Topics in Theoretical Particle Physics and Cosmology

Thesis by

Michael P. Salem

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2007
(Defended May 21, 2007)

Acknowledgements

Much of this research would not have occurred without the hard work and insight of my collaborators. For this I thank Christian Bauer, Matthew Dorsten, Michael Graesser, Lawrence Hall, and Taizan Watari. I have been fortunate to have collaborators who were all willing to patiently explain ideas within and beyond the scope of our research. For this I am especially grateful to Christian, Michael, and Taizan, who have greatly increased my understanding of theoretical physics.

I am pleased to acknowledge my graduate adviser Mark Wise, who has provided invaluable guidance and support throughout my graduate studies. I have benefited much from his suggestions and his pedagogy, while his humor and congeniality helped make graduate school a very enjoyable experience. For his guidance and support I am also grateful to my undergraduate adviser Tanmay Vachaspati. His clever ideas initiated a successful collaboration on my undergraduate thesis, which set the stage for my subsequent progress. I am grateful that both of my advisers have always received me warmly and respectfully.

I have learned much from the other members of Mark's research group. For this I thank Lotty Ackerman, Moira Gresham, Alejandro Jenkins, Jennifer Kile, Sonny Mantry, Donal O'Connell, Sean Tulin, and Peng Wang. Lotty, Moira, and Donal have been especially patient and helpful in answering my questions. I have also benefited from interactions with Michael Ramsey-Musolf, Marc Kamionkowski, and members of Marc's research group, in particular Jonathan Pritchard and Kris Sigurdson. All of these people helped to create a pleasant and productive working environment. For this I am also grateful to my previous and present officemates, in particular Paul Cook and Jie Wang and especially Moira Gresham.

Through most of my graduate studies I have been financially supported primarily by a Caltech teaching assistantship. I am thankful to David Goodstein and especially David Politzer and my Ph1 students for making this responsibility an enjoyable learning experience. I have also received financial support via the Caltech department of physics, the John A. McCone Chair, and the U.S. Department of Energy under contract number DE-FG03-92ER40701. I am grateful to Carol Silberstein and Charlene Cartier for their friendliness and their extensive administrative support.

My loving wife Sara has provided encouragement, emotional support, and patient understanding of the demands of my graduate work. Meanwhile her thoughtful and careful planning for our time together has made my life and my work all the more rewarding. For these, and everything else Sara gives, I am endlessly grateful. I am also thankful for my siblings, Jeff, Joe, Karen, and Sherry, their spouses Jason and Mark, and my parents Frank and Marie. Their constant praise has encouraged me throughout my studies. More importantly, they contributed most to make me who I am.

Finally, I am thankful for the countless researchers who have preceded me. Their elucidation has increased the beauty of the world for me, and without their progress none of this work would have been possible. I appreciate that the knowledge, opportunity, and abundance that makes my life so wonderful was built on the cumulative work of generations of these and other people, who persisted through experiences far more difficult than mine, to create a better world. I can only hope to express my gratitude through my own contributions toward a more knowledgeable, just, and prosperous future for humanity.

Abstract

We first delve into particle phenomenology with a study of soft-collinear effective theory (SCET), an effective theory for Quantum Chromodynamics for when all particles are approximately on their light-cones. In particular, we study the matching of $\mathrm{SCET_I}$ involving ultrasoft and collinear particles onto $\mathrm{SCET_{II}}$ involving soft and collinear particles. We show that the modes in $\mathrm{SCET_{II}}$ are sufficient to reproduce all of the infrared divergences of $\mathrm{SCET_I}$, a result that was previously in contention.

Next we move into early universe cosmology and study alternative mechanisms for generating primordial density perturbations. We study the inhomogeneous reheating mechanism and extend it to describe the scenario where the freeze-out process for a heavy particle is modulated by subdominant fields that received fluctuations during inflation. This scenario results in perturbations that are comparable to those generated by the original inhomogeneous reheating scenarios. In addition, we study yet another alternative to single field inflation whereby the curvature perturbation is generated by interactions at the end of inflation, as opposed to when inflaton modes exit the horizon. We clarify the circumstances under which this process can dominate over the standard one and we show that it may result in a spectrum with an observable level of non-Gaussianities.

We then turn to studies of the landscape paradigm, which hypothesizes that the observed universe is just one among a multitude of possibilities that are realized in separate causal regions. Such a landscape has been used to explain the smallness of the cosmological constant, at least when only it scans across the landscape. We study the scenario where both the cosmological constant and the strength of gravity, parameterized by the effective Planck mass, scan across the landscape. We find that selection effects acting on the cosmological constant are significantly weaker in this scenario and we find the measured value of the Planck mass to be exponentially unlikely under certain plausible assumptions about the landscape. Finally, we study some other models of the landscape as part of a possible explanation for quark-sector flavor parameters in the Standard Model. In this picture quark Yukawa couplings result from overlap integrals involving quark and Higgs wavefunctions in compactified extra dimensions, and the values we measure result from random selection from a landscape of possibilities. We find that many of the salient features of the measured flavor parameters are typical of the landscape distribution.

Contents

List of Figures

Chapter 1

Introduction

This thesis includes all of the published work resulting from my graduate studies at Caltech (and some work to be published in the near future). Since my interests have shifted during these studies, there is no unifying theme that relates all of the chapters in this thesis. Therefore each subject of investigation is presented as a separate chapter with no attempt to unify this work as a whole. To serve as an introduction to this work, a brief background discussion to each of the major topics covered in this thesis is included in the sections below.

1.1 Soft-Collinear Effective Theory

My graduate studies began with an investigation into soft-collinear effective theory (SCET), an effective theory for Quantum Chromodynamics (QCD). A slightly edited version of the publication resulting from this work appears as Chapter 2 of this thesis. In what follows I provide background to this work and describe its context. This involves supplying very brief introductions to SCET and to heavy quark effective theory (HQET). More in-depth descriptions of SCET and HQET can be found in the references cited below. At the end of this section I describe my personal contributions to the work presented in Chapter 2.

We start with a brief description of SCET. The reason for seeking an effective theory for QCD in the first place is that at the energy scales associated with hadrons the QCD coupling "constant" α_s is of order unity. This makes it very difficult to study QCD perturbatively at these energies. SCET [1, 2, 3, 4] identifies a different perturbative quantity that can sometimes be used to study QCD interactions. Specifically, in many QCD interactions the invariant four-momentum dot product p^2 for the constituent particles involved in an interaction is much less than then square of the interaction energy Q^2. This hierarchy can be parameterized using the small quantity λ, with $p^2 \sim Q^2\lambda^2$. We can then study QCD perturbatively with respect to an expansion in powers of λ.

It is convenient to work in light-cone coordinates which refer to the light-cone four-vectors n and $\bar{n}$, with $n^2 = \bar{n}^2 = 0$ and $n \cdot \bar{n} = 2$ (for example, particles moving back-to-back along the z-axis would follow the four-vectors $n = (1\,,0\,,0\,,1)$ and $\bar{n} = (1\,,0\,,0\,,-1)$ where the first component is the time component). In these coordinates the components of a general four-momentum are written

$$p^\mu = (n \cdot p,\ \bar{n} \cdot p,\ p^\perp) \equiv (p^+,\ p^-,\ p^\perp)\,, \tag{1.1}$$

where $p^\perp \equiv p - \frac{1}{2}(\bar{n} \cdot p)n - \frac{1}{2}(n \cdot p)\bar{n}$. In this notation $p^2 = p^+p^- + p_\perp^2$ from which we can see that for off-shellness $p^2 \sim Q^2\lambda^2$ we may have $p^\mu \sim Q(\lambda^2,\, 1,\, \lambda)$ or $p^\mu \sim Q(\lambda,\, \lambda,\, \lambda)$. The former are referred to as collinear momenta and denoted with a subscript c while the latter are referred to as soft momenta are are denoted with a subscript s. Since adding a single soft momentum to a single collinear momentum takes the latter further off-shell (by increasing the p^+ component of the momentum), soft-collinear interactions must involve at least two soft and two collinear particles. On the other hand, a single collinear particle may directly couple to what is referred to as an ultrasoft (usoft) particle, which is defined by the momenta scaling $p^\mu \sim Q(\lambda^2,\, \lambda^2,\, \lambda^2)$ and which is denoted with the subscript us.

The SCET Lagrangian is obtained by integrating out collinear momenta from the QCD Lagrangian. The use of light-cone coordinates is pivotal for making the power counting in λ manifest. We first write $p_c^\mu = \tilde{p}^\mu + k^\mu$ where $\tilde{p}^\mu = \frac{1}{2}p^- n^\mu + p_\perp^\mu$ includes the large components of the collinear momentum and k^μ includes everything else. Fermions can then be expanded according to

$$\psi(x) = \sum_{\tilde{p}} e^{-i\tilde{p}\cdot x}\psi_{n,\tilde{p}}(x) = \sum_{\tilde{p}} e^{-i\tilde{p}\cdot x}\left[\xi_{n,\tilde{p}}(x) + \xi_{\bar{n},\tilde{p}}(x)\right]\,, \tag{1.2}$$

where $\xi_{\bar{n},\tilde{p}} \equiv \frac{1}{4}\not{\bar{n}}\not{n}\psi_{n,\tilde{p}}$ and $\xi_{n,\tilde{p}} \equiv \frac{1}{4}\not{n}\not{\bar{n}}\psi_{n,\tilde{p}}$ are projections of the original fermion onto its collinear and anti-collinear components. The next step in elucidating the power-counting is to separate the gluon fields into collinear, soft, and usoft components $A^\mu = A_c^\mu + A_s^\mu + A_{us}^\mu$, where the components are separated so as to satisfy the power-counting

$$A_c^\mu \sim Q(\lambda^2,\, 1,\, \lambda)\,, \quad A_s^\mu \sim Q(\lambda,\, \lambda,\, \lambda)\,, \quad A_{us}^\mu \sim Q(\lambda^2,\, \lambda^2,\, \lambda^2)\,. \tag{1.3}$$

Furthermore, the large collinear momenta of A_c^μ are factorized analogously to those of ψ as in Eq. (1.2). When the QCD Lagrangian is written in terms of $\xi_{\bar{n},\tilde{p}}$, $\xi_{n,\tilde{p}}$, A_c, A_s, and A_{us}, one finds that at leading order the field $\xi_{\bar{n},\tilde{p}}$ does not interact with soft or usoft degrees of freedom. Therefore at tree level the field $\xi_{\bar{n},\tilde{p}}$ can be replaced by substituting the solution to its equation of motion. Putting all of this together gives the leading-order SCET Lagrangian,

$$\mathcal{L}_{\text{SCET}} = \sum_{\tilde{p},\tilde{q}} e^{-i(\tilde{p}-\tilde{q})\cdot x}\bar{\xi}_{n,\tilde{q}}\left[in \cdot D + (\not{\tilde{p}}_\perp + i\not{D}_\perp^c)\frac{1}{\tilde{p}^- + i\bar{n} \cdot D^c}(\not{\tilde{p}}_\perp + i\not{D}_\perp^c)\right]\frac{\not{\bar{n}}}{2}\xi_{n,\tilde{p}}\,, \tag{1.4}$$

where $D_\mu^c = \partial_\mu - igA_\mu^c$ is a collinear covariant derivative.

The SCET Lagrangian can be simplified by introducing a convenient notation [3]. We define what are referred to as label operators which act according to the definitions

$$\bar{n}\cdot\mathcal{P}\xi_{n,\tilde{p}} \equiv \tilde{p}^-\,\xi_{n,\tilde{p}}\,, \quad \mathcal{P}_\perp^\mu\xi_{n,\tilde{p}} \equiv \tilde{p}_\perp^\mu\xi_{n,\tilde{p}}\,. \tag{1.5}$$

The covariant collinear derivative D^c can then be written in terms of the label operators, that is

$$D_\mu^c = \mathcal{P}_\mu - igA_\mu^c\,. \tag{1.6}$$

One advantage of this notation is that it allows us to absorb the large collinear momenta back into the collinear fields. Thus we define the collinear field

$$\xi_n \equiv \sum_{\tilde{p}} e^{-i\tilde{p}\cdot x}\xi_{n,\tilde{p}}\,. \tag{1.7}$$

The label operators then act to pull out the full collinear momentum of a field χ_n. For example,

$$\bar{n}\cdot\mathcal{P}\xi_n = \sum_{\tilde{p}} \tilde{p}^- e^{-i\tilde{p}\cdot x}\xi_{n,\tilde{p}}\,. \tag{1.8}$$

This allows the SCET Lagrangian to be written in the simpler form,

$$\mathcal{L}_{\rm SCET} = \bar{\xi}_n\left[in\cdot D + i\not{D}_\perp^c\frac{1}{i\bar{n}\cdot D^c}i\not{D}_\perp^c\right]\frac{\not{\bar{n}}}{2}\xi_n\,. \tag{1.9}$$

Note that the leading-order Lagrangian contains no couplings between soft and collinear fields, and that couplings between usoft and collinear fields proceed only via the first term. In fact, even this interaction can be removed via a clever field redefinition [4]:

$$\xi_n = Y_n\xi_n^{(0)}\,, \qquad A = Y_nA^{(0)}Y_n^\dagger\,, \qquad Y_n(x) = \mathrm{P}\exp\left(ig\int_{-\infty}^0 ds\,n\cdot A_{us}(x+ns)\right)\,, \tag{1.10}$$

where the "P" refers to path ordering. This field redefinition essentially converts the $in\cdot D$ in Eq. (1.9) into an $in\cdot D^c$, keeping the remaining form of the Lagrangian the same.

