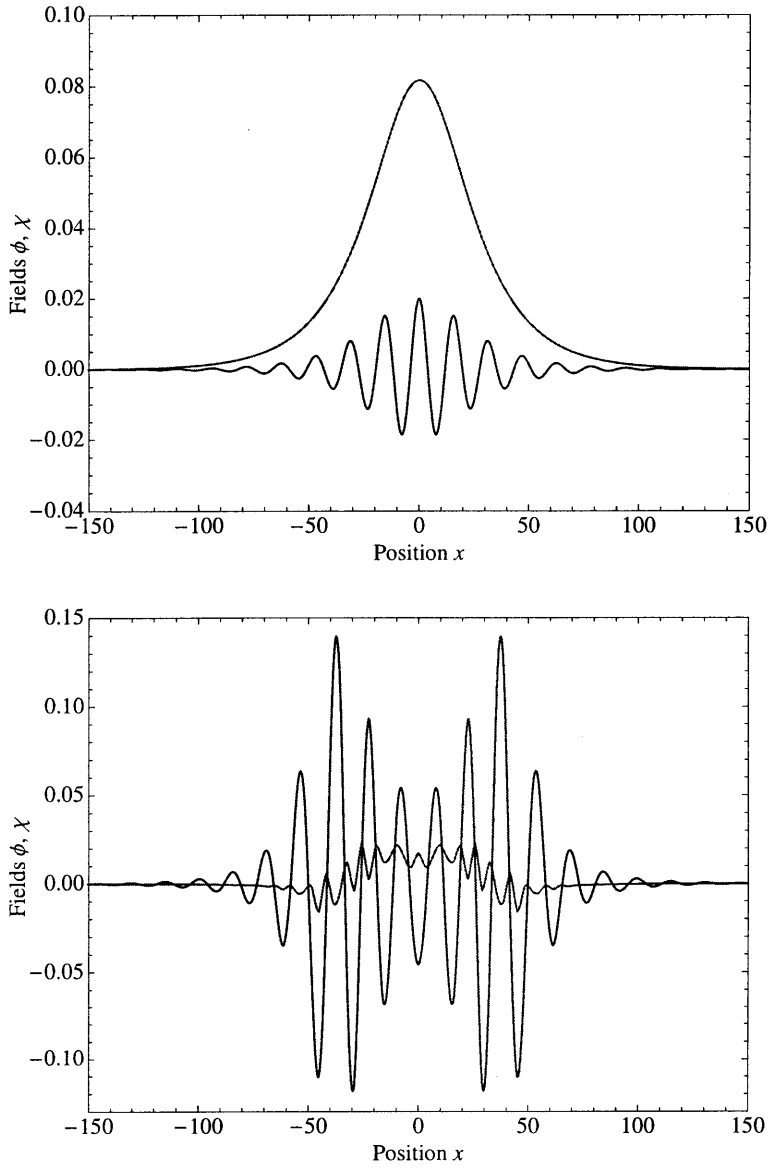

Figure 6-5: Top panel: The χ wave-packet which exhibits exponential growth for $g_1 > g_1^*$ (blue) and the ϕ-oscillon (red). Lower panel: The fields at $t = 80$ after classical evolution, with $g_1 = 0.8\sqrt{\lambda}$, $\epsilon = 0.05$, and $m_\chi = 0.3$.

by $\frac{1}{2}g_2\phi^2\chi^2$. In this case, study of the Mathieu equation reveals $\mu_{\rm h,max} \approx g_2\phi_\epsilon(0)^2/8$, leading to $g_2^* \approx \lambda/(4\,\epsilon)$. Since λ and $g_{1,2}$ are independent parameters, the regime $g_{1,2} > g_{1,2}^*$ is allowed (and easily satisfied for $\lambda\,\hbar \ll 1$, as required for massive oscillons).

If the parameters are in the regime of exponential growth it is interesting to note that substantial parametric resonance can occur from an inhomogeneous clump of energy established by oscillons. This is a form of parametric resonance – explosive energy transfer from a localized clump to a daughter field. Of course this cannot continue indefinitely, since the oscillon has only a finite amount of energy to transfer. Using the initial conditions of Fig. 6-5 (top panel) we have evolved the full coupled $\{\phi,\,\chi\}$ system under the classical equations of motion. We find that exponential growth in χ occurs initially and eventually this results in the destruction of the ϕ-oscillon as seen in Figure 6-5 (lower panel).

Finally, we return to the single field oscillon. In the $\lambda_5\,\phi^5$ model we originally discussed, the resulting (generalized) Mathieu equation reveals $\mu_{\rm h,max} \sim \lambda_5\,\phi_\epsilon(0)^3$, leading to $\lambda_5^* \sim \lambda^{3/2}/\epsilon^2$. However, in our expansion we implicitly assumed λ_5 was $\mathcal{O}(1)$ w.r.t ϵ and therefore we should never enter the regime $\lambda_5 > \lambda_5^*$. Similarly, consider the classic $-\lambda\,\phi^4$ model. The resulting (generalized) Mathieu equation reveals $\mu_{\rm h,max} \sim \lambda^2\phi_\epsilon(0)^4 \sim \epsilon^4$. Comparing this to $\mu_{\rm esc} \sim \epsilon$ we see that it is impossible to obtain exponential growth for small ϵ. (This can change for the wide flat-top oscillons of Ref. [19].) As far as we are aware, this is the first *explanation* of the stability of small amplitude oscillons against exponential growth of short wavelength perturbations. This fact was previously only seen empirically. It is quite interesting that at sufficiently small amplitude, or couplings, the oscillon is stable against exponential growth in perturbations and yet it still has modes that grow linearly with time. This occurs in the limit of degenerate eigenvalues of $M(T)$. These modes seem relatively rare and harmless classically, but they must be integrated over in the quantum theory, resulting in steady decay.

6.8 Collapse Instabilities

Some oscillons are unstable to perturbations with wavelengths comparable to the size of the oscillon ($k = \mathcal{O}(\epsilon)$). Although this is not the focus of our work, we would like to briefly discuss this phenomenon for completeness. These instabilities are so prominent that they often appear in classical simulations starting from initial conditions away from the "perfect oscillon profile", given by the expansion eq. (6.10).

Using the numerical method of the previous Section we have found that for $k = \mathcal{O}(\epsilon)$ there are exponentially growing modes in $d = 3$, but not in $d = 1$, for the $-\lambda\,\phi^4$ theory. Since the corresponding wavelengths are comparable to the size of the oscillon, these instabilities are easily seen in numerical simulations starting from arbitrary initial conditions; an example is displayed in Figure 6-6 for the $V_I \sim -\lambda\,\phi^4$ potential in $d = 3$. In the simulation, we find that the field is growing in the core of the oscillon and we also find that it is spatially collapsing. The existence of this instability is well known in the literature.

To feel more confident that this is the correct behavior of the quantum theory, let

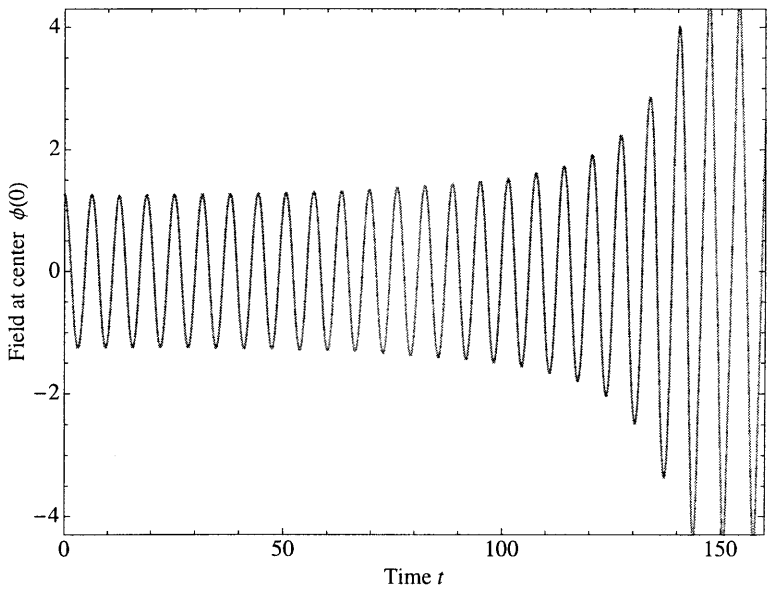

Figure 6-6: Field at center of oscillon over time for $V_I = -\phi^4/4!$ in $d = 3$. We choose $\epsilon = 0.05$ and set up initial conditions with $\phi(0, \rho) = 1.03\,\epsilon\,\phi_1(0, \rho)$, i.e., an initial profile 3% higher than the preferred oscillon profile; this clearly causes an instability by $t \sim 150$.

us turn now to address this problem in a different approximation. At some level we should consider the oscillon as a collection of ϕ-particles in the full quantum theory [34]. Since these "collapse" or "self-focussing" type of instabilities occur at small wavenumbers $k = \mathcal{O}(\epsilon)$, perhaps we can interpret this as resulting from the interaction of non-relativistic particles. The leading interaction is $2\,\phi \rightarrow 2\,\phi$ scattering. Lets consider $V_I = \frac{1}{3!}\lambda_3\,\phi^3 + \frac{1}{4!}\lambda_4\,\phi^4 + \ldots$. At tree-level this scattering process occurs due to 4 diagrams: s, t, and u-channels generated by 2 insertions of the ϕ^3 vertex and a single contact term from the ϕ^4 vertex. The matrix element is easily computed in the non-relativistic limit $i\mathcal{M}(2\,\phi \rightarrow 2\,\phi) = -i\left(\frac{5}{3}\lambda_3^2 - \lambda_4\right)$. This can be recast in position space as a two particle potential $V(\mathbf{r}_1, \mathbf{r_2})$ by taking the inverse Fourier transform and multiplying by $\hbar^2/4$ due to our normalization convention. This gives

$$V(\mathbf{r}_1, \mathbf{r}_2) = -\frac{\hbar^2}{4}\left(\frac{5}{3}\lambda_3^2 - \lambda_4\right)\delta^d(\mathbf{r}_1 - \mathbf{r}_2). \tag{6.85}$$

(The s, t, and u-channels actually produce a type of Yukawa potential if the full relativistic $\mathcal{M}$ is used, but this shouldn't modify our conclusions.) This is a short range force and is attractive if and only if $\frac{5}{3}\lambda_3^2 - \lambda_4 > 0$, which is a condition for the existence of oscillons. (This condition also emerges in the classical small ϵ expansion). In the non-relativistic regime, the wavefunction for N_ϕ particles making up an oscillon will be governed by the Schrödinger equation

$$\left(-\frac{\hbar^2}{2m_\phi}\sum_{i=1}^{N_\phi}\nabla_i^2 + \sum_{i<j}V(\mathbf{r}_i, \mathbf{r}_j)\right)\psi_{N_\phi} = E_{N_\phi}\,\psi_{N_\phi} \tag{6.86}$$

with the above potential V. It is known that a δ-function potential permits unbounded solutions for $d \geq 2$ ($d = 2$ is marginal, being only a logarithmic divergence, but $d \geq 3$ is a power law). A localized gas of ϕ-particles, such as the oscillon, would be expected to be unstable to radial collapse under these conditions. Hence, we expect $d > 2$ to be unstable (with $d = 2$ marginal). For $d = 3$ a collapse time of $\sim 1/(\lambda\,\hbar\,\epsilon^2)$ is naively expected.

There is evidence in the literature [22, 19] that oscillon stability is controlled by the derivative of the oscillon's mass $M_{\rm osc}$ w.r.t amplitude ϕ_a. If the derivative is positive (negative), then the oscillon appears to be stable (unstable) to collapse. Since the canonical small amplitude oscillon satisfies $M_{\rm osc} \sim \phi_a^{2-d}$, we obtain a consistent result. However, this leading order behavior must break down at some point for sufficiently large amplitudes and can be affected by the inclusion of higher order terms in the potential. In fact it is known that beyond a critical amplitude no collapse instability exists in 3-d [22, 19] (also see the Q-ball literature [35, 36, 37]). Collapse in 3-d is also absent in other field theories, such as the SU(2) sector of the standard model [4, 5] and any model where $\phi_\epsilon \sim \epsilon^2$ instead of the canonical $\phi_\epsilon \sim \epsilon$ (e.g., see [18]).

In summary, the collapse of an oscillon is highly model dependent and can be avoided by operating in the appropriate number of dimensions, parameter space, or field theory. However, the radiation we computed in the previous sections is unavoidable.

6.9 Conclusions

We have found that even though an oscillon can have a mass that is much greater than the mass of the individual quanta, the classical decay can be very different to the quantum decay (this point does not appear to have been appreciated in the literature, for instance see the concluding sections of Refs. [5, 6]). The radiation of both classical and quantum oscillons can be understood in terms of forced oscillator equations, see eqs. (6.15, 6.41). We derived the frequency and wavenumber of the outgoing radiation, which were both $\mathcal{O}(1)$ in natural units. Since a classical oscillon has a spread which is $\mathcal{O}(1/\epsilon)$ in position space, it has a spread which is $\mathcal{O}(\epsilon)$ in k-space. Its Fourier modes are therefore exponentially small at the radiating wavenumber and hence such radiation is exponentially suppressed. In the quantum theory, there simply *cannot* be modes whose amplitudes are exponentially suppressed. Instead, zero-point fluctuations ensure that all modes have at least $\mathcal{O}(\hbar)$ amplitude-squared due to the uncertainty principle.

We derived a formula for the quantum lifetime of an oscillon $\sim 1/(\lambda\hbar\epsilon^p)$. The power p is model dependent: $p = 4$ in the $\phi^3 + \ldots$ theory (or $-\phi^4 + \phi^5 + \ldots$), and $p = 6$ in the $-\phi^4 + \phi^6 + \ldots$ theory. Through a Floquet analysis, we explained why the growth of perturbations of small amplitude oscillons is linear in time, as opposed to exponential. The dimensionless $\lambda\hbar$ controls the magnitude of the decay rate, as it should for a leading order in $\hbar$ analysis. For example, the Standard Model Higgs potential has $\lambda\hbar \sim (m_{\rm H}/v_{\rm EW})^2 \sim 0.1\,(m_{\rm H}/100\,{\rm GeV})^2$ and so this is not very small. On the other hand, the effective $\hbar$ of the QCD axion potential is $\lambda\hbar \sim (\Lambda_{\rm QCD}/f_a)^4 \sim 10^{-48}\,(10^{10}\,{\rm GeV}/f_a)^4$ and so oscillons formed from axions, called "axitons" in [8], are governed by classical decay (ignoring coupling to other fields).

We further considered the fate of an oscillon that is coupled to a second scalar χ and found it to either decay or annihilate with a growth in χ that can be exponentially fast, depending on parameters. Since oscillons may form substantially in the early universe [10, 11] this may give rise to interesting phenomenology. At the very least, it presents a plausible cosmological scenario in which a parametric pump field exists that is qualitatively different to the homogeneous oscillations of the inflaton during p/reheating. This is a form of parametric resonance: explosive transfer of energy from a localized clump into bosonic daughter fields. We expect decay into fermions to be quite different (for discussion in the context of Q-balls, see [38]). This may have some cosmological relevance.

It appears that if a field has a perturbative decay channel, then the oscillon will eventually decay through it. This is important because we expect most fields in nature to be perturbatively unstable, including the inflaton, p/reheating fields, Higgs, and most fields beyond the standard model. A good exception is dark matter. This conclusion may seem surprising given that the oscillon is a bound state of particles with a finite binding energy [34]. However, oscillons are formed from fields whose particle number is not conserved. One could imagine a situation in which m_ϕ is only slightly greater than $2\,m_\chi$, and in this case the oscillon's binding energy may prevent direct decays into χ's, but this requires fine tuning and will not forbid $2\,\phi \to 2\,\chi$ or $3\,\phi \to 2\,\phi$ or $4\,\phi \to 2\,\phi$ annihilations.

We conclude that in many scenarios an individual oscillon's lifetime will be shorter than the age of the universe at the time of production (this may prevent individual oscillons from having cosmological significance in such cases). Exceptions include the GUT era [11], inflation [13], and axitons produced at the QCD phase transition [8]. An interesting question for further study is whether oscillons can form and then decay, and then form again repeatedly, like subcritical bubbles in hot water. It is not implausible that such a process could continue over long time scales for cosmic temperatures of order the field's mass; similar to the production and disappearance of unstable particles in a relativistic plasma. This may modify cosmological thermalization.