The Feynman rules for SCET derive from the Lagrangian Eq. (1.9). The rules for interacting (u)soft degrees of freedom are exactly the same as in QCD. Meanwhile, the propagator for a collinear quark with momentum $p = \tilde{p} + k$ is given by

$$\frac{\not{n}}{2}\frac{i\tilde{p}^-}{\tilde{p}^-k^+ + \tilde{p}_\perp^2 + i\epsilon}\,. \tag{1.11}$$

If we do not perform the field redefinitions of Eqs. (1.10) then collinear fields interact with soft gluons via the first term in Eq. (1.9). This gives the vertex factor $\frac{\not{\bar{n}}}{2}ign_\mu T^a$ where T^a is a generator for SU(3). Interactions between collinear quarks and collinear gluons are more complicated and proceed via the second term in Eq. (1.9). This gives the vertex factor

$$\frac{\not{\bar{n}}}{2} ig \left(n_\mu + \frac{\not{\tilde{p}}_\perp}{\tilde{p}^-} \gamma_\mu^\perp + \frac{\not{\tilde{q}}_\perp}{\tilde{q}^-} \gamma_\mu^\perp - \frac{\not{\tilde{p}}_\perp \not{\tilde{q}}_\perp}{\tilde{p}^- \tilde{q}^-} \bar{n}_\mu \right) T^a , \tag{1.12}$$

where the collinear quark carries collinear momentum $\tilde{p}$ and the collinear gluon carries collinear momentum $\tilde{p} - \tilde{q}$. There is no special power counting among the components of (u)soft momenta; therefore interactions between (u)soft fields follow the usual Feynman rules of QCD.

It is now time to discuss some specific applications of SCET. First, consider the inclusive decay of a heavy hadron such as the B meson. Such decays involve collinear particles with off-shellness of order $p_c^2 \sim m_b \Lambda_{\rm QCD}$ and involve non-perturbative degrees of freedom with momenta components of order $\Lambda_{\rm QCD}$ [2]. Identifying the large energy scale with the b mass, $Q \sim m_b$, we see this scenario is described by SCET with $\lambda = \sqrt{\Lambda_{\rm QCD}/m_b}$ where the non-perturbative degrees of freedom correspond to usoft particles. To distinguish it from a second type of application that is described in the next paragraph, this application of SCET is referred to as SCET_I .

A second application of SCET is the study of exclusive decays of heavy hadrons to light hadrons, for example $B \to \pi\pi$ [5, 6]. In this application the off-shellness of collinear particles is $p_c^2 \sim \Lambda_{\rm QCD}^2$, which corresponds to an SCET collinear momentum with $\lambda \to \tilde{\lambda} = \Lambda_{\rm QCD}/m_b$ in the example $B \to \pi\pi$. Meanwhile, the non-perturbative degrees of freedom are still described by particles with momenta components of order $\Lambda_{\rm QCD}$. Thus in this scenario non-perturbative degrees of freedom correspond to soft particles, and usoft particles are unimportant for the processes. This application of SCET is referred to as SCET_II .

Although both SCET_I and SCET_II correspond to the same perturbative expansion with respect to momenta components, SCET_I describes collinear degrees of freedom with off-shellness $p_c^2 \sim Q\Lambda_{\rm QCD}$ whereas SCET_II describes collinear degrees of freedom with off-shellness $p_c^2 \sim \Lambda_{\rm QCD}^2$. Therefore SCET_II can be viewed as a low-energy effective theory to SCET_I . As such, it should be possible to match SCET_I onto SCET_II by integrating out the degrees of freedom with energies greater than those described by SCET_II . These correspond to what would be soft momenta in SCET_I , those components with magnitude of order $\sqrt{Q\Lambda_{\rm QCD}}$. The work in Chapter 2 performs this matching, which involves some unanticipated subtleties.

Although the work in Chapter 2 concerns matching SCET_I onto SCET_II , it also references HQET in the specific examples chosen to elucidate this matching. Therefore we now present a brief introduction to HQET. This is an effective theory for QCD where the perturbative expansion parameter is taken to be the energy of light degrees of freedom over the mass of a heavy quark with which they interact (for a complete review see for example [7]). In the case of a hadron containing the bottom quark, for example, this would be $\varepsilon = \Lambda_{\rm QCD}/m_b$. The large momentum of the heavy quark ψ is factorized as it is in SCET:

$$\psi(x) = e^{-im_\psi v\cdot x}\left[h_v(x) + \tilde{h}_v(x)\right] , \tag{1.13}$$

where $h_v = \frac{1}{2}(1+\not{v})e^{im_\psi v\cdot x}\psi$ and $\tilde{h}_v = \frac{1}{2}(1-\not{v})e^{im_\psi v\cdot x}\psi$ are projections of ψ and v is the velocity of the heavy quark ψ.

It can be shown that the effects of the field $\tilde{h}_v$ are suppressed relative those of h_v by a power of $\epsilon = p_h/m_\psi$, where p_h is the residual momentum of the h_v field. Therefore it is lower order in the expansion and can be ignored. The resulting Lagrangian for h_v is

$$\mathcal{L}_{\rm HQET} = \bar{h}_v\, iv\cdot D\, h_v\, , \tag{1.14}$$

where D is the ordinary covariant derivative. The heavy quark propagator that follows from this Lagrangian is given by

$$\frac{1+\not{v}}{2}\frac{i}{v\cdot k + i\epsilon}\, , \tag{1.15}$$

where k is the residual momentum of the heavy quark, $k = p - m_\psi v$. Meanwhile, the coupling of the heavy quark to a gluon gives the vertex $igv_\mu T^a$. This is all the background to HQET that is required to approach Chapter 2.

Chapter 2 of this thesis is based on “Infrared regulators and SCET(II),” Christian W. Bauer, Matthew P. Dorsten, and Michael P. Salem, *Physical Review D* **69**, 114011 (2004) [8]. My tangible contributions to this collaboration were that I discovered the infrared regulator for SCET that is used to support the main conclusions of this paper, I was the first to verify that this regulator succeeded in disentangling the ultraviolet and infrared divergences that are mixed in pure dimensional regulation, and I was the first to perform some of the other calculations in support of this paper. I verified all of the calculations presented in the paper. Finally, I contributed toward developing a meaningful interpretation of the results of this work.

1.2 Alternative Mechanisms to Generate Primordial Density Perturbations

After studying soft-collinear effective theory, my interests turned toward early universe cosmology. The work that ensued involved the study of alternative mechanisms to generate the primordial density perturbations observed in the cosmic microwave background (CMB). It will put this work in context to first describe the standard mechanism for generating these density perturbations. It should be emphasized that the standard picture is commonly adopted only due to its simplicity, and not because experimental evidence distinguishes it from any other possibilities. For reviews of the standard picture with references to the original literature, see for example Refs. [9, 10, 11].

In the standard picture the primordial density perturbations are ultimately sourced by quantum fluctuations in a single inflaton field φ. If over some volume the field φ dominates the energy density and the non-gradient potential energy of φ is sufficiently greater than the gradient energy and the kinetic energy in φ, then this volume will rapidly expand such that the field gradients and kinetic energy redshift away while the non-gradient potential energy in φ remains relatively constant. The geometry of this volume rapidly approaches the homogeneous and isotropic FRW metric. If the ensuing period of expansion is to resolve the horizon problem then the local physical horizon must expand at a speed less than the physical speed of light. Since the local physical horizon is proportional to the Hubble radius H^{-1}, this constraint is encapsulated in the requirement that the so-called first slow-roll parameter ϵ be less than one:

$$\epsilon \equiv \frac{d}{dt}\left(\frac{1}{H}\right) < 1\,. \tag{1.16}$$

A period of expansion with this relation satisfied is called inflation.

The generation of density perturbations from inflation can be seen qualitatively as follows. In a volume where $\epsilon < 1$ physical modes of constant wavelength can rapidly expand from being well within the local horizon to being much larger than the local horizon. Modes that correspond to quantum fluctuations of the inflaton vacuum expectation value (vev) freeze into classical perturbations about the otherwise homogeneous vev of φ when they expand beyond the local horizon. These perturbations in φ imply that inflation lasts slightly longer in some regions than in others. Meanwhile, the redshift of energy density is highly suppressed during inflation relative the redshift of matter or radiation afterward. Thus the perturbations in φ as modes exit the local horizon translate directly into density perturbations in the radiation after reheating. The observed scale invariance of these density perturbations implies that H does not vary significantly during inflation, which in turn implies $\epsilon \ll 1$. In this limit the quantitative understanding of inflation becomes more transparent.

Clearly the limit $\epsilon \ll 1$ corresponds to a Hubble rate that is nearly constant in time and therefore this limit corresponds to (quasi-) de Sitter space-time. By examining the Einstein field equations it is not hard to see that this limit corresponds to suppressed φ dynamics, $\dot{\varphi}^2 \ll V$, and thus a nearly constant, potential-dominated energy density:

$$H^2 \simeq \frac{V}{3m_{\mathrm{P}}^2}\,. \tag{1.17}$$

About 60 e-folds (a factor of e^{60} in scale factor growth) of such nearly de Sitter expansion is sufficient to solve the horizon and flatness problems of the original hot big bang picture, while being more than enough to redshift away any relics of the grand unification phase transition if inflation occurs below this transition. It is not hard to construct inflaton models to generate over 60 e-folds of inflation; indeed a simple canonical massive scalar $V = \frac{1}{2}m_\varphi^2\varphi^2$ easily generates far more inflation. On the other hand the inflationary paradigm also explains the generation of the small primordial perturbations to the otherwise uniform energy density of the early universe. Constructing "natural" models of inflation that result in perturbations matching those observed in the CMB presents a significant challenge.

We now outline the calculation of these perturbations in the standard picture. This is straightforward to do within the so-called δN formalism [12, 13, 14, 15]. This formalism notes that on super-Hubble scales the number of e-folds of expansion between an initial (time t_0) flat hypersurface and a final (time t) hypersurface of constant density is given by

$$N(t,\mathbf{x}) = \ln\left[\frac{a(t)e^{\zeta(t,\mathbf{x})}}{a(t_0)}\right], \tag{1.18}$$

where a is the homogeneous scale factor and ζ is the gauge-invariant Bardeen curvature perturbation [16, 17]. This result also ignores the effect of anisotropic stress perturbations, but these are not excited by fluctuations in a scalar field (nor by fluctuations in non-relativistic matter and radiation). Rearranging the terms in Eq. (1.18) gives

$$\zeta(t,\mathbf{x}) = N(t,\mathbf{x}) - \ln\left[\frac{a(t)}{a(t_0)}\right] \equiv \delta N(t,\mathbf{x}). \tag{1.19}$$

In the standard picture the fluctuations in ζ stem from fluctuations in φ when modes leave the local horizon. The effect of these fluctuations on N is relatively small such that we can Taylor expand to obtain ζ:

$$\zeta = N'\delta\varphi + \frac{1}{2}N''\delta\varphi^2 - \frac{1}{2}N''\langle\delta\varphi^2\rangle + \dots. \tag{1.20}$$

Here the prime denotes differentiation with respect to φ and $\delta\varphi(t,\mathbf{x})$ is the spatial variation in φ, evaluated on the initial flat hypersurface defined during inflatino. The observation that non-Gaussianities are suppressed in the power spectrum of ζ implies that we only need to keep the first term in this expansion. In the slow-roll approximation the number of e-folds of expansion can be written

$$N \simeq \int_{t_i}^{t_f} H dt \simeq -\frac{1}{m_{\mathrm{P}}^2}\int_{\varphi_i}^{\varphi_f}\frac{V}{V'}d\varphi, \tag{1.21}$$

such that $N' = (1/m_{\mathrm{P}}^2)V/V'$. Thus the power spectrum for ζ is related to the power spectrum for $\delta\varphi$ by $P_\zeta \simeq (1/2\epsilon\, m_{\mathrm{P}}^2)P_{\delta\varphi}$.

The power spectrum for the (real valued) inflaton fluctuations $\delta\varphi$ are calculated using the quantum theory of a scalar field in slightly perturbed de Sitter space. One calculates the two-point correlation function which relates to the power spectrum according to the definition[1]

$$\langle \delta\varphi(t,\mathbf{k})\delta\varphi(t,\mathbf{k}')\rangle \equiv \left(\frac{2\pi}{k}\right)^2 P_{\delta\varphi}(k)\,\delta^{(3)}(\mathbf{k}-\mathbf{k}')\,. \tag{1.22}$$

Here $\mathbf{k}$ labels the Fourier mode of the transformed inflaton φ. The standard result comes from evaluating the correlation function in Eq. (1.22) at tree level in the limit where V'/V and V''/V are roughly constant, which holds when $\epsilon \ll 1$ and $\dot{\epsilon}/\epsilon H \ll 1$. It is customary to define a new small parameter that relates to the second constraint. This is referred to as the second slow-roll parameter (ϵ being the first slow-roll parameter) and it can be defined $\eta \equiv 2\epsilon - \dot{\epsilon}/2\epsilon H$. It is useful to note that when $\epsilon \ll 1$ and $\eta \ll 1$ the slow-roll parameters are simply related to the inflaton potential:

$$\epsilon \simeq \frac{m_{\rm P}^2}{2}\left(\frac{V'}{V}\right)^2, \quad \eta \simeq m_{\rm P}^2\left(\frac{V''}{V}\right)\,. \tag{1.23}$$

Indeed, many authors use these as the definitions for ϵ and η. To leading order in ϵ and η and on scales much larger than the Hubble radius during inflation the power spectrum for $\delta\varphi$ can be written

$$P_{\delta\varphi}(k) \simeq \frac{1}{2}H_k^2(-k\tau)^{2\eta-4\epsilon}\,, \tag{1.24}$$

where τ measures conformal time ($\tau = -1/aH_k$ with the scale factor a normalized to $a = 1$ at $k = H_k$) and H_k is the Hubble rate when the mode k exits the Hubble radius. The weak dependence on k and η imply that the power spectrum of $\delta\varphi$ is nearly scale invariant and does not decay much during the course of inflation. The scale dependence of this power spectrum is parameterized by the tilt $n-1$. Using $H_k \propto k^{-\epsilon}$ one finds

$$n - 1 \equiv \frac{d\ln P_{\delta\varphi}}{d\ln k} = -6\epsilon + 2\eta\,. \tag{1.25}$$

Putting all this together we find the curvature perturbation that results from single field inflation has the approximate power spectrum

$$P_\zeta \simeq \frac{1}{4\epsilon}\frac{H^2}{m_{\rm P}^2}\,, \tag{1.26}$$

with H evaluated during inflation and with a tilt given by Eq. (1.25). This curvature perturbation ultimately sources the fluctuations in temperature that are observed in the CMB, and matching the actual spectrum of CMB fluctuations onto what is expected from the above curvature perturbation implies tight constraints on the inflaton potential. On the other hand, if alternative mechanisms

---

[1] Note that some authors use $\mathcal{P}_{\delta\varphi} \equiv P_{\delta\varphi}/2\pi^2$ for the power spectrum.

exist to generate this curvature perturbation then the CMB power spectrum may not constrain the inflationary dynamics as strongly as the standard picture suggests. This motivates investigation into what other physical processes may generate the primordial curvature perturbation.

Perhaps the simplest generalization of the standard picture is to include a single additional field σ into the inflationary dynamics. As we shall see, processes involving the field σ may generate the principle component of the primordial curvature perturbation even if the energy density associated with σ is sub-dominant during all of inflation. In this case all of the dynamics described above are unchanged, except that the curvature perturbation generated by the inflaton is no longer constrained by the CMB. The basic idea is that analogous to the case with the inflaton φ, the vev of σ will receive fluctuations as modes expand beyond the local horizon. Although σ is sub-dominant during inflation these fluctuations $\delta\sigma$ may be transferred to the dominating form of energy density by a variety of processes that follow inflation. As an introduction to the work of Chapters 3 and 4 we now discuss one of these possibilities in greater detail.

Dvali, Gruzinov, Zaldarriaga, and independently Kofman (DGZK) proposed a mechanism to generate density perturbations that is alternative to the standard picture and that requires only a single light scalar σ in addition to the inflaton φ [18, 19, 20]. As has been mentioned, both σ and φ receive fluctuations to their vevs $\delta\sigma$ and $\delta\varphi$ that behave classically as modes expand beyond the Hubble radius. Since we are interested in providing an alternative to the standard picture for generating density perturbations, we presume the fluctuations $\delta\varphi$ are insignificant to the subsequent evolution of the universe. On the other hand we assume there exist interactions between the fields σ and φ such as those contained in the interaction Lagrangian density

$$\mathcal{L}_{\rm int} = \frac{\lambda_1}{2} M\sigma\,\varphi^2 + \frac{\lambda_2}{4}\,\sigma^2\varphi^2 + \frac{\lambda_3}{2}\frac{\sigma}{M}\,\varphi\,\bar{\psi}\psi + \frac{\lambda_4}{2}\varphi\,\bar{\psi}\psi\,, \tag{1.27}$$

where the λ_i are dimensionless couplings, M is a mass scale, and the field ψ represents a fermion via which the inflaton reheats the universe.

To review a standard picture of reheating consider first the scenario $\lambda_1 = \lambda_2 = \lambda_3 = 0$. In this picture inflation ends with the inflaton φ rocking back and forth at the bottom of its potential well, with effective mass m_φ. During this coherent oscillation the energy density in φ redshifts like non-relativistic matter. Meanwhile, if the fermion ψ is very light next to the inflaton then the inflaton decays into ψ particles at the rate $\Gamma = \lambda_4^2 m_\varphi/32\pi$. However the ψ particles will be relativistic and thus redshift like radiation. It can be shown that the energy density in ψ becomes comparable to the energy density in φ when $\Gamma = H$, which is when the inflaton is said to decay. The energy density in the presumably radiative byproducts of ψ interaction have energy density $(\pi^2/30)g_* T^4$ for g_* effective bosonic degrees of freedom at the temperature T. Thus the reheat temperature is $T_{\rm RH} \sim g_*^{-1/4}\sqrt{\Gamma m_{\rm P}}$.

Now consider the scenario with $\lambda_3 \neq 0$. At tree level the decay proceeds as before except with the new decay rate

$$\Gamma' = \frac{m_\varphi}{32\pi}\left(\lambda_4 + \lambda_3\frac{\langle\sigma\rangle}{M}\right)^2 . \tag{1.28}$$

The vev of σ receives spatial fluctuations $\delta\sigma$ during inflation, and these translate into spatial fluctuations in the decay rate:

$$\delta\Gamma = \frac{m_\varphi}{32\pi}\left[2\lambda_3\lambda_4\left(1 + \frac{\lambda_3}{\lambda_4}\frac{\sigma_0}{M}\right)\frac{\delta\sigma}{M} + \lambda_3^2\frac{\delta\sigma^2}{M^2}\right] , \tag{1.29}$$

where σ_0 is the homogeneous component of $\langle\sigma\rangle$. Since the reheat temperature is proportional to $\sqrt{\Gamma}$, it can be seen immediately that the fluctuations in σ translate into temperature fluctuations after reheating. Alternatively, one can view density perturbations as arising due to greater redshifting of energy density in regions where φ decays to radiation more rapidly.

It is not hard to see how the other interactions in Eq. (1.27) lead to density perturbations. Both the first and second terms here can be seen as introducing a σ-dependent offset to the effective mass of φ. The spatial variations in the vev of σ then translate into spatial variations in the mass of φ. Since the inflaton decay rate Γ is proportional to the mass of φ, these variations correspond to fluctuations in the decay rate and density perturbations result as described above. In fact it is not necessary that φ in Eq. (1.27) be the inflaton. One can envision scenarios where the σ field is so light and long-lived that it lies sub-dominant even as some other massive field S comes to dominate the energy density after reheating occurs in the early universe. If S interacts with σ as φ does in Eq. (1.27), then fluctuations in the effective mass or decay rate of S translate into fluctuations in the duration of S domination and hence density perturbations. Although these generalizations to the original DGZK mechanism were noted in Ref. [19], an important effect was neglected. Specifically, if the field S is to dominate the energy density of the universe, it must freeze-out of equilibrium with the rest of the universe. If S freezes-out while non-relativistic then fluctuations in its mass and its annihilation cross section translate into fluctuations in its relic abundance. These generate density perturbations comparable to those generated by inhomogeneous decay when the decay of S reheats the universe. These effects are described in Chapter 3 of this thesis.

The primordial spectrum of density perturbations that results from the DGZK mechanism has interesting characteristics. For one, it should be noted that the amplitude of the power spectrum of ζ will not be given by Eq. (1.26), but by a spectrum depending on φ-σ coupling parameters, $\langle\sigma\rangle$, and the power spectrum of σ fluctuations $P_{\delta\sigma}$. Since σ is sub-dominant during inflation $P_{\delta\sigma}$ is not required to have the same form as $P_{\delta\varphi}$. However, it turns out that the amplitudes of these power spectra are the same when each is evaluated for a mode leaving the Hubble radius and to leading order in small parameters. On the other hand the spectrum of $\delta\sigma$ differs only in its tilt,

$$n - 1 = -2\epsilon + 2\lambda\,, \tag{1.30}$$

where $\lambda \equiv m_\sigma^2/3H_k^2$ is a new small parameter. Finally, although this point has not been emphasized above, it can be shown that the spectrum of fluctuations in $\delta\varphi$ is always highly Gaussian in its statistical distribution [21]. In contrast, it can be seen for example from Eq. (1.29) that the level of non-Gaussianities in the DGZK mechanism are controlled by the relative sizes of coupling parameters and the ratio σ_0/M.

Other alternatives to the standard picture of density perturbation generation are simply related to the DGZK mechanism. For example, the DGZK mechanism was preceded by the so-called curvaton mechanism [22, 23, 24] in which the light field σ need not interact with any degrees of freedom other than those to which it slowly decays. In this picture the light field σ receives fluctuations to its vev just as in the DGZK mechanism but persists well after inflation as a cold relic. Eventually the curvaton comes to dominate the energy density of the universe and begins coherent oscillations about the bottom of its interaction potential. When these oscillations begin is modulated by the vev of sigma which has received spatial variations; thus density perturbations result from σ decay.

Yet another alternative was proposed by Lyth in Ref. [25]. This also features a light scalar σ, however now the context is hybrid inflation. In hybrid inflation the vacuum energy that drives de Sitter expansion is eliminated when a “waterfall” field χ achieves a non-zero vev. However during inflation the field χ has a large effective mass due to interactions with the inflaton φ, and this effective mass pins χ to $\chi = 0$, which is a point of high potential energy. As in standard inflation the vev of φ slowly decreases and eventually this opens an instability in χ whereby it can cascade to the minimum of its potential. In Lyth's proposal the effective mass of the waterfall field χ is determined by both the vev of φ and the vev of a light field σ. Then fluctuations in the vev of σ modulate the precise timing of when the field χ cascades to its minimum, in turn modulating the duration of inflation. This mechanism for generating density perturbations is discussed in greater detail in Chapter 4 of this thesis.

Chapter 3 is based on “Fluctuating annihilation cross sections and the generation of density perturbations,” Christian W. Bauer, Michael L. Graesser, and Michael P. Salem, *Physical Review D* **72**, 023512 (2005) [26]. My individual contributions to this work include that I developed the general analysis for the evolution of density perturbations that is presented in Section 3.5 (this was later further generalized by Michael Graesser) and I performed the full Boltzmann analysis and obtained the numerical results of Section 3.6. In addition, I performed-first many of the calculations in this paper and I verified all of the work. Finally, I wrote most of the published manuscript.

Chapter 4 is based on “On the generation of density perturbations at the end of inflation,” Michael

P. Salem, *Physical Review D* **72**, 123516 (2005) [27]. Aside from input from some discussions with acknowledged sources, I am entirely responsible for this paper.

1.3 Low-Energy Consequences of a Landscape

Finally, my graduate studies have included investigations into what may be the implications for the low-energy effective theory that describes our universe of an enormous landscape of metastable states in the fundamental theory. These investigations are motivated by the convergence of a few recent results in high energy theoretical physics. First among these is the discovery that string theory, the leading candidate for a fundamental theory of physics, may contain a staggering multitude of meta-stable solutions, each of which may permit a different set of apparent physical laws for the low-energy effective theory (see [28] and references therein). Meanwhile, the phenomenom of eternal inflation [29, 30, 31] provides a means of populating all of the meta-stable states in the fundamental theory in a vast multiverse of distinct "pocket" universes. Finally, these possibilities received motivation in the anthropic argument justifying the observed smallness of the cosmological constant. For background we now describe this argument in greater detail.

Weinberg was the first to make concrete the observation that if the cosmological constant were too large and if all other physics remained the same then non-linear gravitationally bound structures, and hence life, would not readily form in this universe [32, 33]. The basic argument requires the following background. In our universe the spectrum of primordial density perturbations is nearly scale invariant such that the amplitude of perturbations just entering the Hubble radius is nearly constant. During radiation domination the fluctuations in the baryon and radiation energy densities decay after entering the Hubble radius, while fluctuations in the dark matter grow only logarithmically with time. On the other hand, after matter domination fluctuations in the dark matter density grow in proportional to the growth in scale factor, and after recombination the baryons also gather in the resulting potential wells. Under gravitational attraction these perturbations to the energy density continue to grow until cosmological constant dominates the energy density of the universe, after which growth in linear density perturbations essentially halts.

The above evolution of density perturbations implies that perturbations can only grow non-linear due to evolution between matter domination and the domination of cosmological constant. On the other hand, perturbations must grow non-linear before they can separate from the cosmic expansion, collapse, virialize, and further collapse and fragment into galaxies. If the cosmological constant were too large (relative to the energy density in matter at matter-radiation equality) then only very rarely would a fluctuation in the otherwise-constant energy density of the universe have sufficiently large initial amplitude to have become non-linear by the domination of cosmological constant. Thus if we demand that typical fluctuations in the energy density of the universe eventually grow non-linear, then this implies a constraint on the size of the cosmological constant.

By itself this constraint does not explain anything; however it has important implications in the context of a fundamental theory that contains an enormous “landscape” of meta-stable solutions. If the landscape of the fundamental theory is sufficiently large, meta-stable states may exist with a wide range of cosmological constants, including very tiny cosmological constants deriving from chance cancellations among larger terms contributing to the “vaccum” energy of the meta-stable state. Meta-stable states with a cosmological constant as small as ours would seem to be exceedingly rare among the meta-stable states that constitute the full landscape of the theory, and therefore all else being equal one would consider it very unlikely that we would happen to find ourselves in such a state. However, the previously described constraint on the size of the cosmological constant changes this expectation.

Specifically, we should not expect to measure with equal likelihood each value of the cosmological constant that is realized in the landscape. Not only will some values may be realized in many meta-stable states and some meta-stable states be realized more frequently in the multiverse than others, but also the likelihood for observers like us to evolve in a given universe will depend on its value of the cosmological constant, in addition to the other low energy physics that governs the universe. In principle we imagine a probability distribution function that weights each value of the cosmological constant by our likelihood to observe it, and the least presumptuous (non-trivial) assumption regarding the value of the cosmological constant that we measure is that it is typical of the values receiving significant weight in this distribution. Our likelihood to observe a value of the cosmological constant is weighted by the relative likelihood that the value is obtained within a given universe in the multiverse times the relative likelihood that we could have arisen in that universe. Inclusion of the latter factor follows from what is called the anthropic principle [34, 35, 36].