Bibliography

[1] S. Coleman, "Q-balls", Nuc. Phys. B, **262**, 263 - 283 (1985).

[2] H. Segur and M. D. Kruskal, "Nonexistence of small-amplitude breather solutions in ϕ^4 theory", Phys. Rev. Lett. **58**, 747 - 750 (1987).

[3] G. Fodor, P. Forgacs, Z. Horvath and M. Mezei, "Radiation of scalar oscillons in 2 and 3 dimensions," Phys. Lett. B **674** (2009) 319 [arXiv:0903.0953 [hep-th]].

[4] E. Farhi, N. Graham, V. Khemani, R. Markov and R. Rosales, "An oscillon in the SU(2) gauged Higgs model," Phys. Rev. D **72** (2005) 101701 [arXiv:hep-th/0505273].

[5] N. Graham, "An Electroweak Oscillon," Phys. Rev. Lett. **98** (2007) 101801 [Erratum-ibid. **98** (2007) 189904] [arXiv:hep-th/0610267].

[6] N. Graham, "Numerical Simulation of an Electroweak Oscillon," Phys. Rev. D **76** (2007) 085017 [arXiv:0706.4125 [hep-th]].

[7] M. Gleiser and J. Thorarinson, "A Class of Nonperturbative Configurations in Abelian-Higgs Models: Complexity from Dynamical Symmetry Breaking," Phys. Rev. D **79** (2009) 025016 [arXiv:0808.0514 [hep-th]].

[8] E. W. Kolb and I. I. Tkachev, "Nonlinear axion dynamics and the formation of cosmological pseudosolitons", Phys. Rev. D, **49** 10 (1994).

[9] N. Graham and N. Stamatopoulos, "Unnatural oscillon lifetimes in an expanding background," Phys. Lett. B **639** (2006) 541 [arXiv:hep-th/0604134].

[10] E. Farhi, N. Graham, A. H. Guth, N. Iqbal, R. R. Rosales and N. Stamatopoulos, "Emergence of Oscillons in an Expanding Background," Phys. Rev. D **77** (2008) 085019 [arXiv:0712.3034 [hep-th]].

[11] E. J. Copeland, M. Gleiser and H. R. Muller, "Oscillons: Resonant configurations during bubble collapse," Phys. Rev. D **52** (1995) 1920 [arXiv:hep-ph/9503217].

[12] I. Dymnikova, L. Koziel, M. Khlopov and S. Rubin, "Quasilumps from first-order phase transitions," Grav. Cosmol. **6**, 311 (2000) [arXiv:hep-th/0010120].

[13] M. Gleiser, "Oscillons in scalar field theories: Applications in higher dimensions and inflation," Int. J. Mod. Phys. D **16** (2007) 219 [arXiv:hep-th/0602187].

[14] M. Gleiser, B. Rogers and J. Thorarinson, "Bubbling the False Vacuum Away," Phys. Rev. D **77** (2008) 023513 [arXiv:0708.3844 [hep-th]].

[15] M. Hindmarsh and P. Salmi, "Oscillons and Domain Walls," Phys. Rev. D **77**, 105025 (2008) [arXiv:0712.0614 [hep-th]].

[16] M. Alcubierre, R. Becerril, S. F. Guzman, T. Matos, D. Nunez and L. A. Urena-Lopez, "Numerical studies of Φ^2-oscillatons," Class. Quant. Grav. **20** (2003) 2883 [arXiv:gr-qc/0301105].

[17] D. N. Page, "Classical and quantum decay of oscillatons: Oscillating self-gravitating real scalar field solitons," Phys. Rev. D **70** (2004) 023002 [arXiv:gr-qc/0310006].

[18] G. Fodor, P. Forgacs and M. Mezei, "Mass loss and longevity of gravitationally bound oscillating scalar lumps (oscillatons) in D-dimensions," arXiv:0912.5351 [gr-qc].

[19] M. A. Amin and D. Shirokoff, "Flat-top oscillons in an expanding universe," arXiv:1002.3380 [astro-ph.CO].

[20] P. M. Saffin and A. Tranberg, "Oscillons and quasi-breathers in D+1 dimensions," JHEP **0701** (2007) 030 [arXiv:hep-th/0610191].

[21] S. Kichenassamy, "Breather solutions of the nonlinear wave equation", Comm. Pur. Appl. Math. **44**, 789 (1991).

[22] G. Fodor, P. Forgacs, Z. Horvath and A. Lukacs, "Small amplitude quasi-breathers and oscillons," Phys. Rev. D **78** (2008) 025003 [arXiv:0802.3525 [hep-th]].

[23] J. P. Boyd, "A numerical calculation of a weakly non-local solitary wave: the ψ^4 breather", Nonlinearity **3** 177–195 (1990).

[24] J. P. Boyd, "A hyperasymptotic perturbative method for computing the radiation coefficient for weakly nonlocal solitary waves", J. Comput. Phys. 120 15–32 (1995).

[25] G. Fodor, P. Forgacs, Z. Horvath and M. Mezei, "Computation of the radiation amplitude of oscillons," arXiv:0812.1919 [hep-th].

[26] M. Gleiser and D. Sicilia, "Analytical Characterization of Oscillon Energy and Lifetime," Phys. Rev. Lett. **101** (2008) 011602 [arXiv:0804.0791 [hep-th]].

[27] M. Gleiser and D. Sicilia, "A General Theory of Oscillon Dynamics," arXiv:0910.5922 [hep-th].

[28] A. B. Adib, M. Gleiser and C. A. S. Almeida, "Long-lived oscillons from asymmetric bubbles," Phys. Rev. D **66** (2002) 085011 [arXiv:hep-th/0203072].

[29] S. Kasuya, M. Kawasaki and F. Takahashi, "I-balls," Phys. Lett. B **559**, 99 (2003) [arXiv:hep-ph/0209358].

[30] S. Weinberg, "The Quantum Theory of Fields, Vol 1", Cambridge University Press (1995).

[31] L. Kofman, A. D. Linde and A. A. Starobinsky, "Towards the theory of reheating after inflation," Phys. Rev. D **56** (1997) 3258 [arXiv:hep-ph/9704452].

[32] M. Yoshimura, "Catastrophic particle production under periodic perturbation," Prog. Theor. Phys. **94** (1995) 873 [arXiv:hep-th/9506176].

[33] M. Yoshimura, "Decay rate of coherent field oscillation," arXiv:hep-ph/9603356 (1996).

[34] R. F. Dashen, B. Hasslacher and A. Neveu, "The Particle Spectrum In Model Field Theories From Semiclassical Functional Integral Techniques," Phys. Rev. D **11** (1975) 3424.

[35] D. L. T. Anderson, "Stability of time-dependent particlelike solutions in nonlinear field theories. 2," J. Math. Phys. **12** (1971) 945.

[36] T. D. Lee and Y. Pang, "Nontopological solitons," Phys. Rept. **221** (1992) 251.

[37] T. I. Belova and A. E. Kudryavtsev, "Solitons and their interactions in classical field theory," Phys. Usp. **40** (1997) 359 [Usp. Fiz. Nauk **167** (1997) 377].

[38] A. G. Cohen, S. R. Coleman, H. Georgi and A. Manohar, "The Evaporation Of Q-balls," Nucl. Phys. B **272** (1986) 301.

Chapter 7

Finite Contributions to Entanglement Entropy

We show that for a massive scalar field in its ground state the entanglement entropy between the interior and exterior of a spatial domain of arbitrary shape receives a finite contribution that is proportional to the area of the dividing boundary. For systems involving local boundary constraints we demonstrate that there are also subleading finite contributions to the area term: for a field confined to a waveguide in 3+1 dimensions we identify terms proportional to the waveguide's cross sectional area, perimeter length, and curvature. We carry out related calculations for massless fields and comment on the kinds of systems in which these finite contributions may be measurable.

7.1 Introduction

For observations confined to a subsystem A, entanglement entropy is a measure of one's ignorance of the full system due to entanglement between the degrees of freedom in the subsystem and its complement B. It is defined by the von Neumann formula $S = -\mathrm{Tr}_{\mathrm{A}}(\rho_{\mathrm{A}} \ln \rho_{\mathrm{A}})$, where $\rho_{\mathrm{A}} = \mathrm{Tr}_{\mathrm{B}}\rho$ is the reduced density matrix of the subsystem A. This object plays a role in various contexts, including quantum field theory, condensed matter, quantum computing, and black holes.

The entanglement entropy in d+1-dimensional systems typically obeys an area law $S \sim A_{d-1}/\epsilon^{d-1}$ for $d \geq 2$, where A_{d-1} is the $d-1$-dimensional area of the dividing boundary between the subsystem and its complement. However, the constant of proportionality is divergent, since it blows up with the UV cutoff ϵ of the field theory [1]. By contrast, the entanglement entropy of 1+1-dimensional systems is finite, since the ϵ dependence is only logarithmic. For example, the entanglement entropy between a pair of half spaces of a 1+1-dimensional CFT at correlation length ξ was shown to be $S = c/6 \ln \xi$, where c is the central charge of the CFT [2, 3]. This begs the question whether there are finite contributions to the entanglement entropy in $d \geq 2$ dimensions.

In this chapter we demonstrate that there are indeed. For a free scalar field in

d+1-dimensions at finite correlation length ξ (i.e., mass $\mu = 1/\xi$), we show that in addition to the divergent terms, such as $S \sim A_{d-1}/\epsilon^{d-1}$, there is also a finite area law contribution

$$\begin{aligned} \Delta S &= \gamma_d \frac{A_{d-1}}{\xi^{d-1}} \ln \xi, \quad \text{for } d \text{ odd}, \\ \Delta S &= \gamma_d \frac{A_{d-1}}{\xi^{d-1}}, \qquad \text{for } d \text{ even}, \end{aligned} \tag{7.1}$$

where $\gamma_d \equiv (-1)^{\frac{d-1}{2}}[6\,(4\pi)^{\frac{d-1}{2}}((d-1)/2)!]^{-1}$ for d odd and $\gamma_d \equiv (-1)^{d/2}[12\,(2\pi)^{(d-2)/2}(d-1)!!]^{-1}$ for d even. Using heat kernel methods, we exhibit further finite power law corrections for a waveguide geometry through the imposition of boundary conditions; see Fig. 7-1 left panel. This setup allows us to formulate the entropy as an expansion in terms of the geometric properties of the waveguide's cross section. Also, we consider a massless field and compute finite terms for the interval in a waveguide; see Fig. 7-1 right panel.

7.2 Heat Kernel Method

A powerful way to compute the von Neumann entropy, which involves the logarithm of the reduced density matrix, is to use the following identity: $S = -\mathrm{Tr}_{\mathrm{A}}(\rho_{\mathrm{A}} \ln \rho_{\mathrm{A}}) = -\frac{d}{dn} \ln \mathrm{Tr}\rho_{\mathrm{A}}^n|_{n=1}$, known as the "replica trick". Consider, for example, a system in 1 spatial dimension. The quantity $\mathrm{Tr}\rho_{\mathrm{A}}^n$ is a trace over an n-sheeted Riemann surface with cut along the subsystem of interest A. If the subsystem A is a half space, then as explained in Ref. [2] $\mathrm{Tr}\rho_{\mathrm{A}}^n$ is proportional to the partition function Z_δ on a space of deficit angle $\delta = 2\pi(1-n)$. Hence the entropy can be recast as

$$S = \left(2\pi \frac{d}{d\delta} + 1\right) \ln Z_\delta \Big|_{\delta=0}. \tag{7.2}$$

In this chapter we consider the waveguide geometry of Fig. 7-1 left panel. For this geometry we formulate a Euclidean field theory on the space $C_\delta \times \mathcal{M}_{d-1}$, where C_δ is a 2-dimensional cone of radius R (infrared cutoff) and deficit angle δ, and $\mathcal{M}_{d-1}$ is the $(d-1)$-dimensional cross section of the waveguide. The cone's radius corresponds to the physical region in space we are tracing over, and the angular direction is associated with a geometric temperature (imaginary time) direction in the Euclidean path integral. Let Z_δ be the partition function for a field in its ground state defined on this space. Considering a free field theory, the partition function can be automatically evaluated since the path integral is Gaussian. For a field of inverse mass ξ, the partition function is

$$\ln Z_\delta = -\frac{1}{2} \ln \det\left(-\Delta + \xi^{-2}\right), \tag{7.3}$$

where Δ is the Laplacian satisfying the appropriate boundary conditions on $C_\delta \times \mathcal{M}_{d-1}$.

Now lets introduce the heat kernel for the Laplacian operator $\zeta(t) \equiv \mathrm{tr}\left(e^{t\Delta}\right)$. The trace is defined by imposing Dirichlet or Neumann boundary conditions on the waveguide $\partial\mathcal{M}_{d-1}$ and imposing Dirichlet boundary conditions on the cone ∂C_δ. This allows us to rewrite the partition function Z_δ and hence the entropy S in terms of the heat kernel as follows:

$$S = \frac{1}{2}\int_0^\infty \frac{dt}{t}\left(2\pi\frac{d}{d\delta}+1\right)\zeta(t)e^{-t/\xi^2}\Big|_{\delta=0}. \tag{7.4}$$

Since the manifold for the Euclidean field theory of the waveguide is separable into a direct product $C_\delta \times \mathcal{M}_{d-1}$, so too is the heat kernel

$$\zeta(t) = \zeta_\delta(t)\,\zeta_{d-1}(t). \tag{7.5}$$

This simplifies the problem into obtaining expansions of two separate heat kernels. One is an expansion for the 2-dimensional cone $\zeta_\delta(t)$ and the other is an expansion for the $(d-1)$-dimensional cross section of the waveguide $\zeta_{d-1}(t)$. The heat kernel for the cone has the following form [6]

$$\zeta_\delta(t) = \frac{1}{12}\left(\frac{2\pi}{2\pi-\delta}-\frac{2\pi-\delta}{2\pi}\right)+\ldots \tag{7.6}$$

where the dots represent terms that are either annihilated by the $2\pi\frac{d}{d\delta}+1$ operator, or are vanishing in the $R\to\infty$ limit. In either case it is only the piece here that contributes. This gives the following general expression for the entanglement entropy of a waveguide in d spatial dimensions traced over half-space:

$$S = \frac{1}{12}\int_0^\infty \frac{dt}{t}\zeta_{d-1}(t)e^{-t/\xi^2}. \tag{7.7}$$

Hence the entropy is determined by the geometry of the cross section of the waveguide through its heat kernel $\zeta_{d-1}(1)$.

7.3 Waveguide Cross Section

The heat kernel for a closed domain satisfying either Dirichlet ($\eta=-1$) or Neumann ($\eta=+1$) boundary conditions in dimensions 0, 1, and 2 has the following small t expansion [4]:

$$\begin{aligned}
\zeta_0(t) &= 1\\
\zeta_1(t) &= \frac{a}{2\sqrt{\pi t}}+\frac{\eta}{2}+\ldots\\
\zeta_2(t) &= \frac{A}{4\pi t}+\frac{\eta P}{8\sqrt{\pi t}}+\chi+\ldots
\end{aligned} \tag{7.8}$$

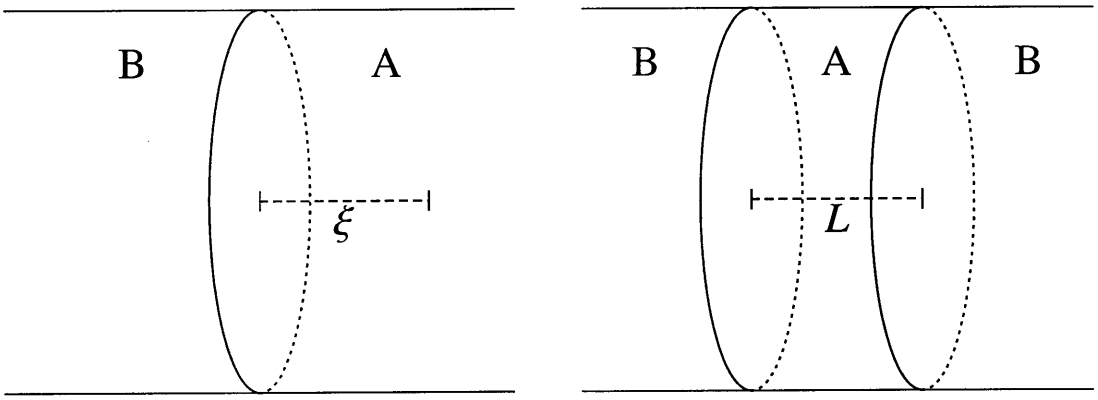

Figure 7-1: Waveguide geometry in $d = 3$. Left: Region A is a half space at finite correlation length ξ. Right: Region A is an interval of length L at criticality.