In the case of the cosmological constant, it will turn out that almost all of the weight of the full probability distribution function lies over values of the cosmological constant that are extremely small compared to the natural scale for this quantity, which is presumed to be comparable to the Planck scale. If zero cosmological constant is not a special point among the landscape of allowed values, then it is reasonable to presume that over the range of values that receive significant weight the distribution of cosmological constant values among the meta-stable states in the fundamental theory is relatively constant. This means the shape of the distribution function for the cosmological constant is determined primarily by the so-called anthropic factor, which weights the relative likelihood for us to have evolved in a given universe. The anthropic factor is in principle extremely complicated to compute, as it should account for all of the environmental selection effects that modulate the likelihood for our existence. However if we restrict attention to a subset of the landscape defined by universes exactly like ours except for the value of the cosmological constant, the problem is dramatically simplified. Such a restriction is only meaningful if the landscape is so vast that this restriction still leaves a very large number of allowed values of the cosmological constant; yet the string theory landscape appears that it may be so vast. It is convenient to assume that the set of allowed values of the cosmological contant may be approximated as a continuous distribution.

The only environmental condition that seems important to our existence and is determined in part by the cosmological constant is the existence of non-linear gravitational structures. This is described above. To account for this quantitatively it is convenient to work within the Press-Schechter formalism [37]. In this picture, one looks at the probability for a randomly selected co-moving volume to have separated from the Hubble flow by some specified time. We are interested in whether non-linear structures ever form, so we take this time to be the infinite future. Whether a given volume ever separates from the Hubble flow is a function of the mass over-density z contained in that volume, calculated assuming linear density perturbation evolution, and evaluated at the infinite future. The quantity z is a Gaussian random variable root-mean-square amplitude $\sigma_\infty(\rho_\Lambda, \mu)$ that depends on both the energy density in cosmological cosntant ρ_Λ and the size of the enclosed volume, which we parameterize by the mass μ that the volume encloses. Meanwhile, for a volume to seperate from the Hubble flow requires $z \gtrsim 1.69$. Thus the likelihood that a volume enclosing at least mass μ will separate from the cosmic expansion is

$$F(\rho_\Lambda, \mu) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_\infty} \int_{1.69}^{\infty} \exp\left[-\frac{1}{2}\frac{z^2}{\sigma_\infty^2}\right] dz = \mathrm{erfc}\left[\frac{1.69}{\sqrt{2}\,\sigma_\infty}\right] , \tag{1.31}$$

where clearly erfc denotes the complementary error function, $\mathrm{erfc}(x) \equiv 2\pi^{-1/2} \int_x^\infty e^{-z^2} dz$.

An overdensity that seperates from the Hubble flow will eventually collapse and virialize. The percentage of over-densities that eventually virialize is a function of the enclosed mass μ because the root-mean-square amplitude of the initial density perturbations depends on μ. Whether a viralized over-density will collapse into a galaxy habitable to observers like us will depend on μ, but not on the cosmological constant. Usually it is assumed that there is some minimum over-density mass μ_0 which will form a habitable galaxy and all larger over-densities are equally suitable for life. Then if only the cosmological constant scans over the landscape, the number of observers in a given universe will be proportional to the mass fraction in suitable galaxies in that universe, which is proportional to F. Under this reasoning $F(\rho_\Lambda, \mu_0)$ gives the likelihood to for a value of the cosmological constant ρ_Λ to be measured. If μ_0 is chosen to coincide with the mass of the Milky Way galaxy, then about 5% of observers measure a value of the cosmological constant smaller than the value we measure.

The above explanation for the size of the cosmological constant is certainly tenuous. For one, it is inappropriate to assume that conditions appropriate for life are independent of μ aside from a cut-off at the Milky Way mass. Allowing for smaller galaxies will make the value of ρ_Λ that we measure more atypical of the full distribution. Furthermore, within string theory the landscape of meta-stable states appears to allow for more parameters than the cosmological constant to scan.

Any parameter on which σ_∞ depends will affect the distribution of ρ_Λ if it too is allowed to scan. One such possibility involves the scanning of the apparent Planck mass over meta-stable states in the fundamental theory. The apparent Planck mass may determine the abundance of baryons and dark matter, in addition to the amplitude of primordial density perturbations. Both of these quantities enter into σ_∞. Exploring this possibility is one of the subjects of Chapter 5.

Although the landscape explanation for the cosmological constant is tenuous, it must be admitted that it is among the most plausible explanations for the observation of dark energy. This, and that an enormous landscape appears to be a natural consequence of string theory, has motivated the investigation of other consequences of a landscape. One example is the study of how inflationary parameters may be selected within the multiverse, which led to the discovery of the so-called runaway inflation problem that is described and extended in Chapter 5 of this thesis. A second example is the subject of Chapter 6 of this thesis. This is the possibility that the flavor structure of the Standard Model might be discribed simply as random selection from the string theory landscape. We now provide some very brief background to this idea.

The paradigm of flavor in the Standard Model of particle physics involves the observed patterns among the elements of the coupling matrices to the following interactions:

$$\mathcal{L}_{\text{flavor}} = \lambda^u_{ij}\bar{u}^i\, q^j_L\, h + \lambda^d_{ij}\bar{d}^i\, q^j_L\, h^* + \lambda^e_{ij}\bar{e}^i\, l^j_L\, h^* + \frac{C_{ij}}{M}\, l^i_L\, l^j_L\, h\, h\,, \tag{1.32}$$

where q_L, l_L (u, d, e) are the left (right) handed quark and lepton fields, h is the Higgs, the indices i and j label generation and we have suppressed gauge symmetry indices. To provide a theoretical description of the spectrum of flavor parameters, i.e. the observable elements of the matrices $\lambda^{u,d,e}$ and C, is a major problem for theoretical physics. Since most of these elements are very small, the historical approach has been to presume some large symmetry that is softly broken in the Standard Model [38, 39]. However, progress invoking theoretically motivated models of approximate flavor symmetry (AFS) has been limited, while the AFS program in general suffers for the extreme flexibility in its ability to explain any possible flavor patterns.

In Chapter 6 an entirely different approach is suggested. It has already been proposed [40] that the observed hierarchy among Standard Model flavor parameters may be attributed to Standard Model Yukawa couplings deriving from overlap integrals involving the zero mode wavefunctions of Standard Model fields over a set of extra dimensions. Such a picture is motivated by the possibility that such interactions could arise due to gauge couplings in a more fundamental theory, after the higher-dimensional components of the gauge fields are integrated out in the low energy effective theory. In the original proposal the central locations of the Standard Model wavefunctions were arbitrarily chosen so as to reproduce the observed hierarchy. In Chapter 6 we demonstrate that this arbitrariness is not necessary. Specifically, for plausible geometries of the compactified extra dimensions, the specific flavor patterns of the Standard Model arise statistically in a landscape ensemble where the central locations of the Standard Model particle wavefunctions are allowed to sit at independent and random positions in the extra dimensional geometry. One needs only assume that the Standard Model particle wavefunctions are highly localized (for example, having Gaussian profile) and that discrepancies between the observed spectrum of flavor parameters and the most likely landscape values are due to random selection from the full statistical ensemble. The investigation of Chapter 6 restricts attention to studying flavor in the quark sector of the Standard Model; a study of the lepton sector is in preperation at the submission time of this thesis.

Chapter 5 is based on "The scale of gravity and the cosmological constant within a landscape," Michael L. Graesser and Michael P. Salem, arXiv:astro-ph/0611694 [41]. I initiated this project and I lead its direction. In addition, I was the first to perform all of the calculations and numerical analyses that are presented in the paper. Finally, I wrote the first draft of the manuscript and I wrote most of the subsequent editions.

Chapter 6 is based on "Quark and lepton masses from the landscape," Lawrence J. Hall, Michael P. Salem, and Taizan Watari, which at the submission time of these thesis is still in preparation [42]. With respect to the content of chapter 6, I wrote the code for some of the numerical analyses, I checked the code for all of the numerical analyses, and I optimized the parameters for the numerical analysis. I verified all of the calculations in the paper, and I performed-first about half of the calculations obtaining approximate analytic results for the flavor parameter distributions. I formulated the precise implementation of statistical tests for the landscape predictions, and I wrote Section 6.4. Finally, I performed extensive editing and reorganizing of the manuscript.

Chapter 2

Infrared Regulators and $\mathrm{SCET_{II}}$

We consider matching from $\mathrm{SCET_I}$, which includes ultrasoft and collinear particles, onto $\mathrm{SCET_{II}}$ with soft and collinear particles at one loop. Keeping the external fermions off their mass shell does not regulate all IR divergences in both theories. We give a new prescription to regulate infrared divergences in SCET. Using this regulator, we show that soft and collinear modes in $\mathrm{SCET_{II}}$ are sufficient to reproduce all the infrared divergences of $\mathrm{SCET_I}$. We explain the relationship between IR regulators and an additional mode proposed for $\mathrm{SCET_{II}}$.

Based on C. W. Bauer, M. P. Dorsten, and M. P. Salem, *Phys. Rev. D* **69**, 114011 (2004).

2.1 Introduction

Soft-collinear effective theory [1, 2, 3, 4] describes the interactions of soft and ultrasoft (usoft) particles with collinear particles. Using light-cone coordinates in which a general four-momentum is written as $p^\mu = (p^+, p^-, p^\perp) = (n{\cdot}p, \bar{n}{\cdot}p, p^\perp)$, where n and $\bar{n}$ are four-vectors on the light cone ($n^2 = \bar{n}^2 = 0$, $n \cdot \bar{n} = 2$), these three degrees of freedom are distinguished by the scaling of their momenta:

$$\begin{aligned} &\text{collinear:} && p_c^\mu \sim Q(\lambda^2, 1, \lambda), \\ &\text{soft:} && p_s^\mu \sim Q(\lambda, \lambda, \lambda), \\ &\text{usoft:} && p_{us}^\mu \sim Q(\lambda^2, \lambda^2, \lambda^2). \end{aligned} \tag{2.1}$$

The size of the expansion parameter λ is determined by the typical off-shellness of the collinear particles in a given problem. For example, in inclusive decays one typically has $p_c^2 \sim Q^2\lambda^2 \sim Q\Lambda_{\mathrm{QCD}}$, from which it follows that $\lambda = \sqrt{\Lambda_{\mathrm{QCD}}/Q}$. For exclusive decays, however, one needs collinear particles with $p_c^2 \sim \Lambda_{\mathrm{QCD}}^2$, giving $\lambda = \Lambda_{\mathrm{QCD}}/Q$. One is usually interested in describing the interactions of these collinear degrees of freedom with non-perturbative degrees of freedom at rest, which satisfy $p^\mu \sim (\Lambda_{\mathrm{QCD}}, \Lambda_{\mathrm{QCD}}, \Lambda_{\mathrm{QCD}})$. Thus inclusive processes involve interactions of collinear and usoft degrees of freedom, while exclusive decays are described by interactions of collinear and soft degrees of freedom. The theory describing the former set of degrees of freedom is called $\mathrm{SCET_I}$, while the latter theory is called $\mathrm{SCET_{II}}$ [5].

Interactions between usoft and collinear degrees of freedom are contained in the leading-order Lagrangian of $\mathrm{SCET_{I}}$,

$$\mathcal{L}_{\mathrm{I}} = \bar{\xi}_n \left[in\!\cdot\! D + i\not{D}_c^{\perp} \frac{1}{i\bar{n}\!\cdot\! D_c} i\not{D}_c^{\perp} \right] \frac{\not{\bar{n}}}{2} \xi_n \, , \tag{2.2}$$

and are well understood. The only interaction between collinear fermions and usoft gluons is from the derivative

$$iD^{\mu} = iD_c^{\mu} + gA_{us}^{\mu} \, . \tag{2.3}$$

These interactions can be removed from the leading-order Lagrangian by the field redefinition [4]

$$\xi_n = Y_n \xi_n^{(0)} \, , \qquad A_n = Y_n A_n^{(0)} Y_n^{\dagger} \, , \qquad Y_n(x) = \mathrm{P} \exp \left(ig \int_{-\infty}^{0} ds \, n\!\cdot\! A_{us}(x + ns) \right) \, . \tag{2.4}$$

However, the same field redefinition has to be performed on the external operators in a given problem, and this reproduces the interactions with the usoft degrees of freedom. Consider for example the heavy-light current, which in $\mathrm{SCET_{I}}$ is given by

$$J_{hl}^{I}(\omega) = \left[\bar{\xi}_n W_n \right]_{\omega} \Gamma h_v \, , \tag{2.5}$$

where h_v is the standard field of heavy quark effective theory [7], the Wilson line W_n is required to ensure collinear gauge invariance [3], and ω is the large momentum label of the gauge invariant $[\bar{\xi}_n W_n]$ collinear system. Written in terms of the redefined fields, this current is

$$J_{hl}^{I}(\omega) = \left[\bar{\xi}_n^{(0)} W_n^{(0)} \right]_{\omega} \Gamma \left[Y_n^{\dagger} h_v \right] \, . \tag{2.6}$$

For exclusive decays, we need to describe the interactions of soft with collinear particles. This theory is called $\mathrm{SCET_{II}}$ [5]. Since adding a soft momentum to a collinear particle takes this particle off its mass shell $(p_c + p_s)^2 \sim (Q\lambda, Q, Q\lambda)^2 \sim Q^2\lambda \sim Q\Lambda_{\mathrm{QCD}}$, there are no couplings of soft to collinear particles in the leading-order Lagrangian.[1] Thus, the Lagrangian is given by [6, 43, 44]

$$\mathcal{L}_{\mathrm{II}} = \bar{\xi}_n \left[in\!\cdot\! D_c + i\not{D}_c^{\perp} \frac{1}{i\bar{n}\!\cdot\! D_c} i\not{D}_c^{\perp} \right] \frac{\not{\bar{n}}}{2} \xi_n \, . \tag{2.7}$$

In this theory, the heavy-light current is given by

$$J_{hl}^{II}(\omega, \kappa) = \left[\bar{\xi}_n^{(0)} W_n^{(0)} \right]_{\omega} \Gamma \left[S_n^{\dagger} h_v \right]_{\kappa} \, , \tag{2.8}$$

---

[1] At higher orders, higher dimensional operators with at least two soft and two collinear particles can appear.

where S_n is a soft Wilson line in the n direction defined by

$$S_n(x) = \mathrm{P}\exp\left(ig\int_{-\infty}^{0} ds\, n\cdot A_s(x+ns)\right)\,. \tag{2.9}$$

This is the most general current invariant under collinear and soft gauge transformations.

This chapter is organized as follows: We first consider the matching of the heavy-light current in $\mathrm{SCET_I}$ onto the heavy-light current in $\mathrm{SCET_{II}}$ using off-shell fermions. While the terms logarithmic in the off-shellness do not agree in the two theories, we argue that this is due to unregulated IR divergences in $\mathrm{SCET_{II}}$. We then discuss IR regulators in SCET in more detail. We first identify the problems with SCET regulators and then propose a new regulator that addresses these issues. Using this regulator we then show that soft and collinear modes in $\mathrm{SCET_{II}}$ are sufficient to reproduce the IR divergences of $\mathrm{SCET_I}$ and explain the relationship between IR regulators and an additional mode proposed for $\mathrm{SCET_{II}}$ [43, 44].

2.2 Matching from $\mathrm{SCET_I}$ onto $\mathrm{SCET_{II}}$

The only difference between $\mathrm{SCET_I}$ and $\mathrm{SCET_{II}}$ is the typical off-shellness of the collinear degrees of freedom in the theory. The theory $\mathrm{SCET_I}$ allows fluctuations around the classical momentum with $p_c^2 \sim Q\Lambda_{\rm QCD}$, while the theory $\mathrm{SCET_{II}}$ allows fluctuations with only $p_c^2 \sim \Lambda_{\rm QCD}^2$. Since both theories expand around the same limit, $\mathrm{SCET_{II}}$ can be viewed as a low energy effective theory of $\mathrm{SCET_I}$. Therefore, one can match from the theory $\mathrm{SCET_I}$ onto $\mathrm{SCET_{II}}$ by integrating out the $\mathcal{O}(\sqrt{Q\Lambda_{\rm QCD}})$ fluctuations.

To illustrate this matching, we consider the heavy-light current. Using the definitions of this current given in Eqs. (2.5) and (2.8), we can write

$$J_{hl}^{I}(\omega) = \int d\kappa\; C(\omega,\kappa)\; J_{hl}^{II}(\omega,\kappa)\,. \tag{2.10}$$

At tree level one finds trivially $C(\omega,\kappa) = 1$. In fact, this Wilson coefficient remains unity to all orders in perturbation theory, as was argued in Ref. [45].

To determine the matching coefficient at one loop, we calculate matrix elements of the current in the two theories. There are two diagrams in $\mathrm{SCET_I}$, shown in Fig. 2.1. We use $\mathrm{SCET_I}$ without performing the field redefinitions of Eq. (2.4) so that the matching between $\mathrm{SCET_I}$ and $\mathrm{SCET_{II}}$ is not trivial. For on-shell external states, we find for the two integrals

$$\begin{aligned} iA^{Ia} &= g^2 C_F \mu^{4-d}\int \frac{\mathrm{d}^d k}{(2\pi)^d}\frac{1}{[-n\cdot k + i0][-v\cdot k + i0][k^2+i0]}\,, && (2.11)\\ iA^{Ib} &= 2g^2 C_F \mu^{4-d}\int \frac{\mathrm{d}^d k}{(2\pi)^d}\frac{\bar{n}\cdot(p_c - k)}{[-\bar{n}\cdot k + i0][k^2 - 2p_c\cdot k + i0][k^2+i0]}\,. && (2.12)\end{aligned}$$

Figure 2.1: Diagrams in $\mathrm{SCET_I}$ contributing to the matching. The solid square denotes an insertion of the heavy-light current.

Figure 2.2: Diagrams in $\mathrm{SCET_{II}}$ contributing to the matching.

The factor of $1/[-\bar{n}\cdot k+i0]$ in iA^{Ib} comes from the Wilson line that appears in the heavy-light current. Meanwhile, the diagrams in $\mathrm{SCET_{II}}$ are shown in Fig. 2.2. For on-shell external states the two integrals are exactly the same as in $\mathrm{SCET_I}$:

$$\begin{aligned} iA^{IIa} &= g^2 C_F \mu^{4-d}\int\frac{\mathrm{d}^d k}{(2\pi)^d}\frac{1}{[-n\cdot k+i0][-v\cdot k+i0][k^2+i0]}\,, \qquad (2.13)\\ iA^{IIb} &= 2g^2 C_F \mu^{4-d}\int\frac{\mathrm{d}^d k}{(2\pi)^d}\frac{\bar{n}\cdot(p_c-k)}{[-\bar{n}\cdot k+i0][k^2-2p_c\cdot k+i0][k^2+i0]}\,. \qquad (2.14)\end{aligned}$$

Since the integrands are exactly the same, the loop diagrams will precisely cancel in the matching calculation. Thus we find that the Wilson coefficient $C(\omega,\kappa)$ remains unity, even at one loop. This confirms the arguments in Ref. [45] to this order.

The fact that both of these integrals are scaleless and therefore that they integrate to zero might bother some readers. The vanishing of these diagrams is due to the cancellation of collinear, infrared (IR) and ultraviolet (UV) divergences. Introducing an IR regulator will separate these divergences, and the UV will be regulated by dimensional regularization. In Ref. [1] a small off-shellness was introduced to regulate the IR divergences of $\mathrm{SCET_I}$. In Refs. [43, 44] the divergence structure of $\mathrm{SCET_{II}}$ was studied keeping both the heavy and the collinear fermions off-shell. Using this IR regulator, the authors of Refs. [43, 44] argued that $\mathrm{SCET_{II}}$ does not reproduce the IR divergences of $\mathrm{SCET_I}$ and introduced a new mode in $\mathrm{SCET_{II}}$ to fix this problem. To gain more insight into their argument, we will go through their calculation in some detail.

In $\mathrm{SCET_I}$ the first diagram is

$$\begin{aligned}
A^{Ia}_{p_c} &= -ig^2C_F\mu^{4-d}\int\frac{\mathrm{d}^dk}{(2\pi)^d}\frac{1}{[\tilde{p}_c-n\cdot k+i0][v\cdot(p_s-k)+i0][k^2+i0]} \\
&= -\frac{g^2C_F}{2\pi}(4\pi)^{1-d/2}\Gamma\left(2-\frac{d}{2}\right)\mu^{4-d}\int_0^\infty \mathrm{d}n\cdot k\,(n\cdot k-\tilde{p}_c)^{-1}\,n\cdot k^{d/2-2}\,(n\cdot k-2v\cdot p_s)^{d/2-2} \\
&= \frac{\alpha_sC_F}{4\pi}\left[-\frac{1}{\epsilon^2}+\frac{2}{\epsilon}\log\frac{-\tilde{p}_c}{\mu}-2\log^2\frac{-\tilde{p}_c}{\mu}+2\log\left(1-\frac{2v\cdot p_s}{\tilde{p}_c}\right)\log\frac{2v\cdot p_s}{\tilde{p}_c}\right. \\
&\qquad\left.+2\,\mathrm{Li}_2\left(\frac{2v\cdot p_s}{\tilde{p}_c}\right)-\frac{3\pi^2}{4}\right], \qquad (2.15)
\end{aligned}$$

where $d=4-2\epsilon$ and $\tilde{p}_c=p_c^2/\bar{n}\cdot p_c$. In going from the first line to the second, we closed the $\bar{n}\cdot k$ contour below, thus restricting $n\cdot k$ to positive values, and performed the Euclidean $k_\perp$ integral. The second diagram gives

$$\begin{aligned}
A^{Ib}_{p_c} &= -2ig^2C_F\mu^{4-d}\int\frac{\mathrm{d}^dk}{(2\pi)^d}\frac{\bar{n}\cdot(p_c-k)}{[-\bar{n}\cdot k+i0][(k-p_c)^2+i0][k^2+i0]} \\
&= \frac{\alpha_sC_F}{4\pi}\left[\frac{2}{\epsilon^2}+\frac{2}{\epsilon}-\frac{2}{\epsilon}\log\frac{-p_c^2}{\mu^2}+\log^2\frac{-p_c^2}{\mu^2}-2\log\frac{-p_c^2}{\mu^2}+4-\frac{\pi^2}{6}\right]. \qquad (2.16)
\end{aligned}$$

In this diagram it is necessary to choose $d<4$ for the $k_\perp$ integral, but one requires $d>4$ for the $\bar{n}\cdot k$ integral. In the former integral, dimensional regularization regulates the divergence at $k_\perp=\infty$, while in the latter it regulates the divergence at $\bar{n}\cdot k=0$. Both of these divergences have to be interpreted as UV, as discussed in Section 2.3. In addition, each of the diagrams contains a mixed UV-IR divergence of the form $\log p_c^2/\epsilon$. This mixed divergence cancels in the sum of the two diagrams and we find, after also adding the wave function contributions,

$$\begin{aligned}
A^{I}_{p_c} &= \frac{\alpha_sC_F}{4\pi}\left[\frac{1}{\epsilon^2}+\frac{2}{\epsilon}\log\frac{\mu}{\bar{n}\cdot p_c}+\frac{5}{2\epsilon}+\log^2\frac{-p_c^2}{\mu^2}-2\log^2\frac{-\tilde{p}_c}{\mu}-\frac{3}{2}\log\frac{-p_c^2}{\mu^2}-2\log\frac{-2v\cdot p_s}{\mu}\right. \\
&\qquad\left.+2\log\left(1-\frac{2v\cdot p_s}{\tilde{p}_c}\right)\log\frac{2v\cdot p_s}{\tilde{p}_c}+2\mathrm{Li}_2\left(\frac{2v\cdot p_s}{\tilde{p}_c}\right)+\frac{11}{2}-\frac{11\pi^2}{12}\right]. \qquad (2.17)
\end{aligned}$$

This reproduces the IR behavior of full QCD.

Now consider the $\mathrm{SCET_{II}}$ diagrams. The second is identical to the one in $\mathrm{SCET_I}$: $A^{IIb}_{p_c}=A^{Ib}_{p_c}$.

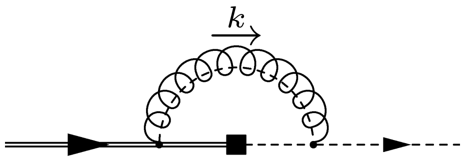


Figure 2.3: Contribution of the additional $\mathrm{SCET_{II}}$ mode proposed in Refs. [43, 44].

The first diagram, however, is different. For this amplitude we find

$$\begin{aligned} A^{IIa}_{p_c} &= -ig^2 C_F \mu^{4-d}\int \frac{\mathrm{d}^d k}{(2\pi)^d}\frac{1}{[-n\cdot k+i0][v\cdot(p_s-k)+i0][k^2+i0]} \\ &= -\frac{g^2 C_F}{2\pi}\mu^{4-d}\int_0^\infty \mathrm{d}n\cdot k\int\frac{\mathrm{d}^{d-2}k_\perp}{(2\pi)^{d-2}}\frac{1}{n\cdot k\,(k_\perp^2+n\cdot k^2-2v\cdot p_s\,n\cdot k)} \\ &= -\frac{\alpha_s C_F}{2\pi}(4\pi)^{2-d/2}\Gamma\left(2-\frac{d}{2}\right)\mu^{4-d}\int_0^\infty \mathrm{d}n\cdot k\, n\cdot k^{d/2-3}(n\cdot k-2v\cdot p_s)^{d/2-2}\,. \end{aligned} \tag{2.18}$$

Note that convergence of this integral at $n\cdot k=\infty$ requires $d<4$, whereas convergence at $n\cdot k=0$ requires $d>4$. Here dimensional regulation is regulating both a UV divergence at $n\cdot k=\infty$, as well as the divergence at $n\cdot k=0$, which is IR in nature, as we will discuss in Section 2.3. Using the variable transformation $x=n\cdot k/(n\cdot k-2v\cdot p_s)$ to relate this integral to a beta function [46] one finds

$$A^{IIa}_{p_c} = \frac{\alpha_s C_F}{4\pi}\left[\frac{1}{\epsilon^2}-\frac{2}{\epsilon}\log\frac{-2v\cdot p_s}{\mu}+2\log^2\frac{-2v\cdot p_s}{\mu}+\frac{5\pi^2}{12}\right]. \tag{2.19}$$

Adding the two diagrams together with the wave function contributions gives

$$\begin{aligned} &\frac{\alpha_s C_F}{4\pi}\left[\frac{3}{\epsilon^2}-\frac{2}{\epsilon}\log\frac{2v\cdot p_s\,p_c^2}{\mu^3}+\frac{5}{2\epsilon}+\log^2\frac{-p_c^2}{\mu^2}+2\log^2\frac{-2v\cdot p_s}{\mu}-\frac{3}{2}\log\frac{-p_c^2}{\mu^2}\right. \\ &\qquad\left.-2\log\frac{-2v\cdot p_s}{\mu}+\frac{11}{2}+\frac{\pi^2}{4}\right]. \end{aligned} \tag{2.20}$$

We can see that in the sum of the two diagrams the terms proportional to $\log p_c^2/\epsilon$ or $\log v\cdot p_s/\epsilon$ do not cancel as they did in $\mathrm{SCET_I}$. Furthermore, the finite terms logarithmic in p_c^2 or $v\cdot p_s$ do not agree with the corresponding terms in the $\mathrm{SCET_I}$ result. This fact prompted the authors of Refs. [43, 44] to conclude that $\mathrm{SCET_{II}}$ does not reproduce the IR divergences of $\mathrm{SCET_{II}}$ and that a new mode is needed in the latter effective theory. However, as we mentioned above, there are problems with IR effects in this diagram. In fact, as we will show in great detail in the next section, the off-shellness of the fermions does not regulate all IR divergences in this diagram. This means that the fact that the terms logarithmic in the fermion off-shellness do not agree between $\mathrm{SCET_I}$ and $\mathrm{SCET_{II}}$ does not imply that the IR divergences are not reproduced correctly since some $1/\epsilon$ poles are IR in origin.