Here a is the cross sectional length of a 2-dimensional waveguide. Also, A, P, and χ is the cross sectional area, perimeter, and curvature of a 3-dimensional waveguide, respectively.[1] The curvature term for an arbitrary piecewise smooth 2-dimensional cross section is given by

$$\chi = \sum_i \frac{1}{24}\left(\frac{\pi}{\alpha_i} - \frac{\alpha_i}{\pi}\right) + \sum_j \frac{1}{12\pi}\int_{\gamma_j} \kappa(\gamma_j) d\gamma_j, \tag{7.9}$$

where α_i is the interior angle of any sharp corners and $\kappa(\gamma_j)$ is the curvature of any smooth pieces. For example, $\chi = 1/6$ for any smooth shape (such as a circle) and $\chi = (n-1)/(n-2)/6$ for any n-sided polygonal (so $\chi = 1/4$ for a square). The curvature term differs from that in Ref. [5] where the curvature piece was argued to be proportional to the number of corners in an arbitrary shape. The result of Ref. [5] is only correct for a square cross section.

7.4 Regularization and Finite Terms

If we were to directly insert the above expansion into eq. (7.7) for the entanglement entropy we would encounter a problem. Integrating from $t = 0^+$ is associated with arbitrarily short distance physics. As is well known this gives rise to a divergence in any number of dimensions that cannot be renormalized away; logarithmic in 1 dimension, linear in 2 dimensions, and quadratic in 3 dimensions [1]. Of course such a divergence may be regulated. For instance, we could impose a hard cutoff on the t integral and integrate from $t = t_c = \epsilon^2$ to $t = \infty$, and find terms that only diverge in the $\epsilon \to 0$ limit. Another procedure is to use Pauli-Villars regularization by subtracting off terms with $\mu = 1/\xi$ replaced by Λ etc and taking Λ large. This is perhaps more appealing as it respects the underlying geometry. However, either approach obtains divergent terms depending on ϵ or Λ, whose coefficients depend on the choice of regularization.

However by returning to eq. (7.7) it is easy to identify finite dependence of the entropy on the inverse correlation length $\mu = 1/\xi$. In general, the leading order behavior of the heat kernel as $t \to 0$ is

$$\zeta_{d-1}(t) = \frac{\alpha}{t^{(d-1)/2}} + \dots \tag{7.10}$$

where $\alpha = A_{d-1}/(4\pi)^{(d-1)/2}$ is a constant. Inserting this into eq. (7.7) reveals that the integrand has the leading order behavior $\sim 1/t^{(d+1)/2}$; giving a divergence of order $d-1$ as $t \to 0^+$ with respect to a cutoff, say ϵ, defined through $t_c = \epsilon^2$ (implicitly the $d-1=0$ case means a logarithmic divergence). But this can be regulated by taking some number of partial derivatives of the entropy with respect to the correlation length ξ, as this pulls down factors of t from the exponential $\exp(-t/\xi^2)$. In particular

[1] This expansion is also of use in computations of the Casimir effect between two partitions in a waveguide, see Refs. [7, 8].

by taking

$$k \equiv \text{Floor}\left[\frac{d+1}{2}\right] \tag{7.11}$$

($k = 1$ for $d = 1$, $k = 1$ for $d = 2$, $k = 2$ for $d = 3$, etc) derivatives of S with respect to ξ^{-2} gives a manifestly finite integral. Hence a dimensionless finite quantity we can define is

$$S_\xi \equiv (-\xi^{-2})^k \frac{\partial^k S}{\partial(\xi^{-2})^k}. \tag{7.12}$$

Using eqs. (7.7) and (7.8) and integrating t in the domain $t \in (0, \infty)$, it is straightforward to obtain

$$\begin{aligned} S_\xi &= \frac{1}{12} && \text{for } d = 1, \\ S_\xi &= \frac{1}{24}\frac{a}{\xi} + \frac{\eta}{24} + \ldots && \text{for } d = 2, \\ S_\xi &= \frac{1}{48\pi}\frac{A}{\xi^2} + \frac{\eta}{192}\frac{P}{\xi} + \frac{\chi}{12} + \ldots && \text{for } d = 3. \end{aligned} \tag{7.13}$$

For $d = 1$ this result is exact [2, 3]. For $d \geq 2$ this expansion is only valid in the domain $a \gg \xi$, where a is a typical cross sectional length; see the appendix for exact results in $d = 2$ and a square cross section in $d = 3$. Note that by doing the appropriate number of anti-derivatives (1 for $d = 2$, 1 for $d = 2$, and 2 for $d = 3$) it is straightforward to obtain contributions to the entropy.

7.5 General Geometries

Although the sub-leading terms in eq. (7.13) are specific to a waveguide geometry, the leading terms have a meaning for arbitrary geometries. In particular, for any boundary in 1 dimension we pick up a contribution of $1/12$ to S_ξ, as is well known [9]. For any closed geometry in 2 dimensions, the leading contribution is $S_\xi = P/(24\xi)$, where P is the perimeter length. While for any closed geometry in 3 dimensions, the leading contribution is $S_\xi = A/(48\pi\xi^2)$. By integrating up these results, we recover the $d = 1, 2, 3$ cases that we reported in eq. (7.1). Furthermore, using the heat kernel in arbitrary dimensions (7.10) we recover the general result for arbitrary dimensions.

This general result differs from estimates made in Section 7 of Ref. [13], where the corresponding term in the entropy was not found. Numerically we have checked this law for the case of the sphere and the cylinder, finding excellent agreement. In fact our numerics indicates that the area term is the only polynomial contribution to S_ξ for large A/ξ^2. We can understand this as follows: in the regime $\xi \ll a$, where a is a typical length scale of curvature of the boundary, the correlations required to feel the curvature are exponentially suppressed. By contrast, if the boundary were to contain a sharp corner, there is expected to be power law corrections. (For

related discussion, see [14, 15]). We have numerically verified this for the square. For smooth geometries, though, local boundary constraints appear to be required for the existence of subleading corrections since they modify the spectrum – as we found for the waveguide.

7.6 Experimental Realization

Measurement of this entanglement requires externally changing the correlation length in such a way that the microphysics is unaffected. One could imagine achieving this for magnetic systems, by operating below the critical temperature. In the absence of external fields, there exist massless modes, but for an externally applied B-field ϕ acquires a mass $\mu = 1/\xi$ that is adjustable. By operating in the regime: $\epsilon^2 \ll \xi^2 \ll A$ (where ϵ is the inter-spin spacing) the area law should be an adequate description and can be numerically quite large; this can be contrasted with the 1-dimensional case $S_\xi = 1/12$, which is small.

7.7 Criticality - Interval

The previous expansion requires the system to be away from criticality. Let us turn now to the critical case ($\xi \to \infty$). To define a finite entropy we must have a length scale, which now comes from an interval of length L; see Fig. 7-1 right panel. In this case we can define the following finite quantity:

$$S_L \equiv L\frac{dS}{dL}. \tag{7.14}$$

The small t heat kernel expansion of eq. (7.6) is insufficient here because we must know the form of $\zeta_\delta(t)$ for not only $t \lesssim L^2$, but also $t \gtrsim L^2$ where t is large.

In general the full form of $\zeta_\delta(t)$ is unknown. However, we do not need $\zeta_\delta(t)$ for arbitrary δ, we only need the specific limit indicated in eq. (7.4). There are powerful tools developed to obtain this, often in $\mathcal{E}$ space rather than t space, which we follow. The derivative of the entanglement entropy can be written in terms of an object defined for 2d conformal field theories known as the c-function C; specifically it is the inverse Laplace transform of $\frac{1}{12}(2\pi\frac{d}{d\delta}+1)\zeta_\delta(t)|_{\delta=0}$. By convolving with the transverse density of states we have

$$S_L = \int d\mathcal{E}\, C\big(L\sqrt{\mathcal{E}}\big)\rho_{d-1}(\mathcal{E}). \tag{7.15}$$

The c-function C has been studied closely, see Refs. [10, 11, 12]. It is known that $C(0) = 1/3$ and it is monotonically decreasing. Using the heat kernel expansion (7.8), we can inverse Laplace transform to obtain a density of states expansion for $\rho_{d-1}(\mathcal{E})$. The quantity S_L is then expressible in terms of a few integrals of C, which have been

computed numerically. We find

$$\begin{aligned} S_L &= \frac{1}{3} && \text{for } d=1, \\ S_L &= k_1 \frac{a}{L} + \frac{\eta}{6} + \ldots && \text{for } d=2, \\ S_L &= k_2 \frac{A}{L^2} + \frac{\eta\, k_1}{4}\frac{P}{L} + \frac{\chi}{3} + \ldots && \text{for } d=3. \end{aligned} \tag{7.16}$$

Here $k_1 \equiv \frac{1}{\pi}\int_0^\infty dx\, C(x)$ and $k_2 \equiv \frac{1}{2\pi}\int_0^\infty dx\, x\, C(x)$. The numerical values are: $k_1 \approx 0.039$ and $k_2 \approx 0.0098$. For $d \geq 2$ this expansion is valid in the domain $a \gg L$; analogous to the previous expansion which was valid in the domain $a \gg \xi$.

7.8 Discussion

We have shown that arbitrary shaped domains have a finite area term, given by eq. (7.1). A cutoff independent quantity comes from taking $k = \text{Floor}[(d+1)/2]$ derivatives of S with respect to ξ^{-2}. This cutoff independent quantity can be defined for a single domain A. This is unlike the strong subadditivity proposal, requiring 2 subsystems of interest A and A', and forming $S_{A\cup A'} + S_{A\cap A'} - S_A - S_{A'}$. For a waveguide geometry we used our construction to obtain an asymptotic expansion of the entropy for small correlation length to cross section width ratio. For arbitrary smooth manifolds the leading order area law should be applicable, large, and perhaps measurable. It would be of interest to extend these results to other fields, such as fermions, and to interacting field theories.

7.9 Appendix: Some Exact Results

In $d = 2$ and for a square cross section of width a in $d = 3$, $\zeta_{d-1}(t)$ is known exactly. The result for S_ξ is found to be

$$\begin{aligned} S_\xi &= \frac{a\coth(a/\xi)}{24\,\xi} + \frac{\eta}{24}, && \text{for } d = 2 \\ S_\xi &= \frac{a^2}{48\pi\xi^2}\left[1 + 2\,\frac{a}{\xi}\sum_{n,m}{}' f_{n,m} K_1(2 f_{n,m} a/\xi)\right] \\ &\quad + \frac{\eta\, a}{48\,\xi}\left[\coth(a/\xi) + \frac{a}{\xi}\,\mathrm{csch}(a/\xi)\right] + \frac{1}{48}, && \text{for } d = 3 \end{aligned}$$

where $f_{n,m} \equiv \sqrt{n^2 + m^2}$, the primed summation means $\{n, m\} \in \mathbb{Z}^2/\{0, 0\}$, and K_1 is the modified Bessel function of the second kind of order 1. For $a \gg \xi$ we recover eq. (7.13) plus exponentially small corrections. (Note that for the square $A = a^2$, $P = 4\,a$, $\chi = 1/4$.)

Bibliography

[1] M. Srednicki, "Entropy and area," Phys. Rev. Lett. **71** (1993) 666 [arXiv:hep-th/9303048].

[2] C. G. . Callan and F. Wilczek, "On geometric entropy," Phys. Lett. B **333** (1994) 55 [arXiv:hep-th/9401072].

[3] C. Holzhey, F. Larsen and F. Wilczek, "Geometric and renormalized entropy in conformal field theory," Nucl. Phys. B **424** (1994) 443 [arXiv:hep-th/9403108].

[4] H. P. Baltes and E. R. Hilf, "Spectra of Finite Systems," (Bibliographisches Institut, Mannheim, 1976).

[5] D. V. Fursaev, "Entanglement entropy in critical phenomena and analogue models of quantum gravity," Phys. Rev. D **73** (2006) 124025 [arXiv:hep-th/0602134].

[6] O. Alvarez, "Theory Of Strings With Boundaries: Fluctuations, Topology, And Quantum Geometry," Nucl. Phys. B **216** (1983) 125.

[7] M. P. Hertzberg, R. L. Jaffe, M. Kardar and A. Scardicchio, "Attractive Casimir Forces in a Closed Geometry," Phys. Rev. Lett. **95** (2005) 250402 [arXiv:quant-ph/0509071].

[8] M. P. Hertzberg, R. L. Jaffe, M. Kardar and A. Scardicchio, "Casimir Forces in a Piston Geometry at Zero and Finite Temperatures," Phys. Rev. D **76** (2007) 045016 [arXiv:0705.0139 [quant-ph]].

[9] P. Calabrese and J. L. Cardy, "Entanglement entropy and quantum field theory," J. Stat. Mech. **0406** (2004) P002 [arXiv:hep-th/0405152].

[10] H. Casini and M. Huerta, "A finite entanglement entropy and the c-theorem," Phys. Lett. B **600** (2004) 142 [arXiv:hep-th/0405111].

[11] H. Casini, C. D. Fosco, and M. Huerta, "Entanglement and alpha entropies for a massive Dirac field in two dimensions," J. Stat. Mech. **0507**, (2005) 007 [arXiv:cond-mat/0505563].

[12] H. Casini and M. Huerta, "Entanglement and alpha entropies for a massive scalar field in two dimensions," J. Stat. Mech. **0512**, (2005) 012 [arXiv:cond-mat/0511014].

[13] S. Ryu and T. Takayanagi, "Aspects of holographic entanglement entropy," JHEP **0608** (2006) 045 [arXiv:hep-th/0605073].

[14] H. Casini and M. Huerta, "Universal terms for the entanglement entropy in 2+1 dimensions," Nucl. Phys. B **764** (2007) 183 [arXiv:hep-th/0606256].

[15] H. Casini and M. Huerta, "Entanglement entropy in free quantum field theory," arXiv:0905.2562 [hep-th].

Chapter 8

Casimir Forces in a Piston Geometry at Zero and Finite Temperatures

We study Casimir forces on the partition in a closed box (piston) with perfect metallic boundary conditions. Related closed geometries have generated interest as candidates for a repulsive force. By using an optical path expansion we solve exactly the case of a piston with a rectangular cross section, and find that the force always attracts the partition to the nearest base. For arbitrary cross sections, we can use an expansion for the density of states to compute the force in the limit of small height to width ratios. The corrections to the force between parallel plates are found to have interesting dependence on the shape of the cross section. Finally, for temperatures in the range of experimental interest we compute finite temperature corrections to the force (again assuming perfect boundaries).

8.1 Introduction

A striking macroscopic manifestation of quantum electrodynamics is the attraction of *neutral* metals. In 1948 Casimir predicted that such a force results from the modification of the ground state energy of the photon field due to the presence of conducting boundary conditions [1]. The energy spectrum is modified in a fashion that depends on the separation between the plates, a. While the zero-point energy is itself infinite, its variation with a gives rise to a finite force. High precision measurements, following the pioneering work of Lamoreaux in 1997 [2], have renewed interest in this subject. A review of experimental attempts to measure the force prior to 1997, and the many improvements since then, can be found in Ref. [3]. As one example, we note experiments by Mohideen *et al.*[4], using an atomic force microscope, which have confirmed Casimir's prediction from 100nm to several μm, to a few percent accuracy. Forces at these scales are relevant to operation of micro-electromechanical systems (MEMS), such as the actuator constructed by Chan *et al.* [5] to control the frequency of oscillation of a nanodevice. They also appear as an undesirable background in precision

experiments such as those that test gravity at the sub-millimeter scale [6].

An undesirable aspect of the Casimir attraction is that it can cause the collapse of a device, a phenomenon known as "stiction" [7]. This has motivated the search for circumstances where the attractive force can be reduced, or even made repulsive [8]. The Casimir force, of course, depends sensitively on shape, as evidenced from comparison of known geometries from parallel plates, to the sphere opposite a plane [9], the cylinder opposite a plane [10], eccentric cylinders [11], the hyperboloid opposite a plane [12], a grating [13], a corrugated plane [14]. The possibility of a repulsive Casimir force between perfect metals can be traced to a computation of energy of a spherical shell by Boyer [15], who found that the finite part of this energy is opposite in sign to that for parallel plates. This term can be regarded as a positive pressure favoring an increased radius for the sphere, *if it were the only consequence of changing the radius.* The same sign is obtained for a square in 2-dimensions and a cube in 3 dimensions [16, 17]. For a parallelepiped with a square base of width b and height a, the finite part of the Casimir energy is positive for aspect ratios of $0.408 < a/b < 3.48$. This would again imply a repulsive force in this regime if there were no other energy contributions accompanying deformations at a fixed aspect ratio. Of course, it is impossible to change the size of a material sphere (or cube) without changing its surface area, and other contributions to its cohesive energy. For example, a spherical shell cut into two equal hemispheres which are then separated has superficial resemblance to the Boyer calculation. However the cut changes the geometry, and it can in fact be shown[18] that the two hemispheres attract.

The piston geometry, first considered by Cavalcanti [19] (in 2 dimensions) and further considered in Refs. [20, 21] (in 3 dimensions), is closely related to the parallelepiped discussed above.[1] As depicted in Fig. 8-1, we examine a piston of height h, with a movable partition at a distance a from the lower base. The simplest case is that of a rectangular base, but this can be generalized to arbitrary cross sections. This set-up is experimentally realizable, and does not require any deformations of the materials as the partition is moved. The force resulting from rigid displacements of this piece is perfectly well defined, and free from various ambiguities due to cut-offs and divergences that will be discussed later. In particular, we indeed find the finite part of the energy can be "repulsive" if only one of the boxes adjoining the partition is considered, while if both compartments are included, the net force on the partition is attractive (in the sense that it is pulled to the closest base).

This chapter expands on a previous brief publication of our results [20], and is organized as follows. Section 8.2 introduces the technical tools preliminary to the calculations, and includes sections on cutoffs and divergences, the optical path approach, and on the decomposition of the electromagnetic (EM) field into two scalar field (transverse magnetic and transverse electric) with Dirichlet and Neumann boundary conditions (respectively). Details of the computation for pistons of rectangular cross section are presented in Section 8.3, and the origin of the cancellations leading to a net attractive force on the partition is discussed in some detail. Interestingly, it is possible to provide results that are asymptotically exact in the limit of small sepa-

[1]The piston geometry was earlier mentioned in Ref. [22].

rations for cross sections of arbitrary shape. As discussed in Section 8.4, there is an interesting dependence on the shape in this limit, related to the resolution with which the cross section is viewed. Finally, in Section 8.5 we present new results pertaining to corrections to the Casimir force at finite temperatures in such closed geometries (for perfect metals). We conclude with a brief summary (Section 8.6), and an Appendix.

8.2 Preliminaries

Before embarking on the calculation of the force on the partition, we introduce some relevant concepts in this Section. Subsection 8.2.1 discusses the general structure of the divergences appearing in the calculation of zero-point energies, and indirectly justifies our focus on the piston geometry. The optical path approach, which is our computational method of choice is reviewed in Sec. 8.2.2. Another important aspect of the piston geometry is that it enables the decomposition of the EM field into Dirichlet and Neumann scalar fields, as presented in Sec. 8.2.3.

8.2.1 Cutoff Dependence

Let us consider an empty cavity made of perfectly conducting material. The Casimir energy of the EM field is a sum over the zero point energies of all modes compared to the energies in the absence of the material $E_C(\Lambda) = E(\Lambda) - E_0(\Lambda) = \frac{1}{2}\sum^{\Lambda} \hbar\omega_m - \frac{1}{2}\sum^{\Lambda} \hbar\omega_m^0$, and is divergent if the upper limit Λ is taken to infinity. In a physical realization, the upper cutoff is roughly the plasma frequency of the metal, as it separates the modes that are reflected and those that are transmitted and are hence unaffected by the presence of the metallic boundaries. Based on general results for the density of states in a cavity [23], we know $E(\Lambda)$ has an asymptotic form, with a leading term proportional to the volume V of the cavity and the fourth power of Λ and sub-leading terms proportional to its surface area S, a length L which is related to the average curvature of the walls (in a cavity with edges but otherwise flat, like a parallelepiped, it is the total length of the edges) proportional to Λ^3 and Λ^2 respectively, and so forth. For example in the case of a scalar field with Dirichlet boundary conditions, we find

$$E(\Lambda) = \frac{3}{2\pi^2} V\Lambda^4 - \frac{1}{8\pi} S\Lambda^3 + \frac{1}{32\pi} L\Lambda^2 + ... + \widetilde{E}, \tag{8.1}$$

where "..." denote lower order cutoff dependences,[2] and $\widetilde{E}$ is the important finite part in the limit of $\Lambda \to \infty$. The EM field also enjoys a similar expansion, although some terms may be absent.

Although the volume term is cancelled by an identical term in E_0, this is not obviously the case for the other divergent terms (surface area, perimeter, and so on). The energy of an isolated cavity is therefore dependent on the physical properties of

[2] For general geometries, there are also linear and logarithmic terms in Λ, but for the class of geometries examined in this chapter (pistons) there are no further terms in Λ.

the metal. A determination of the stresses in a single closed cavity requires detailed considerations of the metal, and its extrapolation to the perfectly conducting limit will be problematic [24]. It is tempting to ignore these cutoff dependent terms, and to remove them in analogy to the renormalization of ultraviolet divergences in quantum field theories. This is unjustified as there are no boundary counter-terms to cancel them, see Ref. [25]. If however, we are interested in the force between rigid bodies, then any surface, perimeter, *etc.* terms are independent of the distance between them, and a well defined (finite) force exists in the perfect conductor limit.

While the piston geometry considered in this chapter is closely related to the parallelepiped cavities considered in the literature, it does not suffer from problems associated with changes in shape. The overall volume, surface, and perimeter contributions are all unchanged as the height of the partition is varied, and the force acting on it is finite and well defined. The same observations led Cavalcanti [19] to consider a rectangular (2-dimensional) piston. He found that the force on the partition, though weakened relative to parallel lines, is attractive.

8.2.2 Optical Approach

The Casimir energy can be expressed as a sum over contributions of *optical paths* [26], and much intuition into the problem is gained by classifying the corresponding paths. For generic geometries this approach yields only an approximation to the exact result that ignores diffraction. Fortunately, it is exact for rectilinear geometries if reflections from edges and corners are properly included.

Consider a free scalar field in spatial domain $\mathcal{D}$ obeying some prescribed boundary conditions (Dirichlet or Neumann) on the boundary $\mathcal{B} = \partial\mathcal{D}$. The Casimir energy is defined as the sum over the zero point energies, $E = \sum \frac{1}{2}\hbar\omega$, where ω are the eigenfrequencies in $\mathcal{D}$ (we refrain from subtracting E_0 for the moment). This expression needs to be regularized, as explained in the previous section, by some cutoff Λ. We choose to implement this by a smoothing function $S_\Lambda(k) = e^{-k/\Lambda}$, and thus examine $E(\Lambda) = \sum_k \frac{1}{2}\hbar\omega(k)e^{-k/\Lambda}$.

The Casimir energy can be expressed in terms of the spectral Green's function $G(\mathbf{x}, \mathbf{x}', k)$ which satisfies the Helmholtz equation in $\mathcal{D}$ with a point source,

$$(\nabla'^2 + k^2)G(\mathbf{x}', \mathbf{x}, k) = -\delta^3(\mathbf{x}' - \mathbf{x})\,, \tag{8.2}$$

and subject to the same boundary conditions on $\mathcal{B}$ as the field. The Casimir energy of a scalar field is then given by the integral over space and wavenumber of the imaginary part of G, in the coincidence limit $\mathbf{x}' \to \mathbf{x}$ [25] ($\hbar = c = 1$), as

$$E(\Lambda) = \frac{1}{\pi}\Im \int_0^\infty dk\, k^2 e^{-k/\Lambda} \int_{\mathcal{D}} d\mathbf{x}\, G(\mathbf{x}, \mathbf{x}, k)\,. \tag{8.3}$$

The knowledge of the Helmholtz Greens function at coincident points allows us to calculate the Casimir energy of the configuration.

It is convenient to introduce a fictitious time t and a corresponding space-time propagator, $G(\mathbf{x}', \mathbf{x}, t)$, defined as the Fourier transform of $G(\mathbf{x}', \mathbf{x}, k)$. The propagator

$G(t)$ can be expressed as the functional integral of a free quantum particle of mass $m = 1/2$ with appropriate phases associated with paths that reflect off the boundaries.

In the optical approach, the path integral is approximated as a sum over *classical* paths of $\exp[iS_{p_r}(\mathbf{x}', \mathbf{x}, t)]$, weighted by the Van Vleck determinant $D_r(\mathbf{x}', \mathbf{x}, t)$ [27]. Here $S_{p_r}(\mathbf{x}', \mathbf{x}, t)$ is the classical action of a path p_r from $\mathbf{x}$ to $\mathbf{x}'$ in time t, composed of straight segments and undergoing r reflections at the walls. For rectilinear geometries, like the parallelepiped that we will discuss, this is exact and effectively generalizes the method of images to the Helmholtz equation.

For definiteness consider a scalar field satisfying either Dirichlet or Neumann boundary conditions, introducing a parameter η, which is -1 for the Dirichlet and $+1$ for the Neumann case. The Green's function is then given by

$$G(\mathbf{x}', \mathbf{x}, k) = \sum_{p_r} \frac{\phi(p_r, \eta)}{4\pi l_{p_r}(\mathbf{x}', \mathbf{x})} e^{ikl_{p_r}(\mathbf{x}', \mathbf{x})}, \tag{8.4}$$

where l_{p_r} is the length of the path from $\mathbf{x}$ to $\mathbf{x}'$ along p_r. There is a phase factor $\phi(p_r, \eta) = \eta^{n_s + n_c}$ with n_s and n_c the number of surface and corner reflections, respectively. Note that reflections from an *edge* do not contribute to the phase.

Since paths without reflections or with only one reflection can have zero length, they require a frequency cutoff Λ, implemented by the smoothing function $S_\Lambda(k) = e^{-k/\Lambda}$. Then the $\mathbf{x}$ and k integrals can be exchanged, the k integral performed and the Casimir energy written as

$$E_\eta(\Lambda) = \frac{1}{2\pi^2} \sum_{p_r} \phi(p_r, \eta) \int_{\mathcal{D}} d\mathbf{x} \frac{\Lambda^4 (3 - (l_{p_r}(\mathbf{x})\Lambda)^2)}{(1 + (l_{p_r}(\mathbf{x})\Lambda)^2)^3}. \tag{8.5}$$

The limit $\Lambda \to \infty$ can be taken in each term of the sum, unless a path has zero length, which can occur only for cases with $r = 0$ or $r = 1$. After isolating these two contributions, we set

$$E_\eta(\Lambda) = E_0(\Lambda) + E_1(\Lambda, \eta) + \widetilde{E}_\eta. \tag{8.6}$$

The zero reflection path has exactly zero length, and contributes the energy $E_0(\Lambda) = \frac{3}{2\pi^2} V \Lambda^4$, where V is the volume of the space. This is a constant and therefore does not contribute to the Casimir force. The one reflection paths (energy $E_1(\Lambda, \eta)$) generate cutoff dependent terms, but generically, also cutoff-independent terms. We will show however that such one reflection terms do not contribute to the force when specialized to the piston geometry.

For paths undergoing multiple reflections $r > 1$, the length l_{p_r} is always finite, and we can safely send $\Lambda \to \infty$ in eq. (8.5), resulting in the simpler and cutoff independent contribution

$$\widetilde{E}_\eta = -\frac{1}{2\pi^2} \sum_{p_r > 1} \phi(p_r, \eta) \int_{\mathcal{D}} d\mathbf{x} \frac{1}{l_{p_r}(\mathbf{x})^4}. \tag{8.7}$$

This is a finite contribution to the energy in the limit $\Lambda \to \infty$. The derivative of $\widetilde{E}_\eta$ gives the finite force between the rigid bodies.

8.2.3 Electromagnetic Field Modes

In the previous section we defined the optical approach for a scalar field. Although a similar definition can be made for the electromagnetic field in an arbitrary geometry, the Helmholtz equation becomes matrix–valued, complicating the treatment even in a semiclassical approximation. However, in the piston geometry, with arbitrary cross section, the EM field can be separated into transverse magnetic (TM) and transverse electric (TE) modes, that satisfy Dirichlet and Neumann boundary conditions.

At the surface of an ideal conductor, the $\mathbf{E}$ and $\mathbf{B}$ fields satisfy the boundary conditions, $\mathbf{E} \times \mathbf{n} = \mathbf{0}$ and $\mathbf{B} \cdot \mathbf{n} = 0$, where $\mathbf{n}$ is the normal vector at the surface. The normal modes of the piston consist of a TM set, which satisfy

$$E_x = \psi(y,z)\cos(n\pi x/a), \quad n = 0,1,2,\cdots, \tag{8.8}$$

where ψ vanishes on the boundaries of the domain, and therefore satisfies Dirichlet conditions on the 2-dimensional boundary; and a TE set, with

$$B_x = \phi(y,z)\sin(n\pi x/a), \quad n = 1,2,3,\cdots, \tag{8.9}$$

where ϕ satisfies Neumann boundary conditions. The other components of $\mathbf{E}$ and $\mathbf{B}$ can be computed from Maxwell's equations, and are easily shown to obey conducting boundary conditions. There is, however, one important exception: the TE mode built from the trivial Neumann solution, $\phi =$ constant, does not satisfy conducting boundary conditions unless the constant (and all components of $\mathbf{E}$ and $\mathbf{B}$) are zero. We must ensure that the corresponding set of modes in eq. (8.9) are not included in the Casimir summation.

Equations (8.8) and (8.9) enable us to list the spectrum of the electromagnetic field. Denote the spectra of the TM modes as the set $\Omega(N_I \otimes D_S) \subset \mathbb{R}$. Here N_I indicates that the component on the interval satisfies Neumann boundary conditions, and D_S indicates that the component on the cross section satisfies Dirichlet boundary conditions. Similarly, we denote the spectra of the TE modes as $\Omega(D_I \otimes N_S)$ in a similar notation. However, as explained above, we must remove $\Omega(D_I)$, which are the frequencies with $\phi =$ constant. Hence, the electromagnetic spectra is the set

$$\Omega_C = \Omega(N_I \otimes D_S) \cup \Omega(D_I \otimes N_S) \setminus \Omega(D_I). \tag{8.10}$$

Note that $\Omega(D_I) = \{\pi/a, 2\pi/a, \ldots\}$ is the set of eigenfrequencies in 1-dimension. The Dirichlet and Neumann spectra on the interval are identical except for the $n = 0$ mode (see eqs. (8.8) and (8.9)), but the energy of this mode is independent of a and does not contribute to the Casimir force. So we may replace $N_I \to D_I$ in the TM spectrum and $D_I \to N_I$ in the TE spectrum, with the result

$$\Omega_C \approx \Omega(D_I \otimes D_S) \cup \Omega(N_I \otimes N_S) \setminus \Omega(D_I), \tag{8.11}$$

where the notation $\approx$ indicates equality up to terms independent of a. Thus, the EM spectrum is the union of Dirichlet and Neumann spectra in the 3-dimensional

domain, $\mathcal{D}$, except that the Dirichlet spectrum on the interval must be taken out.

8.3 Rectangular Piston

8.3.1 Derivation

The piston geometry is depicted in Fig. 8-1. The domain $\mathcal{D}$ consists of the whole parallelepiped, the union of Regions I and II. Only the partition, located a distance a from the base and $h-a$ from the top, is free to move. We study the scalar field for both Dirichlet and Neumann boundary conditions and the electromagnetic field. According to eq. (8.11), the EM Casimir energy arises from the sum of the Dirichlet and Neumann energies in 3 dimensions minus the Dirichlet Casimir energy in one dimension, $E = h\Lambda^2/2\pi - \zeta(2)/(4\pi a) - \zeta(2)/(4\pi(h-a))$ (a standard result). In total, then, the EM Casimir force on the partition is

$$F_C = F_D + F_N + \frac{\zeta(2)}{4\pi a^2} - \frac{\zeta(2)}{4\pi(h-a)^2}, \tag{8.12}$$

where the final term vanishes if we take $h \to \infty$.

Let us initially focus on Region I, the parallelepiped of size $a \times b \times c$, below the partition. The optical energy receives contributions from the sum over all closed paths p_r in domain $\mathcal{D}_I$: Each path is composed of straight segments with equal angles of incidence and reflection when bouncing off the walls. There are four distinct classes of paths: $E_{\rm per}$, from periodic orbits reflecting off faces (*e.g.* paths (c), (d), (i)); $E_{\rm aper}$, from aperiodic tours off faces (*e.g.* paths (a), (e), (f)); $E_{\rm edge}$, from closed paths involving reflections off edges (*e.g.* paths (b), (g)); and $E_{\rm cnr}$, from closed paths with reflections off corners (*e.g.* path (h)). To each path p_r, we associate a vector $\mathbf{l}_{p_r}$ pointed along the *initial* heading of the path, and of length $|\mathbf{l}_{p_r}| = l_{p_r}$.