We also calculate the diagram in $\mathrm{SCET_{II}}$ containing the additional mode proposed in Refs. [43, 44]. The new messenger mode has momenta scaling $p_{sc}^\mu \sim (\Lambda^2_{\mathrm{QCD}}/Q,\, \Lambda_{\mathrm{QCD}},\, \Lambda^{3/2}_{\mathrm{QCD}}/Q^{1/2})$. (Note that the invariant mass of this term satisfies $p_{sc}^2\sim\Lambda^3_{\mathrm{QCD}}/Q \ll \Lambda^2_{\mathrm{QCD}}$.) The diagram is shown in Fig. 2.3 and for its amplitude we find

$$\begin{aligned} A^{IIc}_{p_c} &= -2ig^2 C_F\mu^{4-d}\int\frac{\mathrm{d}^d k}{(2\pi)^d}\frac{1}{[\tilde{p}_c-n\cdot k+i0][2v\cdot p_s-\bar{n}\cdot k+i0][k^2+i0]} \\ &= \frac{\alpha_s C_F}{4\pi}\left[-\frac{2}{\epsilon^2}+\frac{2}{\epsilon}\log\frac{2v\cdot p_s\,\tilde{p}_c}{\mu^2}-\log^2\frac{2v\cdot p_s\,\tilde{p}_c}{\mu^2}-\frac{\pi^2}{2}\right]. \end{aligned} \tag{2.21}$$

Adding this term to the Eq. (2.20) cancels the terms proportional to $\log(2v\cdot p_s\, p_c^2/\mu^3)/\epsilon$ and we find

$$\begin{aligned} A^{II}_{p_c} &= \frac{\alpha_s C_F}{4\pi}\left[\frac{1}{\epsilon^2}+\frac{2}{\epsilon}\log\frac{\mu}{\bar{n}\cdot p_c}+\frac{5}{2\epsilon}+\log^2\frac{-p_c^2}{\mu^2}-\frac{3}{2}\log\frac{-p_c^2}{\mu^2}-2\log\frac{-2v\cdot p_s}{\mu}\right. \\ &\qquad \left. +2\log^2\left(\frac{-2v\cdot p_s}{\mu}\right)-\log^2\left(\frac{2v\cdot p_s\tilde{p}_c}{\mu^2}\right)+\frac{11}{2}-\frac{\pi^2}{4}\right]. \end{aligned} \tag{2.22}$$

This result does not agree with the $\mathrm{SCET_I}$ expression in Eq. (2.17). However, this is expected, since the off-shellness in $\mathrm{SCET_{II}}$ satisfies $\tilde{p}_c \ll v\cdot p_s$. In this limit the $\mathrm{SCET_I}$ result in Eq. (2.17) agrees with the result in Eq. (2.22).

2.3 Infrared Regulators in SCET

2.3.1 Problems with Known IR Regulators

One of the most important properties of $\mathrm{SCET_I}$ is the field redefinition given in Eq. (2.4), which decouples the usoft from the collinear fermions. It is the crucial ingredient for proving factorization theorems. Furthermore, only after performing this field redefinition is it simple to match from $\mathrm{SCET_I}$ onto $\mathrm{SCET_{II}}$, since one can identify the Wilson line Y_n in $\mathrm{SCET_I}$ with the Wilson line S_n in $\mathrm{SCET_{II}}$. However, it is a well known fact that field redefinitions only leave on-shell Green functions invariant [47]. Hence, the off-shellness of the collinear quark p_c^2 used to regulate the IR in $\mathrm{SCET_I}$ takes away our ability to perform this field redefinition. Since no field redefinition is performed on the heavy quark, one is free to give it an off-shellness.

IR divergences appear in individual diagrams, but they cancel in the set of diagrams contributing to a physical observable. More specifically, the IR divergences in virtual loop diagrams are cancelled against those in real emissions, which physically have to be included due to finite detector resolutions. From this it is obvious that the IR divergences in the heavy-light current originate from regions of phase space where either the gluon three-momentum $|\mathbf{k}|$ or the angle θ between the gluon and the light fermion goes to zero. Other divergences arise if the three-momentum of the gluon goes to infinity or θ goes to π. These divergences are UV. To check if the IR divergences match between the two theories one has to use an IR regulator that regulates all IR divergences in both theories. To get insight into the behavior of the three-momentum and the angle, it will be instructive to perform the required loop integrals by integrating over k_0 using the method of residues, and then integrating over the magnitude of the three-momentum and the solid angle. This will allow us to identify clearly the IR divergences as described above. Let us illustrate this method by showing that all $1/\epsilon$ divergences in the SCET$_\mathrm{I}$ one-loop calculation of the previous section are UV. For the first diagram we find

$$\begin{aligned} A^{Ia}_{p_c} &= -ig^2 C_F \mu^{4-d} \int \frac{\mathrm{d}^d k}{(2\pi)^d} \frac{1}{[\tilde{p}_c - n\cdot k + i0][v\cdot(p_s - k) + i0][k^2 + i0]} \\ &= -\frac{g^2 C_F}{2} \frac{\Omega_{d-2}}{(2\pi)^{d-1}} \mu^{4-d} \int_0^\infty \mathrm{d}|\mathbf{k}||\mathbf{k}|^{d-2} \int_{-1}^{1} \frac{\mathrm{d}\!\cos\theta \, \sin^{d-4}\theta}{(|\mathbf{k}|(1-\cos\theta) - \tilde{p}_c)\,(|\mathbf{k}| - v\cdot p_s)|\mathbf{k}|} \,. \end{aligned} \tag{2.23}$$

Performing the remaining integrals, we of course reproduce the result obtained previously, but this form demonstrates that all divergences from regions $|\mathbf{k}| \to 0$ and $(1-\cos\theta) \to 0$ are regulated by the infrared regulators and thus all $1/\epsilon$ divergences are truly UV.

The second diagram is

$$\begin{aligned} A^{Ib}_{p_c} &= -2ig^2 C_F \mu^{4-d} \int \frac{\mathrm{d}^d k}{(2\pi)^d} \frac{\bar{n}\cdot(p_c - k)}{[-\bar{n}\cdot k + i0][(p_c - k)^2 + i0][k^2 + i0]} \\ &= -2ig^2 C_F \left[I_1 + I_2 \right] \,, \end{aligned} \tag{2.24}$$

where I_1 and I_2 are the integrals with the $\bar{n}\cdot p$ and the $\bar{n}\cdot k$ terms in the numerator, respectively. The integral I_2 is standard and we find

$$I_2 = \frac{i}{16\pi^2} \left[\frac{1}{\epsilon} - \log\frac{-p_c^2}{\mu^2} + 2 \right] \,, \tag{2.25}$$

where ϵ regulates only UV divergences. For the first integral we again perform the k_0 integral by contours and we find

$$\begin{aligned} I_1 &= \frac{i\bar{n}\cdot p_c}{2} \frac{\Omega_{d-2}}{(2\pi)^{d-1}} \mu^{4-d} \int_0^\infty \mathrm{d}|\mathbf{k}||\mathbf{k}|^{d-2} \int_{-1}^{1} \mathrm{d}\!\cos\theta \, \sin^{d-4}\theta \\ &\quad \times \Bigg[-\frac{1}{\mathbf{k}^2(1+\cos\theta)[2|\mathbf{k}|(p_0 - |\mathbf{p}|\cos\theta) - p_c^2]} \\ &\qquad + \frac{1}{a[p_0 + a + |\mathbf{k}|\cos\theta][2p_0^2 + 2p_0 a - 2|\mathbf{k}||\mathbf{p}|\cos\theta - p_c^2]} \Bigg] \,, \end{aligned} \tag{2.26}$$

where $p_c = (p_0, \mathbf{p})$ and $a = \sqrt{\mathbf{k}^2 + \mathbf{p}^2 - 2|\mathbf{k}||\mathbf{p}|\cos\theta}$. From this expression we can again see that all IR singularities from $|\mathbf{k}| \to 0$ and $(1-\cos\theta) \to 0$ are regulated by the off-shellness, and all remaining divergences are UV. Note furthermore that in the limit $|\mathbf{k}| \to \infty$, with unrestricted θ, the two terms cancel each other, so that there is no usual UV divergence. This agrees with the fact that there are five powers of k in the denominator of the integrand in Eq. (2.24). However, in the limit $|\mathbf{k}| \to \infty$ with $|\mathbf{k}|(1+\cos\theta) \to 0$ the second term of Eq. (2.26) remains finite, whereas the first term develops a double divergence. Thus, it is this region of phase space that gives rise to the double pole in this diagram. The presence of the square roots makes the evaluation of the remaining integrals difficult, but we have checked that we reproduce the divergent terms of the result given in Eq. (2.16).

From the above discussion it follows that the off-shellness of the external fermions regulates all the IR divergences, and that the $1/\epsilon$ divergences all correspond to divergences of UV origin. The situation is different in $\mathrm{SCET}_{\mathrm{II}}$, since the off-shellness of the light quark does not enter diagram (a). In this case we find

$$\begin{aligned} A^{IIa}_{p_c} &= -ig^2C_F\mu^{4-d}\int\frac{\mathrm{d}^dk}{(2\pi)^d}\frac{1}{[-n\cdot k+i0][v\cdot(p_s-k)+i0][k^2+i0]} \\ &= -\frac{g^2C_F}{2}\frac{\Omega_{d-2}}{(2\pi)^{d-1}}\mu^{4-d}\int_0^\infty\mathrm{d}|\mathbf{k}||\mathbf{k}|^{d-2}\int_{-1}^1\frac{\mathrm{d}\cos\theta\,\sin^{d-4}\theta}{\mathbf{k}^2(1-\cos\theta)(|\mathbf{k}|-v\cdot p_s)}\,. \end{aligned} \tag{2.27}$$

The IR divergence originating from the limit $(1-\cos\theta)\to 0$ is not regulated by the off-shellness. Thus part of the $1/\epsilon$ divergences in Eq. (2.19) are of IR origin. In other words, the fact that the terms logarithmic in the off-shellness in the $\mathrm{SCET}_{\mathrm{I}}$ amplitude Eq. (2.17) are not reproducing the corresponding terms in the $\mathrm{SCET}_{\mathrm{II}}$ amplitude Eq. (2.20) does not imply that the IR divergences do not match between the two theories. In order to check whether the IR divergences of the two theories match, one needs a regulator that regulates all IR divergences in both $\mathrm{SCET}_{\mathrm{I}}$ and $\mathrm{SCET}_{\mathrm{II}}$.

As an alternative IR regulator one could try to use a small gluon mass. Consider the first diagram in $\mathrm{SCET}_{\mathrm{I}}$ again, this time with a gluon mass. We find

$$\begin{aligned} A^{Ia}_{m} &= -ig^2C_F\mu^{4-d}\int\frac{\mathrm{d}^dk}{(2\pi)^d}\frac{1}{[-n\cdot k+i0][v\cdot(p_s-k)+i0][k^2-m^2+i0]} \\ &= -\frac{g^2C_F}{2}\frac{\Omega_{d-2}}{(2\pi)^{d-1}}\mu^{4-d}\int_0^\infty\mathrm{d}|\mathbf{k}||\mathbf{k}|^{d-2}\int_{-1}^1\mathrm{d}\cos\theta\,\sin^{d-4}\theta \\ &\qquad\times\frac{1}{(\mathbf{k}^2+m^2-v\cdot p_s\sqrt{\mathbf{k}^2+m^2})(\sqrt{\mathbf{k}^2+m^2}-|\mathbf{k}|\cos\theta)}\,. \end{aligned} \tag{2.28}$$

Again, all divergences $|\mathbf{k}|\to 0$ and $(1-\cos\theta)\to 0$ are regulated by the gluon mass, but in the limit $|\mathbf{k}|\to\infty$ with $|\mathbf{k}|(1-\cos\theta)\to 0$ the integrand becomes

$$\frac{|\mathbf{k}|^{d-4}\sin^{d-4}\theta}{|\mathbf{k}|(1-\cos\theta)+\frac{m^2}{2|\mathbf{k}|}}\,, \tag{2.29}$$

so that the term that could potentially regulate the $(1-\cos\theta)\to 0$ divergence goes to zero as $|\mathbf{k}|\to\infty$. This is why a gluon mass cannot be used to regulate the IR of SCET.

2.3.2 A New Regulator for SCET

The gluon mass is not an appropriate IR regulator for SCET because it appears in the combination $m^2/|\mathbf{k}|$ in the expression Eq. (2.29). Instead of using a gluon mass, consider adding the terms

$$\mathcal{L}^c_{\mathrm{reg}} = -\frac{\delta}{2}A^c_\mu\bar{\mathcal{P}}A^\mu_c\,,\qquad \mathcal{L}^{(u)s}_{\mathrm{reg}} = -\frac{\delta}{2}A^{(u)s}_\mu i\bar{n}\cdot\partial A^\mu_{(u)s} \tag{2.30}$$

to the collinear and (u)soft gluon Lagrangians. Here, $\bar{\mathcal{P}}$ is the label operator which picks out the large momentum label of the collinear gluon field. Both of these terms are generated if a similar term is added to the full QCD gluon action before constructing SCET. Note that these terms preserve the invariance of the theory under the field redefinitions given in Eq. (2.4).

We here note that an alternative regulator has previously been introduced in Ref. [48]. In that paper a quark mass is used in conjunction with an "analytic" regulator, which regulates the $(1-\cos\theta) \to 0$ divergence. The conclusions about the soft-collinear mode in Ref. [43, 44] are similar to the ones drawn here. However, we believe that a regulator such as the one introduced here is advantageous, since it can naturally be defined at the level of the Lagrangian, and a single dimensionful parameter regulates all IR divergences.