First we consider the periodic orbits, which are paths that involve an even number of reflections off faces, with $r = \{0, 2, 4, \ldots\}$ (*e.g.* paths (c), (d), (i)). As the central point is varied throughout $\mathcal{D}_I$, the length l_{p_r} of each periodic path remains fixed, making the integration trivial, *i.e.* $\int_{\mathcal{D}_I} d^3x \to abc = V$. We index the paths by integers n, m, l, so $\mathbf{l}_{p_r} = (2na, 2mb, 2lc)$, with $l_{nml} = \sqrt{(2na)^2 + (2mb)^2 + (2lc)^2}$. The $n = m = l = 0$ term gives $E_0 = \frac{3}{2\pi^2} V\Lambda^4$ (see eq. (8.5) with $l_{p_r} = 0$), while all others are evaluated using eq. (8.7), giving:

$$\begin{aligned} E^I_{\rm per}(\Lambda) &= \frac{3}{2\pi^2} V\Lambda^4 - \frac{abc}{32\pi^2} Z_3(a, b, c; 4) && (8.13) \\ &= \frac{3}{2\pi^2} V\Lambda^4 - \frac{\zeta(4)}{16\pi^2}\frac{A}{a^3} + \Gamma(b, c) + \mathcal{O}\left(e^{-2\pi g/a}\right), \text{ as } a \to 0 && (8.14) \end{aligned}$$

where $Z_d(a_1, \ldots, a_d; s)$ is the Epstein zeta function defined in the Appendix (eq. (8.50)), and $\Gamma(b, c)$ does not depend on a and hence does not contribute to the force on the piston. In eq. (8.14) $g \equiv \min(b, c)$, and $A = bc$ is the area of the base. The leading cutoff independent piece as $a \to 0$ is the Casimir energy for parallel plates, coming

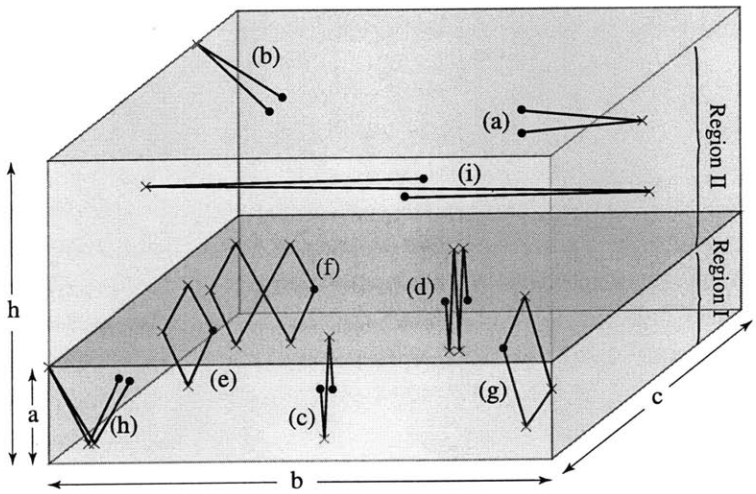

Figure 8-1: The 3-dimensional piston of size $h \times b \times c$. A partition at height a separates it into Region I and Region II. A selection of representative paths are given in (a)–(i). Several of these paths (namely, (a,b,c,d,h,i)) have start and end points that actually coincide, but we have slightly separated them for clarity.

from orbits that reflect off both the base and partition, see paths (c), (d), *etc.* in Fig. 8-1. To extract this behavior we have used

$$Z_d(a_1,\ldots,a_d;s) = \frac{2\zeta(s)}{a_1^s} + \mathcal{O}\left(\frac{1}{a_1}\right), \tag{8.15}$$

in the regime $a_1 \ll a_2,\ldots,a_d$ (see Appendix).

We next consider the contribution of the aperiodic orbits that involve an odd number of reflections off faces, with $r = \{1,3,5,\ldots\}$. Examples in the figure include paths (a), (e), and (f). For each such path, when we vary the point of integration over $\mathcal{D}_I$ one of the Cartesian components of the length vector $\mathbf{l}_{p_r}$ changes and the other two components are fixed. For example, only the x component varies for those paths that undergo an odd number of reflections off walls parallel to the yz-plane and an even number of reflections off walls parallel to both the xy and xz-planes. The x-component $\mathbf{l}_{p_r}$ increases by $2a$ each time that the number of reflections off the yz-planes increases, so that $\mathbf{l}_{p_r} = (2a(n-1)+2\xi(x),\, 2bm,\, 2cl)$, where $\xi(x) = x$ or $\xi(x) = a - x$ depending on the direction of the path. The summation over n and the x-integral $\int_0^a dx$ together combine to form an integral over x from $-\infty$ to $+\infty$. So we introduce $l_{p_r}(x) = \sqrt{(2x)^2+(2bm)^2+(2cl)^2}$ in terms of which the integration over the fixed components y and z is trivial $\int dydz = bc$, and the x-integration is elementary. The above example singled out the x-component. To include all such paths, the analysis must be repeated for the other two components under the cyclic interchange of a, b, and c. Employing eq. (8.5) for $\{n,m\} = \{0,0\}$ and eq. (8.7) for $\{n,m\} \neq \{0,0\}$ we obtain,

$$\begin{aligned} E^I_{\text{aper}}(\Lambda) &= \frac{\eta}{8\pi}S\Lambda^3 - \frac{\eta}{64\pi}\Big(abZ_2(a,b;3) + acZ_2(a,c;3) + bc\,Z_2(b,c;3)\Big) && (8.16)\\ &= \frac{\eta}{8\pi}S\Lambda^3 - \eta\frac{\zeta(3)}{64\pi}\frac{P}{a^2} + \Upsilon(b,c) + \mathcal{O}\big(e^{-2\pi g/a}\big)\,, \text{ as } a\to 0 && (8.17)\end{aligned}$$

where Υ does not depend on a, $S = 2(ab+ac+bc) = aP + 2A$ is the total surface area, and $P = 2b + 2c$ is the perimeter of the base. The leading cutoff independent piece as $a \to 0$ comes from paths that reflect once off a side wall and off both the base and partition, see paths (e), (f), *etc.* in Fig. 8-1.

Next we calculate the contribution of even reflection paths which intersect an edge of Region I. Examples include (b) and (g) in Fig. 8-1. In this case it is only the component of $\mathbf{l}_{p_r}$ *parallel* to the edge that remains fixed, while the other 2 components vary as the point of origin varies over $\mathcal{D}_I$. For example, suppose the reflecting edge is parallel to the z-axis. Then the z-integration is trivial, $\int_0^c dz = c$, and the path vector is a function of x and y given by $\mathbf{l}_{p_r} = (2a(n-1)+2\xi(x),\, 2b(m-1)+2\psi(y),\, 2cl)$, where $\xi(x) = x$ or $\xi(x) = a-x$, $\psi(y) = y$ or $\psi(y) = b-y$, depending on the quadrant that $\mathbf{l}_{p_r}$ lies in: up or down in x, right or left in y, respectively. In this case we can replace both the summations over n and m and the double integral over x and y by an integral over the whole xy-plane. This integral is most easily performed in polar co-ordinates, using a path length that may be written as $l_{p_r}(r) = \sqrt{(2an)^2+(2r)^2}$.

The contribution to the Casimir energy is found to be

$$E^I_{\text{edge}}(\Lambda) = \frac{1}{32\pi} L\Lambda^2 - \frac{\zeta(2)}{16\pi}\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}\right), \tag{8.18}$$

where $L = 4(a+b+c) = 4a+2P$ is the total perimeter length. The cutoff independent piece $\sim 1/a$ comes from paths that reflect once off a side edge and off both the base and partition, see path (g), *etc.* in Fig. 8-1.

Finally, we consider the paths which reflect off a corner (E_{cnr}). In this case, as the integration variable moves throughout its domain, *all* components of the distance vector $\mathbf{l}_{p_r}$ vary. Hence, we can incorporate all such paths by extending our integral over all space in x, y and z. This leaves no dependence on the geometry of the parallelepiped (i.e., it is independent of a, b, and c), and only contributes a constant that is of no interest, which we ignore.

Adding together all these contributions, we obtain the Casimir energy of a scalar field in Region I as

$$E^I_\eta(\Lambda) = \frac{3}{2\pi^2} V\Lambda^4 + \frac{\eta}{8\pi} S\Lambda^3 + \frac{1}{32\pi} L\Lambda^2 + \widetilde{E}^I_\eta, \tag{8.19}$$

where $\widetilde{E}^I_\eta$ gives the cutoff independent piece, from eqs. (8.13), (8.16), and (8.18). We note that the cutoff dependent terms agree with the leading terms obtained by integrating Balian and Bloch's asymptotic expansion of the density of states [23].

We obtain the Casimir energy for the entire piston by adding to eq. (8.19) the analogous expression for Region II obtained by the replacements: $a \to h-a$, $V \to hA$, $S \to hP + 4A$, and $L \to 4h + 4P$. It is easy to see that after including Region II the sum of all cutoff dependent terms is independent of partition height a. Therefore the force on the partition is well defined and finite in the limit $\Lambda \to \infty$. Also, of course, the contribution to the Casimir energy from the region outside the piston is independent of a and can be ignored entirely. The force on the partition is given by the partial derivative with respect to a of the cutoff independent terms as

$$F_\eta = -\frac{\partial}{\partial a}\left(\widetilde{E}_\eta(a,b,c) + \widetilde{E}_\eta(h-a,b,c)\right), \tag{8.20}$$

where we have defined $\widetilde{E}^I_\eta = \widetilde{E}_\eta(a,b,c)$.

We focus on the $h \to \infty$ limit in which the expression for the contribution from Region II simplifies. Consider the periodic, aperiodic, and edge paths whose cutoff independent contribution to the energy is given in eqs. (8.13), (8.16), and (8.18). Replacing $a \to h - a$, taking $h \to \infty$, and using eq. (8.52) of the appendix in these

equations gives

$$\begin{aligned}
\widetilde{E}^{II}_{\rm per} &\to -\frac{h-a}{32\pi^2 A} Z_2(b/c, c/b; 4), \\
\widetilde{E}^{II}_{\rm aper} &\to -\eta \frac{h-a}{32\pi} \left(\frac{1}{b^2} + \frac{1}{c^2}\right) \zeta(3), \\
\widetilde{E}^{II}_{\rm edge} &\to 0,
\end{aligned} \tag{8.21}$$

where we have not reported terms independent of a, since they do not affect the force. Also, we re-express the Region I energy $\widetilde{E}_\eta(a,b,c)$ in a fashion that is useful for $a \ll b, c$, using eq. (8.51) of the appendix. The net force on the partition due to quantum fluctuations of the *scalar field* is then

$$\begin{aligned}
F_\eta &= -\frac{3\zeta(4)}{16\pi^2}\frac{A}{a^4} - \eta\frac{\zeta(3)}{32\pi}\frac{P}{a^3} - \frac{\zeta(2)}{16\pi a^2} - \frac{J_\eta(b/c)}{32\pi^2 A} \\
&+ \eta\frac{\pi}{2a^3}\sum_{m,n=1}^{\infty} n^2 \left(K_0\left(2\pi mn\, b/a\right) b + (b \leftrightarrow c)\right) \\
&+ \frac{\pi^2}{32}\frac{A}{a^4}{\sum_{m,n}}' \frac{\coth(f_{mn}(b/a, c/a))}{f_{mn}(b/a, c/a)\sinh^2(f_{mn}(b/a, c/a))},
\end{aligned} \tag{8.22}$$

where the primed summation is over $\{m,n\} \in \mathbb{Z}^2 \setminus \{0,0\}$ and K_0 is the zeroth order modified Bessel function of the second kind. Here we have defined $J_\eta(x) \equiv Z_2(x^{1/2}, x^{-1/2}; 4) + \pi\eta(x + x^{-1})\zeta(3)$ and $f_{mn}(x,y) \equiv \pi\sqrt{(mx)^2 + (ny)^2}$. The first four terms of eq. (8.22) dominate for $a \ll b, c$, while the following terms are exponentially small in this regime. The first term arises from periodic orbits reflecting off walls (see eq. (8.14)), the second term from aperiodic tours bouncing off walls (see eq. (8.17)), the third term from reflections off edges (see eq. (8.18)), the fourth term from Region II paths (see eq. (8.21)). Note that the infinite series, involving exponentially small terms, is convergent for any a, b, c.

The electromagnetic case is closely related to the scalar Dirichlet and Neumann cases, which we discussed in detail in Section 8.2.3. According to eq. (8.12), the EM Casimir energy in Region I is related to the Dirichlet energy E_D and the Neumann energy E_N by

$$E^I_C(\Lambda) = E^I_D(\Lambda) + E^I_N(\Lambda) - \sum_{d=a,b,c} E_1(d, \Lambda), \tag{8.23}$$

where $E_1(d,\Lambda) = d\Lambda^2/2\pi - \zeta(2)/(4\pi d)$ is the energy of a scalar field in 1-dimensions obeying Dirichlet boundary conditions in a region of length d. The contribution from Region II follows from replacing $a \to h - a$. Combining previous results, the

electromagnetic Casimir force is found to be,

$$F_C = -\frac{3\zeta(4)}{8\pi^2}\frac{A}{a^4} + \frac{\zeta(2)}{8\pi a^2} - \frac{J_C(b/c)}{32\pi^2 A} + \frac{\pi^2}{16}\frac{A}{a^4}\sum_{m,n}{}' \frac{\coth(f_{mn}(b/a, c/a))}{f_{mn}(b/a, c/a)\sinh^2(f_{mn}(b/a, c/a))}, \tag{8.24}$$

where $J_C(x) \equiv J_{-1}(x) + J_{+1}(x) = 2Z_2(x^{1/2}, x^{-1/2}; 4)$.

8.3.2 Discussion

Here we address the implications of eqs. (8.22) and (8.24) for the force on the partition in more detail. To begin, we discuss the important issue of attraction versus repulsion. We are interested in comparing the force on the partition (F_Γ, where $\Gamma = D, N$ or C for Dirichlet, Neumann, or EM boundary conditions respectively) to the force reported in the literature for a single cavity, which we denote $F_{\Gamma,\text{box}}$ [17]. The latter is obtained by the following prescription: calculate the energy in a single rectilinear cavity, drop the cutoff dependent ("divergent") terms, ignore contributions from the region exterior to the cavity, and differentiate with respect to a to obtain a force. We emphasize that there is no justification for dropping the cutoff dependent terms, so although we refer to this result, for convenience, as F_{box}, it does not apply to the physical case of a rectilinear box.

For the piston geometry, we note that the *sole* contribution from Region II is the a–independent term denoted by J. In fact this is *the only term that distinguishes* F from F_{box}, *i.e.*,

$$F_\Gamma = F_{\Gamma,\text{box}} - J_\Gamma(b/c)/(32\pi^2 A). \tag{8.25}$$

Naively, the difference by a constant may not seem important. Indeed it is not too important for small values of the ratio $a/(b, c)$. However it is very important for $a \gtrsim (b, c)$. In Fig. 8-2 we plot both these forces for a square cross section ($b = c$) as a function of a/b. (The plots include scalar as well as EM cases.) Note that in all cases $F \to 0$, while $F_{\text{box}} \to J(1)/(32\pi^2 A)$ (a constant) as $a/b \to \infty$. For this geometry $J_D(1) \approx -1.5259$, $J_N(1) \approx 13.579$, and $J_C(1) \approx 12.053$, so J is negative for Dirichlet and positive for both Neumann and EM. We see that F is always attractive, while F_{box} can change sign. It is always attractive for Dirichlet, but becomes repulsive for Neumann when $a/b > 1.745$ and for EM when $a/b > 0.785$. This is the consequence of ignoring Region II and the cutoff dependence. Indeed, it is easy to show that the piston force is attractive for any choice of a, b, c, h. A final comment is that for h finite and $a = h/2$, the partition sits at an unstable equilibrium position. This comment was made in Ref. [28], although the above detailed results were not derived there.

With the explicit form for F, we can more closely compare the piston with Casimir's original parallel plate geometry. For better comparison in Figs. 8-3 and 8-4, we have plotted the forces for the scalar and EM fields, after dividing by the parallel plates results, $F_{(D,N)\|} = -3\zeta(4)A/(16\pi^2 a^4)$ or $F_{C\|} = -3\zeta(4)A/(8\pi^2 a^4)$. First, note that for the EM case not only does $F_C \to 0$ as $a/b \to \infty$ but it does so

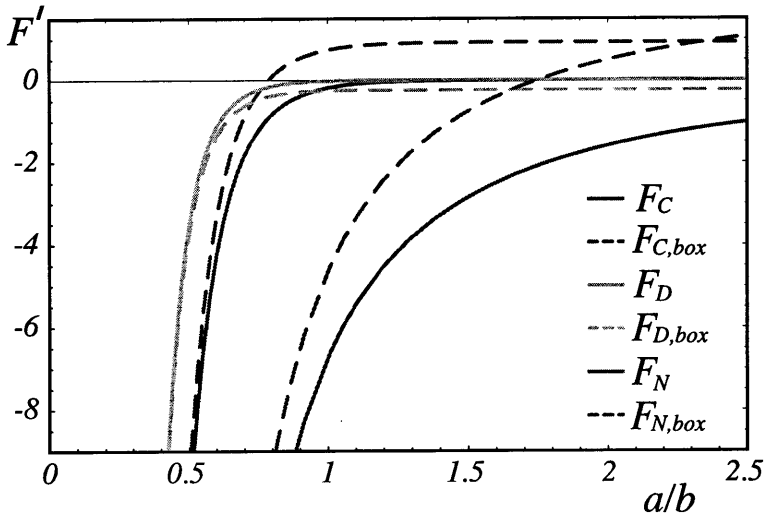

Figure 8-2: The force F on a square piston ($b = c$) due to quantum fluctuations of a field subject to Dirichlet, Neumann, or conducting boundary conditions, as a function of a/b, rescaled as $F' \equiv 16\pi^2 AF/(3\zeta(4))$ ($F' \equiv 8\pi^2 AF/(3\zeta(4))$) for scalar (EM) fields. The solid lines are for the piston, solid middle = F_C, solid upper = F_D, and solid lower = F_N, while their dashed counterparts are for the box.

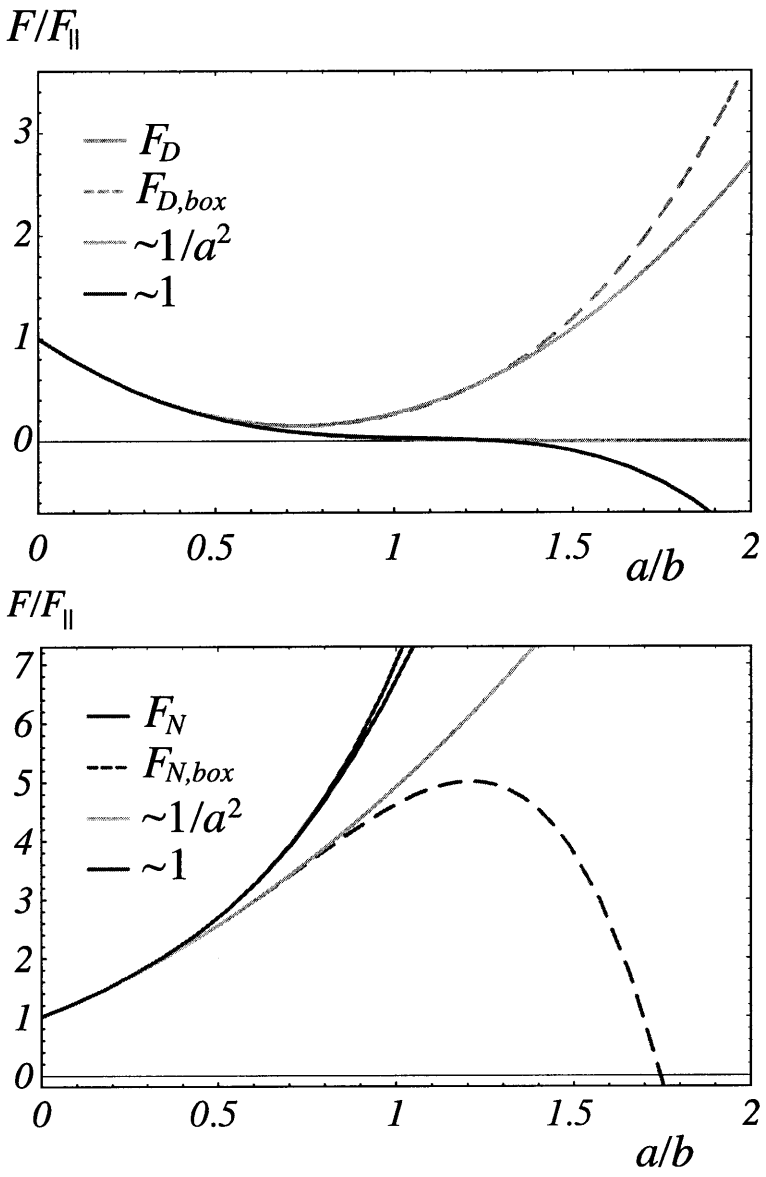

Figure 8-3: The force F on a square partition ($b = c$) due to quantum fluctuations of a scalar field as a function of a/b, normalized to the parallel plates force $F_{\|}$. Left figure is Dirichlet; solid middle = F_D (piston), dashed = $F_{D,\text{box}}$ (box), solid upper = $\{1/a^4, 1/a^3, 1/a^2\}$ terms, solid lower = $\{1/a^4, 1/a^3, 1/a^2, 1\}$ terms. Right figure is Neumann; solid middle = F_N (piston), dashed = $F_{N,\text{box}}$ (box), solid lower = $\{1/a^4, 1/a^3, 1/a^2\}$ terms, solid upper = $\{1/a^4, 1/a^3, 1/a^2, 1\}$ terms.

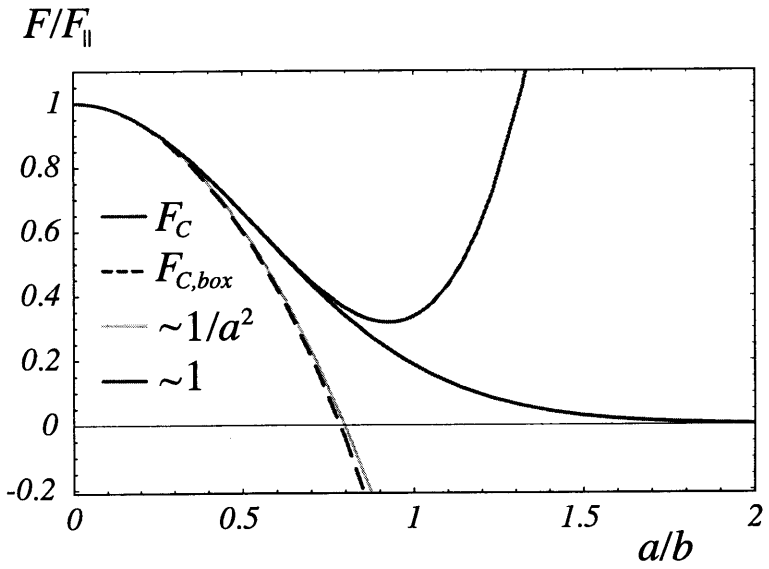

Figure 8-4: The force F on a square partition ($b = c$) due to quantum fluctuations of the EM field as a function of a/b, normalized to the parallel plates force $F_{\|}$. Solid middle = F_C (piston), dashed = $F_{C,\mathrm{box}}$ (box), solid lower = $\{1/a^4, 1/a^2\}$ terms, solid upper = $\{1/a^4, 1/a^2, 1\}$ terms.

rather quickly. Since $F_{C\|}$ vanishes as $1/a^4$, it is clear from Fig. 8-4 that F_C vanishes even more rapidly. In fact it vanishes exponentially fast, as $e^{-2\pi a/b}$ for $a \gg b$. We can understand this as follows: In this limit the most important paths are those that reflect off the top and bottom plates once, and therefore travel a distance $2a$. The transverse wavenumber $k = \pi/b$ due to the finite cross section, acts as an effective mass for the system, and damps the contribution of these paths exponentially. In fact for any rectangular cross section we find

$$F_C \approx -\frac{\pi}{2}\left(\frac{1}{\sqrt{ab^3}}e^{-2\pi a/b} + \frac{1}{\sqrt{ac^3}}e^{-2\pi a/c}\right), \quad \text{as } a \to \infty. \tag{8.26}$$

Experimentally, values of $a/b \sim 1$ are not yet realizable. Instead, typical experimental studies of Casimir forces have transverse dimensions that are roughly 100 times the separation between the "plates". This means that the leading order corrections to $F_{C\|}$ are more likely to be detected experimentally. In Figs. 8-3 and 8-4 we show the result of including successive corrections to $F_\|$ for scalar and EM cases, respectively; we plot the curve which includes $\{1/a^4, 1/a^3, 1/a^2\}$ terms and another curve that includes $\{1/a^4, 1/a^3, 1/a^2, 1\}$ terms. In the EM case we note that the $1/a^3$ term that appears in the expansion for Dirichlet and Neumann boundary conditions is canceled. In the next Section, we will demonstrate that this is a general phenomenon for any cross section (see ahead to eq. (8.37)). Hence the first correction to the EM result scales as $1/a^2$, which is $\mathcal{O}(a^2/A)$ compared to $F_{C\|}$. We see that this correction is quite accurate up to $a/b \sim 0.3$. We suspect this regime of accuracy to be roughly valid for any cross section.

8.4 General Cross Sections

The piston for arbitrary cross section cannot be solved exactly, but we can obtain much useful information from an asymptotic expansion for small separation a. The generalized piston maintains symmetry along the vertical axis, and its geometry is the product of $I \otimes \mathcal{S}$ of the interval $I = [0, a]$ and some 2-dimensional cross section $\mathcal{S} \subset \mathbb{R}^2$. Let us denote by $\mathcal{E} = k^2$ the eigenvalues of the Laplacian on the piston base $\mathcal{S}$ and the interval I, separately, with appropriate boundary conditions

$$-\Delta_{\mathcal{S},I}\psi_{\mathcal{S},I} = \mathcal{E}\psi_{\mathcal{S},I}. \tag{8.27}$$

The corresponding densities of states are denoted by by ρ_I and $\rho_\mathcal{S}$, respectively. Then the density of states (per unit "energy", $\mathcal{E}$) of the problem in the 3-dimensional region $I \otimes \mathcal{S}$ is $\rho(\mathcal{E})$ and can be written as the convolution

$$\rho(\mathcal{E}) = \int_0^\infty d\mathcal{E}' \rho_\mathcal{S}(\mathcal{E} - \mathcal{E}')\rho_I(\mathcal{E}'). \tag{8.28}$$

The 2-dimensional density $\rho_\mathcal{S}$ is not known in general, since the wave equation can not be solved in full generality in an arbitrary domain $\mathcal{S}$. However, for small height to width ratios, the smallness of a translates to high energies $\mathcal{E}$, and we will see that

the asymptotic behavior of $\rho_{\mathcal{S}}$ is sufficient for extracting an asymptotic expansion for the force is powers of $1/a$.

The number of eigenstates with energy less than $\mathcal{E}$ in $\mathcal{S}$ has the asymptotic expansion at large $\mathcal{E}$ [29],

$$N_{\mathcal{S}}(\mathcal{E}) = \left(\frac{A}{4\pi}\mathcal{E} + \eta\frac{P}{4\pi}\sqrt{\mathcal{E}} + \chi + r_N(\mathcal{E})\right)\Theta(\mathcal{E}). \tag{8.29}$$

Here, χ is related to the shape of the domain $\mathcal{S}$ through

$$\chi = \sum_i \frac{1}{24}\left(\frac{\pi}{\alpha_i} - \frac{\alpha_i}{\pi}\right) + \sum_j \frac{1}{12\pi}\int_{\gamma_j}\kappa(\gamma_j)d\gamma_j, \tag{8.30}$$

where α_i is the interior angle of each sharp corner and $\kappa(\gamma_j)$ is the curvature of each smooth section. It is easy to check that $\chi = 1/4$ for a rectangle and $\chi = 1/6$ for any smooth shape (for example a circle). Note that we have included the step function $\Theta(\mathcal{E})$ in the expression for $N_{\mathcal{S}}(\mathcal{E})$, ensuring that only $\mathcal{E} > 0$ contributes. Here $r_N(\mathcal{E})$ is a function which designates lower order terms (remainder) in an $\mathcal{E} \to \infty$ asymptotic expansion. For any polygonal shape r_N is exponentially small, $r_N(\mathcal{E}) = \mathcal{O}(e^{-c\mathcal{E}})$ ($c > 0$ is a constant) [30]. However, we are aware of only a much weaker estimate for smooth shapes, as $r_N(\mathcal{E}) = \mathcal{O}(1/\sqrt{\mathcal{E}})$ [29]. The derivative of $N_{\mathcal{S}}(\mathcal{E})$ is the density of states[3]

$$\rho_{\mathcal{S}}(\mathcal{E}) = \left(\frac{A}{4\pi} + \eta\frac{P}{8\pi}\frac{1}{\sqrt{\mathcal{E}}}\right)\Theta(\mathcal{E}) + \chi\,\delta(\mathcal{E}) + r_\rho(\mathcal{E}), \tag{8.31}$$

where we have used $\Theta'(\mathcal{E}) = \delta(\mathcal{E})$, and $\mathcal{E}\,\delta(\mathcal{E}) = \sqrt{\mathcal{E}}\,\delta(\mathcal{E}) = 0$ for all $\mathcal{E}$.

The other function in the convolution, the 1-dimensional density of states, is known exactly: it is simply a sum over delta functions, which we rewrite in terms of its Poisson summation

$$\rho_I(\mathcal{E}) = \sum_{n=1}^{\infty}\left(\mathcal{E} - \frac{n^2\pi^2}{a^2}\right) \tag{8.32}$$

$$= \frac{a}{2\pi}\frac{\Theta(\mathcal{E})}{\sqrt{\mathcal{E}}} + 2\sum_{m=1}^{\infty}\int_0^{\infty} dx\cos(2\pi m x)\,\delta\left(\mathcal{E} - \frac{x^2\pi^2}{a^2}\right). \tag{8.33}$$

The first term in eq. (8.33) is the smooth contribution to the density of states, and the second is the oscillatory component. The leading contributions (as $a \to 0$) to the 3-dimensional density of states come from convolving the smooth part of $\rho_{\mathcal{S}}$ with ρ_I,

[3] In Eqs. (8.31) and (8.34) we have denoted the various remainders by $r_\rho(\mathcal{E}), r_1(\mathcal{E}), r_2(\mathcal{E})$. We discuss the size of the remainders in the expansion of the forces F_η & F_C following eq. (8.37).

giving[3]

$$\begin{aligned}\rho(\mathcal{E}) &= a\left(\frac{1}{4\pi^2}A\sqrt{\mathcal{E}}+\frac{\eta}{16\pi}P+\frac{1}{2\pi}\frac{\chi}{\sqrt{\mathcal{E}}}+r_1(\mathcal{E})\right)\\ &+\sum_{m=1}^{\infty}\Big(\frac{A}{4\pi^2 m}\sin(2ma\sqrt{\mathcal{E}})+\eta\frac{aP}{8\pi}J_0(2ma\sqrt{\mathcal{E}})\\ &+\frac{a\chi}{\pi\sqrt{\mathcal{E}}}\cos(2ma\sqrt{\mathcal{E}})+r_2(\mathcal{E})\Big).\end{aligned} \tag{8.34}$$

The first line agrees precisely with the Balian and Bloch theory of the density of states [23], and gives the cutoff dependent terms in the Casimir energy

$$E(\Lambda)=\frac{1}{2}\int_0^\infty d\mathcal{E}\rho(\mathcal{E})\sqrt{\mathcal{E}}\,e^{-\sqrt{\mathcal{E}}/\Lambda}. \tag{8.35}$$

The cutoff dependent contributions have no effect on the Casimir force when Region II is included, since they are linear in a, as explained earlier. The second line in eq. (8.34) gives the leading three terms in an asymptotic expansion of the force

$$F_\eta=-\frac{3\zeta(4)}{16\pi^2a^4}A-\eta\frac{\zeta(3)}{32\pi a^3}P-\frac{\zeta(2)\chi}{4\pi a^2}+r_\eta(a). \tag{8.36}$$

Also, even for these general cross sections, the EM energy can be related using eq. (8.12) to Dirichlet and Neumann energies, as

$$F_C=-\frac{3\zeta(4)}{8\pi^2a^4}A+\frac{\zeta(2)(1-2\chi)}{4\pi a^2}+r_C(a). \tag{8.37}$$

In eqs. (8.36) and (8.37) we have written the remainder terms as $r_\eta(a)$ and $r_C(a)$ $(=r_{-1}(a)+r_{+1}(a))$, respectively. Following our earlier estimates for $r_N(\mathcal{E})$ that appears in $N_S(\mathcal{E})$, and noting that there is always an $\mathcal{O}(1)$ term that comes from Region II, we have $r_{\eta,C}(a)=\mathcal{O}(1)$ for polygonal shapes and $r_{\eta,C}(a)=\mathcal{O}(1/a)$ for smooth shapes, as $a\to 0$.