The infinitesimal, dimensionful parameter δ suffices to regulate all IR divergences in SCET, unlike the gluon mass. Following the same steps as in Eq. (2.28). We find

$$A_\delta^{Ia} \;=\; -\frac{g^2 C_F}{2}\frac{\Omega_{d-2}}{(2\pi)^{d-1}}\mu^{4-d}\int_0^\infty \mathrm{d}|\mathbf{k}||\mathbf{k}|^{d-2}\int_{-1}^{1}\frac{8\,\mathrm{d}\cos\theta\,\sin^{d-4}\theta}{b(\delta+b)(\delta+b-2|\mathbf{k}|\cos\theta)}\,, \tag{2.31}$$

where we have introduced

$$b=\sqrt{4\mathbf{k}^2+\delta^2+4|\mathbf{k}|\delta\cos\theta}\,. \tag{2.32}$$

Obviously, the parameter δ regulates the divergences $|\mathbf{k}| \to 0$ and $(1-\cos\theta) \to 0$, just as the gluon mass did. Expanding around the limit $|\mathbf{k}| \to \infty$ with $|\mathbf{k}|(1-\cos\theta) \to 0$ the integrand becomes

$$\frac{|\mathbf{k}|^{d-4}\sin^{d-4}\theta}{|\mathbf{k}|(1-\cos\theta)+\delta}\,, \tag{2.33}$$

and this IR region is therefore regulated as well. Even though δ is enough to regulate all IR divergences in SCET, we will keep the heavy quark off its mass-shell for later convenience.

Performing the integrals using the method above is difficult. While performing the k_0 integration using the method of residues gives insight into the divergence structure of the loop integrals, it is simpler to perform the integrals using the variables $n\cdot k$ and $\bar{n}\cdot k$ instead. The first diagram in $\mathrm{SCET_I}$ with this new regulator is then given by

$$\begin{aligned}
A_\delta^{Ia} &= -ig^2C_F\mu^{4-d}\int\frac{\mathrm{d}^dk}{(2\pi)^d}\frac{1}{[-n\cdot k+i0][v\cdot(p_s-k)+i0][k^2-\delta\bar{n}\cdot k+i0]}\\
&= -\frac{g^2C_F}{2\pi}(4\pi)^{1-d/2}\Gamma\left(2-\frac{d}{2}\right)\mu^{4-d}\int_\delta^\infty \mathrm{d}n\cdot k\; n\cdot k^{-1}\left(n\cdot k-\delta\right)^{d/2-2}\left(n\cdot k-2v\cdot p_s\right)^{d/2-2}\\
&= \frac{\alpha_sC_F}{4\pi}\left[-\frac{1}{\epsilon^2}+\frac{2}{\epsilon}\log\frac{\delta}{\mu}-2\log^2\frac{\delta}{\mu}+2\log\left(1-\frac{2v\cdot p_s}{\delta}\right)\log\frac{2v\cdot p_s}{\delta}\right.\\
&\qquad\left.+2\,\mathrm{Li}_2\left(1-\frac{2v\cdot p_s}{\delta}\right)-\frac{3\pi^2}{4}\right]\,.
\end{aligned} \tag{2.34}$$

Similarly, it is possible to show that the parameter δ regulates all IR divergences in the second diagram, for which we find

$$\begin{aligned} A_\delta^{Ib} &= -2ig^2C_F\mu^{4-d}\int\frac{\mathrm{d}^dk}{(2\pi)^d}\frac{\bar{n}\cdot(p_c-k)}{[-\bar{n}\cdot k+i0][k^2-2k\cdot p_c+i0][k^2-\delta\bar{n}\cdot k+i0]} \\ &= \frac{\alpha_sC_F}{4\pi}\left[\frac{2}{\epsilon^2}+\frac{2}{\epsilon}-\frac{2}{\epsilon}\log\frac{\delta\,\bar{n}\cdot p_c}{\mu^2}+\log^2\frac{\delta\,\bar{n}\cdot p_c}{\mu^2}-2\log\frac{\delta\,\bar{n}\cdot p_c}{\mu^2}+4-\frac{\pi^2}{6}\right]. \end{aligned} \tag{2.35}$$

The mixed UV-IR divergences cancel in the sum of the two diagrams,

$$\begin{aligned} A_\delta^{I} &= \frac{\alpha_sC_F}{4\pi}\left[\frac{1}{\epsilon^2}+\frac{2}{\epsilon}\log\frac{\mu}{\bar{n}\cdot p_c}+\frac{5}{2\epsilon}+\log^2\frac{\delta\bar{n}\cdot p_c}{\mu^2}-2\log^2\frac{\delta}{\mu}-\frac{3}{2}\log\frac{\delta\bar{n}\cdot p_c}{\mu^2}\right. \\ &\qquad -2\log\frac{\delta-2v\cdot p_s}{\mu}+2\log\left(1-\frac{2v\cdot p_s}{\delta}\right)\log\frac{2v\cdot p_s}{\delta} \\ &\qquad \left.+2\mathrm{Li}_2\left(1-\frac{2v\cdot p_s}{\delta}\right)+\frac{11}{2}-\frac{11\pi^2}{12}\right], \end{aligned} \tag{2.36}$$

and one can show that this result reproduces the IR behavior of full QCD.

Since the regulator is in the gluon action, it is the same for $\mathrm{SCET_I}$ and $\mathrm{SCET_{II}}$, and the two diagrams in $\mathrm{SCET_{II}}$ are identical to those in $\mathrm{SCET_I}$ since the integrands are exactly equal:

$$A_\delta^{II}=A_\delta^{I}\,. \tag{2.37}$$

Therefore, the IR divergences in $\mathrm{SCET_{II}}$ are exactly the same as those in $\mathrm{SCET_I}$.

While in $\mathrm{SCET_I}$ it is possible to choose the scaling $\delta\sim Q\lambda^2$ such that both the contributions to the collinear and the usoft gluon action are leading order in the power counting, the same is not true in $\mathrm{SCET_{II}}$. Choosing $\delta\sim Q\lambda^2$ to make the IR regulator leading order in collinear gluon Lagrangian makes it suppressed by one power of λ in the soft Lagrangian. This can be understood physically, since in going from $\mathrm{SCET_I}$ to $\mathrm{SCET_{II}}$ the typical scaling of the (u)soft momenta remains of order Λ_{QCD}, while the off-shellness of the collinear particles is lowered. However, the IR divergence from $n\cdot k\to 0$ corresponds to $(1-\cos\theta)\to 0$, and the typical cutoff on $(1-\cos\theta)$ is set by the collinear scales. Since $n\cdot k_c\ll n\cdot k_s$ it is natural that any cutoff κ regulating the $n\cdot k_s\to 0$ divergence will satisfy $\kappa\ll n\cdot k_s$. This is not a problem, since the IR regulator does not introduce a new scale into the effective theory.

If one insists on keeping the scaling manifest, one is forced to drop the regulator term in the soft gluon Lagrangian. In this case, the diagram (a) in $\mathrm{SCET_{II}}$ no longer includes the IR regulator δ and is therefore not regulated properly. The calculation then reduces to the result given in Eq. (2.19). Part of the $1/\epsilon$ divergences in this result are from true UV divergences, but others are due to the unregulated $(1-\cos\theta)\to 0$ IR divergences, which arise from physics at the scale $n\cdot k\sim Q\lambda^2$. These IR divergences can be recovered by adding a diagram containing a gluon scaling as $n\cdot k\sim Q\lambda^2$,

$\bar{n}\cdot k \sim Q\lambda$. Requiring $n\cdot k\,\bar{n}\cdot k\ \sim k_\perp^2$, this is the soft collinear messenger mode introduced in [43, 44]. The resulting diagram (c) gives

$$\begin{aligned} A_\delta^{IIc} &= -2ig^2 C_F \mu^{4-d}\int\frac{\mathrm{d}^d k}{(2\pi)^d}\frac{1}{[-n\cdot k+i0][2v\cdot p_s-\bar{n}\cdot k+i0][k^2-\delta\bar{n}\cdot k+i0]} \\ &= \frac{\alpha_s C_F}{4\pi}\left[-\frac{2}{\epsilon^2}+\frac{2}{\epsilon}\log\frac{-2v\cdot p_s\,\delta}{\mu^2}-\log^2\frac{-2v\cdot p_s\,\delta}{\mu^2}-\frac{\pi^2}{2}\right]. \end{aligned} \tag{2.38}$$

Adding all the diagrams we find

$$\begin{aligned} &\frac{\alpha_s C_F}{4\pi}\left[\frac{1}{\epsilon^2}+\frac{2}{\epsilon}\log\frac{\mu}{\bar{n}\cdot p_c}+\frac{5}{2\epsilon}+\log^2\frac{\delta\bar{n}\cdot p_c}{\mu^2}-\frac{3}{2}\log\frac{\delta\bar{n}\cdot p_c}{\mu^2}-2\log\frac{-2v\cdot p_s}{\mu}\right. \\ &\qquad\left.+2\log^2\left(\frac{-2v\cdot\,p_s}{\mu}\right)-\log^2\left(\frac{-2v\cdot\,p_s\delta}{\mu^2}\right)+\frac{11}{2}-\frac{\pi^2}{4}\right], \end{aligned} \tag{2.39}$$

which again reproduces the $\mathrm{SCET_I}$ result for $\delta \ll v\cdot p_s$. From this discussion it follows that the presence of the soft collinear messenger mode depends on the precise implementation of the IR regulator in the theory. Since the definition of an effective theory should be independent of the regulator used for an explicit calculation, one can view the soft-collinear messenger mode as part of the IR regulator.

The term added to the gluon Lagrangian breaks gauge invariance. However, in this regard it is on the same footing as a gluon mass. Since the coupling of gluons to fermions is via a conserved current, this breaking of gauge invariance is only a problem once gluon self-interactions are taken into account. For the renormalization of operators such as the heavy-light current considered in this work, this only arises at the two-loop level. In matching calculations, the IR divergences always cancel. Hence any IR regulator, including the one proposed here, is applicable to matching calculations at any order in perturbation theory. More care has to be taken when using this regulator to calculate operator mixing, and in this case gauge non-invariant operators have to be included beyond one loop order. The main advantage of the new regulator is that it preserves invariance of $\mathrm{SCET_I}$ under the field redefinition given in Eq. (2.4).

2.4 Conclusions

We have considered the matching of the heavy-light current in $\mathrm{SCET_I}$ onto the corresponding current in $\mathrm{SCET_{II}}$, in particular addressing the question whether all of the long distance physics in $\mathrm{SCET_I}$ is correctly reproduced in $\mathrm{SCET_{II}}$. Using the off-shellness of the external heavy and light fermions, it was argued in Refs. [43, 44] that a new collinear-soft messenger mode is required in $\mathrm{SCET_{II}}$ to reproduce all the long distance physics of $\mathrm{SCET_I}$. Regulating the IR divergences in $\mathrm{SCET_{II}}$ with an off-shellness is problematic, since the off-shellness prevents performing the field redefinition required to decouple the usoft gluons from the collinear particles, which allows the matching onto

$\mathrm{SCET_{II}}$ easily. In this work we investigated the relationship between IR regulators and the definition of $\mathrm{SCET_{II}}$. By performing the k_0 loop integral by contours and then writing the remaining integrals as $\mathrm{d}|\mathbf{k}|\,\mathrm{d}\cos\theta$, we showed explicitly that the off-shellness leaves the IR angular divergence $(1-\cos\theta)\to 0$ unregulated in $\mathrm{SCET_{II}}$.

We then introduced a new regulator for SCET that regulates soft ($|\mathbf{k}|\to 0$) and collinear ($\cos\theta\to 1$) IR divergences in both $\mathrm{SCET_{I}}$ and $\mathrm{SCET_{II}}$. This regulator modifies the gluon action, much like a gluon mass, and thus preserves the field redefinition required to decouple usoft gluons from collinear particles in SCET. Using this regulator, we showed explicitly that $\mathrm{SCET_{II}}$ as formulated in Refs. [4, 45] reproduces all the IR divergences of $\mathrm{SCET_{I}}$. We also argued that any cutoff κ regulating the collinear divergence has to satisfy $\kappa \ll n\cdot k_s$. Regulating $\mathrm{SCET_{II}}$ this way therefore naturally requires keeping a formally subleading regulator in the theory.

We also showed that a soft-collinear messenger mode is required in the definition of the IR regulator if one insists on power counting the regulator in the same way as kinetic terms in the action. In this case, there are unregulated IR divergences left in soft diagrams, which are corrected by additional contributions from the soft-collinear mode.

The new regulator introduced in this work preserves the invariance of $\mathrm{SCET_{I}}$ under the field redefinitions of Eqs. (2.4), and is therefore useful in studying factorization theorems beyond tree level.

Acknowledgments

We would like to thank Thomas Becher, Richard Hill, Bjorn Lange, Michael Luke, Matthias Neubert, David Politzer, Ira Rothstein, Iain Stewart, and Mark Wise for useful discussions. This work was supported by the Department of Energy under the contract DE-FG03-92ER40701.

Chapter 3

Fluctuating Annihilation Cross Sections and the Generation of Density Perturbations

Fluctuations in the mass and decay rate of a heavy particle which for some period dominates the energy density of the universe are known to lead to adiabatic density perturbations. We show that generically the annihilation cross section of the same particle also receives fluctuations, which leads to entropy perturbations at freeze-out. If the particle comes to dominate the energy density of the universe and subsequently decays, this leads to an additional source of adiabatic density perturbations. On the other hand, non-adiabatic density perturbations result when the particle does not decay but contributes to the observed dark matter.

Based on C. W. Bauer, M. L. Graesser, and M. P. Salem, *Phys. Rev. D* **72**, 023512 (2005).

3.1 Introduction

Measurements of the cosmic microwave background radiation [49, 50, 51] have revealed a highly uniform energy density background with super-horizon perturbations on the order of one part in 10^5. In the standard inflationary paradigm (for reviews, see for example Refs. [9, 52, 10]), these density perturbations were created in the inflationary epoch when quantum fluctuations of the inflaton field expanded beyond the Hubble radius and were converted into density perturbations upon inflaton decay. However, to obtain the observed level of density perturbations from this mechanism requires tight constraints on the inflaton potential [53].