The generalization to arbitrary cross sections in eq. (8.37) has interesting features. The correction to the parallel plates result depends on geometry through the parameter χ, which depends sensitively on whether the cross section is smooth or has sharp corners. For example, $\chi=\frac{1}{6}$ for all smooth shapes and $\chi=\frac{1}{6}\frac{n-1}{n-2}$ for an n-sided polygon of equal interior angles. Given unavoidable imperfections in any experimental realization, one may wonder what precisely constitutes "smooth" or "sharp." Note that for any deformation with local radius of curvature R ($R=0$ for perfectly sharp corners), we have the dimensionless quantity R/a, where a is the base-partition height. Given that our expansion is valid for small a, we conclude that $R\gg a$ is a smooth deformation, while $R\ll a$ can be regarded as a sharp corner. As a simple example, consider a shape that is roughly square (4-sided polygon) if viewed from large distances, but is in fact rounded with radius R at the "corners" if examined closely. Let us also imagine that the overall width (b) is much larger

than R. Then, since the corresponding term in the Casimir force goes as $1 - 2\chi$ (see eq. (8.37)), we expect the correction to be $\zeta(2)/(6\pi a^2)$ for $a/R \ll 1$ and decrease to $\zeta(2)/(8\pi a^2)$ for $a/R \gg 1$. A more interesting example would be a self-similar (fractal or self-affine) perimeter, in which the number of sharp corners deceases as a power of the resolution a. For such a case, we expect a correction scaling as a non-trivial power of $1/a$, reminiscent of results in Ref. [31]. It would be interesting to see if such corrections are experimentally accessible.

Another noteworthy feature of eq. (8.37) is that the leading correction to the EM force (compared to parallel plates) is smaller by order of a^2/A. By contrast the corrections are only of order $a/\sqrt{A}$ for scalar fields with either Dirichlet or Neumann boundary conditions. However, the latter corrections are exactly equal and opposite in sign, and cancel for the EM force. It is interesting to inquire if this precise cancellation applies only to perfect metallic boundary conditions, or remains when the effects of finite conductivity are taken into account. More work is necessary to understand the finite conducting piston. Yet another case is for side walls made of dielectrics, where a simple modification of the optical path method, which replaces the sign factor η with the reflection coefficients for TM and TE modes, suggests that the cancellation does not occur. A piston that is made entirely of a uniform dielectric is examined in Ref. [32]

8.5 Thermal Corrections

The question of the leading corrections to the Casimir force at finite temperatures T has generated recent interest, both from the practical need to evaluate the accuracy of experiments, and due to fundamental issues. In particular, there is controversy pertaining to the appropriate model for the metallic walls, which we shall ignore in this chapter. Instead, we shall compute corrections to the Casimir force due to finite temperature excitations of the modes in the piston, while continuing to treat its walls as perfect metals [33].

8.5.1 Rectangular Piston

We first answer this question for the piston with rectangular cross section. In units with $\hbar = c = k_B = 1$, the inverse temperature $\beta = 1/T$ introduces a new length scale whose size relative to the dimensions a, b, and c of the piston (we imagine, as earlier, that $h \to \infty$) sets the importance of thermal corrections. (More precisely, $\pi\beta$ is the appropriate length scale.) In typical experiments $a \sim 1\mu m$, $b, c \sim 100\mu m$, and at room temperature $\pi\beta \sim 20\mu m$. Thus the regime of most experimental interest is where the length scales satisfy $a \ll \pi\beta \lesssim b,\ c$. In light of this we focus on thermal lengths much larger than the base–partition height, *i.e.* $a \ll \pi\beta$. To fully investigate the low temperature regime, we assume $a \ll \pi\beta, b, c \ll h$, but will allow $\pi\beta$ to be less than or greater than b or c.

Each mode of the field can be regarded as an independent harmonic oscillator,

and by summing the corresponding contributions, we find the free energy

$$\mathcal{F}_{\rm tot} = \frac{-1}{\beta}\sum_m \ln\left(\frac{e^{-\frac{1}{2}\beta\omega_m}}{1-e^{-\beta\omega_m}}\right) = E + \delta\mathcal{F}. \tag{8.38}$$

We have separated out the the zero-temperature Casimir energy E, from the finite temperature "correction" $\delta\mathcal{F} = \delta E - T\delta S$, and focus on the latter for calculating finite temperature effects.

First, a note of caution is in order regarding the scalar field with Neumann boundary conditions. In *any cavity*, there is a trivial solution to the Neumann problem, namely a constant field with $\omega = 0$. This means that whenever β is finite ($T > 0$) then $\delta\mathcal{F} = -\infty$, which signals condensation of the scalar field into the ground state. We note that this phenomenon occurs for *closed* geometries where the spectrum is discrete and not in general for *open* geometries in which the region near $\omega = 0$ is integrable due to phase space suppression. We will proceed by calculating the free energy of a scalar field with both Dirichlet and Neumann boundary conditions, ignoring the mode with $\omega = 0$ for the latter. We then use eq. (8.12) to obtain the EM force. This procedure is valid since the offending Neumann mode is specifically excluded from the EM spectrum.

For a Dirichlet scalar field in Region I, since all modes satisfy $\omega_m > \pi/a$, their Boltzmann weights are small in the limit of $a \ll \pi\beta$, and

$$\delta\mathcal{F}^I = \mathcal{O}\big(e^{-\pi\beta/a}\big) \tag{8.39}$$

is exponentially small. Similarly, the a-dependent terms of the *electromagnetic* free energy in region I are exponentially small. This is true for any cross section and reflects the fact that thermal wavelengths $\sim \pi\beta$ are *excluded* from Region I [34]. However, a power law contribution to the free energy and force will come from Region II. We use the optical expansion, which remains exact for the free energy in rectilinear geometries, to compute this contribution for scalar fields [34], as

$$\delta\mathcal{F}^{II} = -\frac{1}{2\pi^2}\sum_{p_r}\phi(p_r,\eta)\sum_q{}' \int_{\mathcal{D}} d\mathbf{x}\frac{1}{\left[l_{p_r}(\mathbf{x})^2 + (q\beta)^2\right]^2}. \tag{8.40}$$

Note that here the sum ranges over $q \in \mathbb{Z}\setminus\{0\}$ — the $q = 0$ term is just the Casimir energy (see eq. (8.7)).

It is natural to break the energy up into the familiar four classes of paths: periodic orbits, aperiodic tours off faces, reflections off edges, and reflections off corners. However, summing each set separately gives a logarithmic divergence (that cancels among the different classes for Dirichlet boundary conditions). Fortunately, this problem can be ignored in the $h \to \infty$ limit, as can be seen, for example, by considering the contribution from the sum over periodic orbits (paths (c), (d), etc in Fig. 8-1). Noting

that $h-a$ is the height of the piston in Region II, we have

$$\delta\mathcal{F}^{II}_{\rm per} = -\frac{1}{16\pi^2}\sum_{q=1}^{\infty}\sum_{n,m,l=-\infty}^{\infty}\frac{(h-a)bc}{[(n(h-a))^2+(mb)^2+(lc)^2+(q\beta/2)^2]^2}. \tag{8.41}$$

This expression is logarithmically divergent, but if we take $h\to\infty$, only the $n=0$ term contributes and the remaining summation over $\{q,m,l\}$ is finite. Strictly speaking, the interchange of the limit $h\to\infty$ with the summations, which eliminates the logarithmic divergence, is formally problematic. However a more rigorous analysis justifies this step for the Dirichlet case through the cancellation among the different classes, but always gives $-\infty$ for the Neumann case as anticipated. Performing this interchange gives the following result for the contribution of periodic orbits

$$\begin{aligned}\delta\mathcal{F}^{II}_{\rm per} = &-\frac{\zeta(4)(V_p-Aa)}{\pi^2\beta^4}-\frac{(h-a)A}{32\pi\beta}Z_2(b,c;3)\\ &+\frac{(h-a)A}{32\pi^2}Z_2(b,c;4)+\mathcal{O}\left(e^{-4\pi g/\beta}\right)\end{aligned} \tag{8.42}$$

with $g\equiv\min(b,c)$ and V_p as the total piston volume. Here we have expanded for small β relative to $g=\min(b,c)$. We note that although the third term is independent of β, this really is part of $\delta\mathcal{F}$. The reader that is interested in the opposite limit of $\beta\to\infty$, i.e., the low temperature limit, should look ahead to Section 8.5.3.

Proceeding in a similar fashion for all contributions to the free energy of a scalar field we find (ignoring the exponentially small contribution of Region I)

$$\begin{aligned}\delta\mathcal{F}_\eta = &-\frac{\zeta(4)(V_p-Aa)}{\pi^2\beta^4}-\eta\frac{\zeta(3)(S_p-Pa)}{8\pi\beta^3}-\frac{\zeta(2)(h-a)}{4\pi\beta^2}-\frac{M_\eta(b/c)(h-a)}{32\pi\beta\sqrt{A}}\\ &+\frac{J_\eta(b/c)(h-a)}{32\pi^2A}-\frac{\pi^2(V_p-Aa)}{8\beta^4}{\sum_{m,n}}'\frac{1+2f_{mn}(\bar b,\bar c)-e^{-f_{mn}(\bar b,\bar c)}}{f^3_{mn}(\bar b,\bar c)\sinh^2(f_{mn}(\bar b,\bar c))}\\ &-\eta\frac{(h-a)}{\beta^2}\sum_{m,n=1}^{\infty}\frac{n}{m}\left(K_1(2\pi mn\bar b)+K_1(2\pi mn\bar c)\right)\end{aligned} \tag{8.43}$$

where we have defined $M_\eta(x)\equiv Z_2(x^{3/2},x^{-3/2};3)+4\eta(x^{1/2}+x^{-1/2})\zeta(2)$, $\bar b\equiv 2b/\beta, \bar c\equiv 2c/\beta$, and S_p is the total surface area of the piston. It is important to note that while $\delta\mathcal{F}_{-1}=\delta\mathcal{F}_D$, $\delta\mathcal{F}_{+1}=\delta\mathcal{F}_N$ is not strictly correct as we have ignored the $\omega=0$ Neumann mode. Although $\delta\mathcal{F}_N=-\infty$, as stated earlier, this expression correctly gives the a-dependence in $\delta\mathcal{F}_N$.

The EM case can be handled in a similar fashion. Repeating our earlier decomposition, we note that $\delta\mathcal{F}_{\rm EM}=\delta\mathcal{F}_{-1}+\delta\mathcal{F}_{+1}+\zeta(2)(h-a)/\beta^2$, since the spectral decomposition in eq. (8.11) *correctly* leaves out the $\omega=0$ mode of the Neumann spectrum. We thus find (again ignoring the exponentially small contribution of Region

I)

$$\delta\mathcal{F}_{\rm EM} = -\frac{2\zeta(4)(V_p - Aa)}{\pi^2\beta^4} + \frac{\zeta(2)(h-a)}{2\pi\beta^2} - \frac{M_C(b/c)(h-a)}{32\pi\beta\sqrt{A}} + \frac{J_C(b/c)(h-a)}{32\pi^2 A}$$
$$- \frac{\pi^2(V_p - Aa)}{4\beta^4}\sum_{m,n}{}' \frac{1 + 2f_{mn}(\bar{b},\bar{c}) - e^{-2f_{mn}(\bar{b},\bar{c})}}{f_{mn}^3(\bar{b},\bar{c})\sinh^2(f_{mn}(\bar{b},\bar{c}))} \tag{8.44}$$

where $M_C(x) \equiv M_{-1}(x) + M_{+1}(x) = 2Z_2(x^{3/2}, x^{-3/2}; 3)$.

In Eqs. (8.43) and (8.44) we have written the expansion as a series in increasing powers of β. The result, though, is correct (up to exponentially small terms in $a/\pi\beta$) for any ratio of β to b or c, and for h much larger than any of the other scales. The infinite summations that appear are convergent for all finite values of $\{\beta, b, c\}$. The leading term in eq. (8.44) is the Stefan–Boltzmann energy, and the following terms are corrections due to geometry. The term independent of β is equal to but opposite in sign to that appearing in the Casimir energy. Note that the appearance of a term independent of β is an artifact of performing a small β expansion. All terms depend linearly on a and provide a constant force on the partition. Note that the first five terms in $\delta\mathcal{F}_\eta$ and the first four terms in $\delta\mathcal{F}_{\rm EM}$ have power law dependences on β, while the remaining terms (summations) are exponentially small for $\pi\beta < (b, c)$.

8.5.2 General Cross Section

If we consider general $I \otimes \mathcal{S}$ geometries, as in Section 8.4, we may use the smooth 3-dimensional Balian and Bloch density of states in Region II to obtain the leading terms in the free energy. Specifically, we use the first line of eq. (8.34) with the replacement $a \to h - a$ for $\rho(\mathcal{E})$, and calculate the free energy from

$$\delta\mathcal{F} = \frac{1}{\beta}\int_0^\infty d\mathcal{E}\,\rho(\mathcal{E})\ln(1 - \exp(-\beta\sqrt{\mathcal{E}})). \tag{8.45}$$

Since we only know the first three terms in the expansion for the density of states, we will obtain contributions proportional to the volume, surface, and perimeter of the piston, but nothing at $\mathcal{O}(1/\beta)$. It is fairly straightforward to get

$$\delta\mathcal{F}_D = -\frac{\zeta(4)(V_p - Aa)}{\pi^2\beta^4} + \frac{\zeta(3)(S_p - Pa)}{8\pi\beta^3} - \frac{\zeta(2)\chi(h-a)}{\pi\beta^2} + \mathcal{O}\left(\frac{1}{\beta}\right), \tag{8.46}$$

$$\delta\mathcal{F}_{\rm EM} = -\frac{2\zeta(4)(V_p - Aa)}{\pi^2\beta^4} + \frac{\zeta(2)(1-2\chi)(h-a)}{\pi\beta^2} + \mathcal{O}\left(\frac{1}{\beta}\right). \tag{8.47}$$

We again see the effect of the modes excluded from Region I due to $a \ll \pi\beta$, in the factors $V_p - Aa$, $S_p - Pa$, and $h - a$. These leading terms provide thermal contributions to the quantum force on the partition, given in eqs. (8.36) and (8.37).

Let us comment on the relationship between the Casimir and thermal contributions to the force. We begin by focusing on the regime that is perhaps of most experimental interest: $a \ll \pi\beta \ll \sqrt{A}$. If we include both Casimir and thermal

contributions to the force, as given in eqs. (8.37) & (8.47),

$$F_{\rm EM} = -\frac{3\zeta(4)}{8\pi^2}\left(\frac{1}{a^4}+\frac{1}{(\beta/2)^4}\right)A+\frac{\zeta(2)(1-2\chi)}{4\pi}\left(\frac{1}{a^2}+\frac{1}{(\beta/2)^2}\right)+\cdots \quad (8.48)$$

Note that the leading contributions are related to terms in the Casimir energy by the interchange $a \leftrightarrow \beta/2$, but this connection ceases for higher order corrections. We have only calculated further terms for the parallelepiped and we can compare them in this limit. In particular, eq. (8.44) includes a contribution of order $1/\beta$ which has no counterpart (*i.e.* a term of order $1/a$) in the Casimir force. A term of order $1/a$ can only come from the derivative of $\sim \ln a$, which is absent from the EM Casimir energy.

8.5.3 Low Temperature Limit

Equation (8.48) provides the leading terms in the Casimir force in the limit $a \ll \pi\beta \ll b, c$ (or more generally $a \ll \pi\beta \ll \sqrt{A}$ for non-rectangular cross sections). We may more accurately refer to this as a "medium temperature" regime, as opposite to a lower temperature regime with $\pi\beta \gg \sqrt{A}$. In fact, for the rectangular piston we obtained in eqs. (8.43) and (8.44) results that are valid for $a \ll \{\pi\beta, b, c\}$ for any ratio of β to b or c, and will now examine their lower temperature limit. A naive application of the proximity-force approximation gives always a thermal correction to the force that vanishes as $\sim 1/\beta^4 = T^4$ in the $T \to 0$ limit [16]. However, in Ref. [34] it is argued that for open geometries this limit is quite subtle and is sensitive to the detailed shape of each surface. In fact it is reasonable to argue that for the cases relevant to experiments there may be weaker power laws, *i.e.*, $1/\beta^\alpha$ with $\alpha < 4$. But in our closed geometry another scenario is natural: If $T \to 0$, so that $\beta \gg a, \sqrt{A}$, modes are excluded from *both* regions due to a gap in the spectrum, resulting in an exponentially small free energy, which (for the rectangular piston) is

$$\delta\mathcal{F}_{\rm EM} = -\frac{(h-a)}{\sqrt{2}\beta^{3/2}}\left(\frac{1}{\sqrt{b}}e^{-\pi\beta/b}+\frac{1}{\sqrt{c}}e^{-\pi\beta/c}\right), \quad \text{as } \beta\to\infty. \quad (8.49)$$

A plot of the force $F_{T\Gamma} \equiv -d\delta\mathcal{F}_\Gamma/da$ (where $\Gamma = D$ or C as for $T = 0$), derived from eqs. (8.43) and (8.44) is given in Fig. 8-5. The force is normalized to the Stefan–Boltzmann term, $F_{SB} = -\zeta(4)A/(\pi^2\beta^4)$ $(-2\zeta(4)A/(\pi^2\beta^4))$ for Dirichlet (EM) fields. Having taken $a \ll \pi\beta$ in our analysis, the a dependence is ignorable, and we plot the force as a function of β/b ($b = c$). The high β/b asymptotic curves (eq. (8.49) is the EM case) are also included. Note that from eq. (8.48) we can read off the small $\beta/\sqrt{A}$ corrections to F_{SB} for arbitrary cross sections.

8.6 Conclusions

In this work we have obtained an exact, analytic result for the Casimir force for a piston geometry. Exact, analytic results are rare in this field but nonetheless partic-

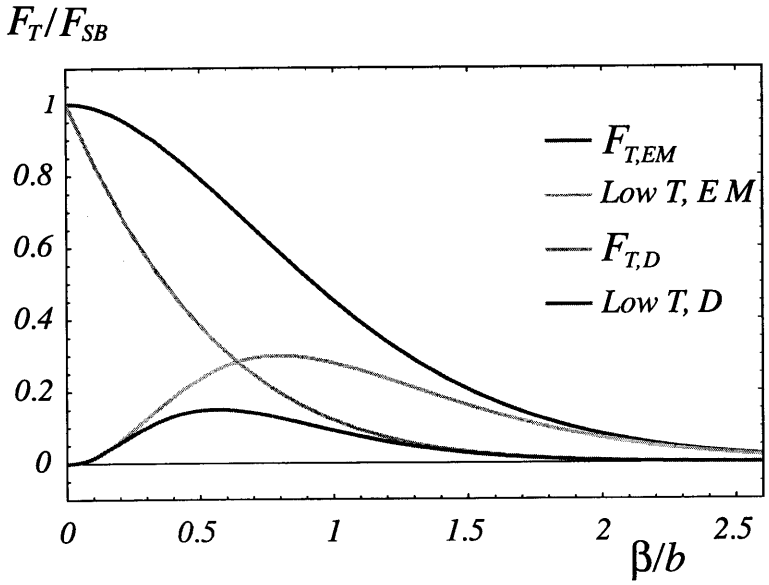

Figure 8-5: The force F_T from thermal fluctuations on a square partition ($b = c$), normalized to the Stefan–Boltzmann expression $F_{SB} = -\zeta(4)A/(\pi^2\beta^4)$ $(-2\zeta(4)A/(\pi^2\beta^4))$ for Dirichlet (EM) fields, as a function of β/b. This is valid in the regime: $a \ll \{\pi\beta, b, c\}$. Starting from a normalized value of 1, the full result for Dirichlet (electromagnetic) is the lower (upper) curve. Also, starting from a normalized value of 0, the exponentially small asymptote (as $\beta/b \to \infty$) for Dirichlet (electromagnetic) is the lower (upper) curve.

ularly useful for comparison with the approximations needed to describe real systems and more complicated geometries.

We have obtained analytic expressions for the force acting on the partition in a piston with perfect metallic boundaries. The results are exact for the rectangular piston, and in the form of an asymptotic series in $1/a$ for arbitrary cross section. We find that the partition is always attracted to the (closer) base; consistent with a more general result obtained in Ref. [18]. Since the piston geometry is closely related to single cavity for which a repulsive force has been conjectured, we are able to shed some light on this question. In particular, we emphasize that to avoid unphysical deformations (and closely related issues on cutoffs and divergences) it is essential to examine contributions to the force from both sides of the partition. The cutoff independent contribution from a single cavity (that we call F_{box}) approaches a constant for large a. However, in the piston geometry compensating contributions from the second cavity cancel both the cutoff dependent terms and part of the cutoff independent term, to cause a net attraction.

For general cross sections we find interesting dependence on geometrical features of the shape, such as its sharp corners and curved segments. We have obtained the first three terms for scalar fields and the first two terms for EM fields (one less due to cancellation) in an expansion in powers of a. By comparison to our calculated exact result for a rectangular cross section we estimate that this expansion is valid for $a/b \approx 0.3$. This covers the conventional experimentally accessible regime, and is therefore a useful result for a large class of geometries. We have also obtained thermal corrections which cover the experimentally accessible regime.

8.7 Appendix: Zeta Functions

The general Epstein Zeta function is defined as

$$Z_d(a_1,\ldots,a_d;s) \equiv \sum_{n_1,\ldots,n_d}{}' \left((n_1a_1)^2+\cdots+(n_da_d)^2\right)^{-s/2}, \tag{8.50}$$

where the summation is over $\{n_1,\ldots,n_d\} \in \mathbb{Z}^d \setminus \{0,\ldots,0\}$. Note that the Riemann Zeta function is a special case of this, namely $\zeta(s) = Z_1(1;s)/2$.

In eq. (8.15) we pointed out that such functions could be approximated by a term involving the Riemann zeta function and a power of a_1, as $a_1 \to 0$. An exact representation, as discussed in Ref. [17], is

$$\begin{aligned} Z_d(a_1,\ldots,a_d;s) &= \frac{2\zeta(s)}{a_1^s} + \frac{\Gamma\left(\frac{s-1}{2}\right)\sqrt{\pi}}{\Gamma\left(\frac{s}{2}\right)a_1} Z_{d-1}(a_2,\ldots,a_d;s-1) \\ &+ \frac{4\pi^{s/2}}{\Gamma\left(\frac{s}{2}\right)a_1^s} \sum_{n=1}^{\infty} \sum_{n_2,\ldots,n_d}{}' n^{(s-1)/2} K_{(s-1)/2}\left(2\pi n \frac{\sqrt{(a_2n_2)^2+\cdots+(a_dn_d)^2}}{a_1}\right) \\ &\times \left(\frac{\sqrt{(a_2n_2)^2+\cdots+(a_dn_d)^2}}{a_1}\right)^{(1-s)/2} \end{aligned} \tag{8.51}$$

where K_ν is the modified Bessel function of the second kind. This is useful in Region I where a_1 is small (with $a_1 \to a$).

For Region II it is important to examine the limit in which one of the lengths is infinite, say $a_1 \to \infty$ (with $a_1 \to h - a$). In this limit the order of the zeta function is reduced:

$$Z_d(a_1,\ldots,a_d;s) \to Z_{d-1}(a_2,\ldots,a_d;s). \tag{8.52}$$

Bibliography

[1] H. B. G. Casimir, "On the Attraction Between Two Perfectly Conducting Plates," K. Ned. Akad. Wet. Proc. **51**, 793 (1948).

[2] S. K. Lamoreaux, "Demonstration of the Casimir force in the 0.6 to 6 micrometers range," Phys. Rev. Lett. **78**, 5 (1997).

[3] R. Onofrio, "Casimir forces and non-Newtonian gravitation," New Journal of Physics **8**, 237 (2006) [arXiv:hep-ph/0612234].

[4] U. Mohideen and A. Roy, "Precision measurement of the Casimir force from 0.1 to 0.9 μm," Phys. Rev. Lett. **81**, 4549 (1998) [arXiv:physics/9805038]; A. Roy, C. Y. Lin and U. Mohideen, "Improved Precision Measurement of the Casimir Force," Phys. Rev. D **60**, 111101 (1999) [arXiv:quant-ph/9906062].

[5] H. B. Chan, V. A. Aksyuk, R. N. Kleiman, D. J. Bishop, and F. Capasso, "Quantum Mechanical Actuation of Microelectromechanical System by the Casimir Force," Science, **291**, 1941-1944 (2001);

[6] E. G. Adelberger, B. R. Heckel, and A. E. Nelson, "Tests of the gravitational inverse-square law," Ann. Rev. Nucl. Part. Sci. **53** 77-121 (2003) [arXiv:hep-ph/0307284]; S. R. Beane, "On the importance of testing gravity at distances less than 1-cm," Gen. Rel. Grav. **29**, 945-951 (1997) [arXiv:hep-ph/9702419].

[7] F. M. Serry, D. Walliser, and G. J. Maclay, "The Anharmonic Casimir Oscillator," J. Microelectromech. Syst. **4**, 193-205 (1995); F. M. Serry, D. Walliser, and G. J. Maclay, "The role of the casimir effect in the static deflection and stiction of membrane strips in microelectromechanical systems (MEMS)," J. Applied Phys. **84**, 2501-2506 (1998).

[8] O. Kenneth, I. Klich, A. Mann, and M. Revzen, "Repulsive Casimir forces," Phys. Rev. Lett. **89**, 033001 (2002) [arXiv:quant-ph/0202114].

[9] H. Gies and K. Langfeld, "Casimir effect on the worldline," JHEP, 0306, 018 (2003) [arXiv:hep-th/0303264].

[10] T. Emig, R. L. Jaffe, M. Kardar, and A. Scardicchio, "Casimir interaction between a plate and a cylinder," Phys. Rev. Lett. **96** 080403 (2006) [arXiv:cond-mat/0601055].

[11] D. A. R. Dalvit, F. C. Lombardo, F. D. Mazzitelli, R. Onofrio, "Casimir force between eccentric cylinders," Europhys. Lett. **67**, 4, 517-523 (2004) [arXiv:quant-ph/0406060]; D. A. R. Dalvit, F. C. Lombardo, F. D. Mazzitelli, and R. Onofrio, "Exact Casimir interaction between eccentric cylinders," Phys. Rev. A **74**, 020101(R) (2006) [arXiv:quant-ph/0608033].

[12] O. Schröder, A. Scardicchio, and R. L. Jaffe, "The Casimir Energy for a Hyperboloid Facing a Plate in the Optical Approximation," Phys. Rev. A **72**, 012105 (2005) [arXiv:hep-th/0412263].

[13] T. Emig, A. Hanke, R. Golestanian, and M. Kardar, "Probing the Strong Boundary Shape Dependence of the Casimir Force," Phys. Rev. Lett. **87**, 260402 (2001) [arXiv:cond-mat/0106028];

[14] R. Golestanian and M. Kardar, Phys. Rev. A **58**, 1713 (1998). "Path Integral Approach to the Dynamic Casimir Effect with Fluctuating Boundaries," [arXiv:quant-ph/9802017];

[15] T. H. Boyer, "Quantum electromagnetic zero point energy of a conducting spherical shell and the Casimir model for a charged particle," Phys. Rev. **174**, 1764 (1968).

[16] S. G. Mamaev and N. N. Trunov, "Dependence Of The Vacuum Expectation Values Of The Energy Momentum Tensor On The Geometry And Topology Of The Manifold," Theor. Math. Phys. **38**, 228 (1979) [Teor. Mat. Fiz. **38**, 345 (1979)]; W. Lukosz, Physica **56**, 109 (1971); V. M. Mostepanenko and N. N. Trunov, "The Casimir Effect And Its Applications," Sov. Phys. Usp. **31**, 965 (1988) [Usp. Fiz. Nauk **156**, 385 (1988)]; S. Hacyan, R. Jauregui, and C. Villarreal, Phys. Rev. A bf 47, 4204 (1993). G. J. Maclay, Phys. Rev. A **61**, 052110 (2000); M. Bordag, U. Mohideen and V. M. Mostepanenko, "New developments in the Casimir effect," Phys. Rept. **353**, 1 (2001) [arXiv:quant-ph/0106045].

[17] J. Ambjorn and S. Wolfram, "Properties Of The Vacuum. 1. Mechanical And Thermodynamic," Annals Phys. **147**, 1 (1983).

[18] O. Kenneth and I. Klich, "Opposites Attract - A Theorem About The Casimir Force," Phys. Rev. Lett **97**, 160401 (2006) [arXiv:quant-ph/0601011].

[19] R. M. Cavalcanti, "Casimir force on a piston," Phys. Rev. D **69**, 065015 (2004) [arXiv:quant-ph/0310184].

[20] M. P. Hertzberg, R. L. Jaffe, M. Kardar, and A. Scardicchio, "Attractive Casimir Forces in a Closed Geometry," Phys. Rev. Lett. **95**, 250402 (2005) [arXiv:quant-ph/0509071].

[21] V. N. Marachevsky, "Casimir energy of two plates inside a cylinder," (2006) [arXiv:hep-th/0609116]; A. Edery, "Casimir piston for massless scalar fields in three dimensions," Phys. Rev. D 75, 105012 (2007) [arXiv:hep-th/0610173].

[22] E. A. Power, "Introductory Quantum Electrodynamics," (Elsevier, New York, 1964), Appendix I.

[23] R. Balian and C. Bloch, "Distribution Of Eigenfrequencies For The Wave Equation In A Finite Domain. 1. Three-Dimensional Problem With Smooth Boundary Surface," Annals Phys. **60**, 401 (1970).

[24] J. Baacke and G. Krusemann, "Perturbative Analysis Of The Divergent Contributions To The Casimir Energy," Z. Phys. C **30**, 413 (1986); G. Barton, "Perturbative Casimir energies of dispersive spheres, cubes and cylinders," J. Phys. A **34**, 4083 (2001).

[25] N. Graham, R. L. Jaffe, V. Khemani, M. Quandt, O. Schroeder and H. Weigel, "The Dirichlet Casimir problem," Nucl. Phys. B **677**, 379 (2004) [arXiv:hep-th/0309130].

[26] R. L. Jaffe and A. Scardicchio, "The Casimir Effect and Geometric Optics," Phys. Rev. Lett. **92**, 070402 (2004) [arXiv:quant-ph/0310194]; A. Scardicchio and R. L. Jaffe, "Casimir Effects: An Optical Approach I. Foundations and Examples," Nucl. Phys. B **704**, 552 (2005) [arXiv:quant-ph/0406041].

[27] K. Gottfried and T. M. Yan, "Quantum Mechanics: Fundamentals," 2nd edn, Springer (2003).

[28] G. J. Maclay and J. Hammer, Proc. 7th Intl. Conf. on Squeezed States and Uncertainty Relations, (2001).

[29] H. P. Baltes and E. R. Hilf, "Spectra of Finite Systems," Bibliographisches Institut, Mannheim, (1976).

[30] P. B. Bailey and F. H, Brownell, J. Math. Anal. Appl., **4**, 212 (1962).

[31] H. Li and M. Kardar, "Fluctuation induced forces between manifolds immersed in correlated fluids," Phys. Rev. A **46**, 6490 (1992).

[32] G. Barton, "Casimir Piston And Cylinder, Perturbatively," Phys. Rev. D **73**, 065018 (2006).

[33] B. Geyer, G. L. Klimchitskaya, and V. M. Mostepanenko, "Surface-impedance approach solves problems with the thermal Casimir force between real metals," Phys. Rev. A **67**, 062102 (2003); M. Boström and Bo E. Sernelius Phys. Rev. Lett. **84**, 4757 (2000); J. R. Torgerson and S. K. Lamoreaux Phys. Rev. **E 70**, 047102 (2004).

[34] A. Scardicchio and R. L. Jaffe, "Casimir Effects: an Optical Approach II. Local Observables and Thermal Corrections," Nucl. Phys. B **743**, 249-275 (2006) [arXiv:quant-ph/0507042].