Recently, Dvali, Gruzinov, Zaldarriaga and independently Kofman (DGZK) proposed a new mechanism [18, 19, 20] for producing density perturbations. A nice feature of their scenario is that the only requirements on the inflaton potential are to produce the required e-foldings of inflation and at a scale consistent with WMAP data. The DGZK mechanism posits the existence of some heavy particle S with a mass and decay rate that depend on the vacuum expectation value of some light field χ. Here χ is presumed to have acquired super-horizon fluctuations during the inflationary epoch; however χ never contributes significantly to the energy density of the universe Nevertheless, the fluctuations in χ persist and result in fluctuations in the mass and decay rate of S, so long as the χ mass m_χ is less than the Hubble rate H at the time at which fluctuations are transferred to radiation. In the DGZK mechanism the field S comes to dominate the energy density of the universe and decays into radiation while $m_\chi < H$. Fluctuations in the mass and decay rate of S result in fluctuations in the duration of S energy domination, which in turn lead to adiabatic density perturbations since the energy of a massive S field redshifts more slowly than that of radiation.

The DGZK mechanism is closely related to the previously proposed "curvaton" scenario [22, 23, 24]. In the curvaton mechanism, the field χ is not coupled to any other fields. Instead, well after inflation has ended, χ itself comes to dominate the energy density of the universe. The fluctuations in χ modulate the duration of χ domination which leads to adiabatic density perturbations on χ decay. This work of this chapter is not concerned with this possibility.

The DGZK mechanism has been studied extensively. For example, the evolution of the density perturbations that result from this mechanism has been studied in detail using gauge invariant formalisms in Refs. [54, 55, 56]. These perturbations are shown to possess a highly scale invariant spectrum in Ref. [57] and are shown to contain significant non-Gaussianities in Ref. [58]. The original DGZK mechanism has also been extended to apply to preheating as studied in Ref. [59]. For discussions of the limitations of this mechanism see for example Refs. [60, 61, 62].

In the original DGZK scenario [18, 19] it is assumed that S decouples while being relativistic. In this work we generalize this to apply to the case where S freezes-out of equilibrium with a fluctuating annihilation rate $\langle\sigma v\rangle$. We use the term "freeze-out" to refer specifically to the scenario where S decouples from thermal equilibrium after it has become non-relativistic. In this case the number density of S at a temperature T after freeze-out is

$$n_S \simeq \frac{T^3}{m_S m_{\rm pl}\langle\sigma v\rangle}\,, \tag{3.1}$$

where $m_{\rm pl}$ is the Planck mass and m_S is the mass of S. Therefore we expect fluctuations in the mass and annihilation rate of S during freeze-out to result in fluctuations in the number density of S. If S lives long enough to dominate the energy density of the universe and subsequently decays, these entropy perturbations are converted into adiabatic perturbations. These add to the ones produced by the original DGZK mechanism and the quantum fluctuations of the inflaton.

The remainder of this chapter is organized as follows. In Section 3.2 we describe the density perturbations produced by our generalized DGZK mechanism. Sections 3.3 and 3.4 contain explicit models for implementing our mechanism and for producing the fluctuating masses and coupling constants, respectively. In Section 3.5 an alternate analytical description is given which allows to track the evolution of the perturbations, while in Section 3.6 Boltzmann equations are derived and solved numerically to confirm the analytical arguments presented in other sections of this chapter. Conclusions are given in Section 3.7.

3.2 Analytical Determination of the Perturbations

Our generalized DGZK mechanism includes a heavy particle S with mass m_S, decay rate Γ and annihilation cross section $\langle \sigma v \rangle$, where S decays to and interacts with radiation. We begin by identifying several key temperature scales. The temperature at which S begins to thermalize with radiation is denoted as $T_{\rm therm}$. We assume for simplicity that S particles are produced only as they thermalize from radiation annihilation below $T = T_{\rm therm}$. We also define:

1. $T_{\rm f.o.}$: Temperature at which S freezes-out of thermal equilibrium;
2. $T_{\rm dom}$: Temperature at which S begins to dominate the energy density of the universe;
3. $T_{\rm dec}$: Temperature at which S decays.

Since the number density of S particles falls off exponentially after S becomes non-relativistic, $T_{\rm f.o.}$ is typically within an order of magnitude of m_S. Therefore in this work we always take $T_{\rm f.o.} \simeq m_S$. In terms of m_S, Γ and $\langle \sigma v \rangle$ we also find

$$T_{\rm dom} \simeq \frac{1}{m_{\rm pl} \langle \sigma v \rangle} \,, \qquad T_{\rm dec} \simeq m_{\rm pl}\, \Gamma^{2/3} \langle \sigma v \rangle^{1/3} \,, \tag{3.2}$$

where we have assumed $T_{\rm dec} < T_{\rm dom}$ in the last equation. This condition is necessary for significant density perturbations to be produced by this mechanism. In Eq. (3.2) the cross section is to be evaluated at the freeze-out temperature $T_{\rm f.o.}$. Note that for S particles to be produced in the first place we require $T_{\rm therm} > T_{\rm f.o.}$.

As described in Refs. [18, 19], the period of S domination between $T_{\rm dom}$ and $T_{\rm dec}$ gives rise to an enhancement of the resulting energy density compared to a scenario where the S domination is absent. Comparing energy densities at common scale factor one finds that after S decays

$$\rho \;=\; \left(\frac{\rho_{\rm dom}}{\rho_{\rm dec}} \right)^{1/3} \rho_{\rm rad} \,, \tag{3.3}$$

where

$$\rho_{\rm dom} \;\simeq\; T_{\rm dom}^4 \,, \qquad \rho_{\rm dec} \simeq \frac{T_{\rm dec}^3}{m_{\rm pl} \langle \sigma v \rangle} \,, \tag{3.4}$$

and $\rho_{\rm rad}$ is the energy density which would result without any period of matter domination. As discussed in detail in Section 3.4, couplings to an additional field χ can give rise to fluctuations in

m_S, Γ, and $\langle\sigma v\rangle$:

$$m_S = \overline{m}_S\,(1+\delta_m)\ , \qquad \langle\sigma v\rangle = \langle\overline{\sigma v}\rangle\left(1+\delta_{\langle\sigma v\rangle}\right)\ , \qquad \Gamma = \overline{\Gamma}\,(1+\delta_\Gamma)\ , \tag{3.5}$$

where the barred quantities refer to background values. According to Eqs. (3.2–3.5), these fluctuations give rise to fluctuations in $T_{\rm dom}$ and $T_{\rm dec}$ which result in energy density perturbations

$$\frac{\delta\rho}{\rho} \;=\; -\frac{2}{3}\,\delta_\Gamma - \frac{4}{3}\,\delta_{\langle\sigma v\rangle}\,. \tag{3.6}$$

Note that although $\delta\rho/\rho$ contains no explicit dependence on δ_m, both $\delta_{\langle\sigma v\rangle}$ and δ_Γ are in general functions of δ_m.

Comparing the energy density at a common scale factor corresponds to choosing a gauge where the perturbation in the scale factor vanishes, $\psi = 0$. Thus the fluctuation in the energy density computed here can be directly related to the gauge invariant Bardeen parameter [16, 17]

$$\zeta \;=\; -\psi + \frac{\delta\rho}{3(\rho+p)}\,. \tag{3.7}$$

Thus we find after S decays

$$\zeta = -\frac{1}{6}\,\delta_\Gamma - \frac{1}{3}\,\delta_{\langle\sigma v\rangle}\,. \tag{3.8}$$

We can obtain the same result in synchronous gauge, where different regions all have the same global time. Since $\rho \sim 1/t^2$ in both matter and radiation dominated universes, one finds that $\delta\rho = 0$ on surfaces of constant time. Thus the Bardeen parameter is

$$\zeta = -\psi = \frac{\delta a}{a}\,. \tag{3.9}$$

To obtain ζ, we only need to determine $a(t, \Gamma, \langle\sigma v\rangle, m)$ and then compare two regions at fixed t, but different Γ, $\langle\sigma v\rangle$ and m_S. Assuming the S particles freeze-out while non-relativistic and decay after dominating the energy density of the universe, this gives

$$a(t) \;=\; \frac{a(t)}{a(t_{\rm dec})}\frac{a(t_{\rm dec})}{a(t_{\rm dom})}\frac{a(t_{\rm dom})}{a(t_{\rm f.o.})}\frac{a(t_{\rm f.o.})}{a(t_0)}a(t_0) \;=\; \left(\frac{t}{t_0}\right)^{1/2}\left(\frac{t_{\rm dec}}{t_{\rm dom}}\right)^{1/6} a(t_0)\,, \tag{3.10}$$

where $t_{\rm dec} \simeq \Gamma^{-1}$ is the time when S decays, $t_{\rm dom} \simeq m_{\rm pl}^3\langle\sigma v\rangle^2$ is the time at which it dominates the energy density of the universe, and $t_{\rm f.o.}$ is the time at which it freezes-out. Substituting gives

$$a(t) \;=\; \left(\frac{t}{t_0}\right)^{1/2} m_{\rm pl}^{-1/2}\Gamma^{-1/6}\langle\sigma v\rangle^{-1/3}a(t_0)\,. \tag{3.11}$$

Using this result and Eq. (3.9) we again obtain Eq. (3.8).

The above discussion is approximate and requires that S completely dominates the energy density of the universe. Obtaining the perturbations when S does not dominate requires that we include the matter contribution to the scale factor or energy density during radiation domination. This is done in Section 3.5 using a different formalism. In Section 3.6 we confirm these analytic results using a numerical calculation of the density perturbations using Boltzmann equations.

3.3 Explicit Models for Coupling S to Radiation

It is important to verify that models exist which exhibit the features discussed in the previous section. We present two models in which the annihilation cross section is determined by renormalizable and non-renormalizable operators, respectively.

The first model is given by the Lagrangian

$$\mathcal{L} \;=\; \sqrt{-g}\left[\frac{(\partial_\mu S)^2}{2} + \frac{(\partial_\mu X)^2}{2} - \frac{m_S^2}{2}S^2 - \frac{m_X^2}{2}X^2 - \frac{g\,m_S}{2}\,S\,X^2 - \frac{\lambda}{4}\,S^2X^2\right]. \tag{3.12}$$

We assume that X is in thermal equilibrium with the remaining radiation and that S particles are only produced through their coupling to X. The interaction terms in the above model yield an S decay rate and cross section

$$\Gamma \;\sim\; g^2 m_S\,, \qquad \langle\sigma v\rangle \;\sim\; \frac{\lambda^2}{M^2}\,, \tag{3.13}$$

where

$$M \;\simeq\; \begin{cases} T & \text{when } T > m_S\,, \\ m_S\,, & \text{when } T < m_S\,. \end{cases} \tag{3.14}$$

Note that we neglect the $\mathcal{O}(g^4)$ contribution to the cross section. This is justified given the limits on the coupling constants derived below.

The requirement that $T_{\text{therm}} > m_S$ and that S remains in thermal equilibrium down to $T \simeq m_S$ gives the condition on the coupling λ

$$\lambda \;>\; \sqrt{\frac{m_S}{m_{\text{pl}}}}\,. \tag{3.15}$$

On the other hand the condition $T_{\text{dec}} < T_{\text{dom}}$ implies

$$g^2\lambda^4 \;<\; \frac{m_S^3}{m_{\text{pl}}^3}\,. \tag{3.16}$$

Thus a necessary (but not sufficient) condition on g to satisfy both Eq. (3.15) and Eq. (3.16) is

$$g \quad < \quad \sqrt{\frac{m_S}{m_{\rm pl}}}\,. \tag{3.17}$$

Finally, we require that the period of S domination does not disrupt big bang nucleosynthesis (BBN). Thus the decay of S must reheat the universe to a temperature $T_{\rm rh} > T_{\rm BBN}$, where $T_{\rm rh} \simeq \sqrt{\Gamma m_{\rm pl}}$. This gives the constraint

$$g^2 \quad > \quad \frac{T_{\rm BBN}^2}{m_S m_{\rm pl}}\,. \tag{3.18}$$

Using $T_{\rm BBN} \simeq 10^{-21} m_{\rm pl}$, the above relations provide the constraint $m_S \gtrsim 10^{-21} m_{\rm pl}$. Given any m_S satisfying this constraint, limits on λ and g are calculated using Eq. (3.15) and Eq. (3.16).

Note that in this model the S particles are produced at $T = T_{\rm therm}$ and remain in thermal equilibrium with the radiation until they freeze-out at $T \simeq m_S$. This is different from the assumption made in Ref. [19], where S starts in thermal equilibrium and decouples while still relativistic. In order to achieve the latter scenario, the coupling of S to radiation has to proceed via a higher dimensional operator; that is via the propagation of an intermediate particle with mass much greater than m_S.

This brings us to our second model. Consider a heavy fermion ψ_S and a light fermion ψ_X, coupled via an additional heavy scalar ϕ_H with mass m_H,

$$\mathcal{L}_{\rm int} = g_S\, \bar{\psi}_S \psi_S\, \phi_H + g_X\, \bar{\psi}_X \psi_X\, \phi_H\,. \tag{3.19}$$

We assume that the fermion ψ_S decays to radiation with rate Γ. The annihilation cross section is

$$\langle \sigma v \rangle \sim \frac{g_S^2 g_X^2}{m_H^4} M^2\,, \tag{3.20}$$

where M is defined in Eq. (3.14). In this case, thermalization occurs for temperatures bounded by

$$m_S \left(\frac{m_H^4}{m_S^3 m_{\rm pl}} \frac{1}{g_S^2 g_X^2} \right)^{1/3} < T < g_S^2 g_X^2 m_{\rm pl}\,. \tag{3.21}$$

The conditions that S is in in thermal equilibrium when it reaches $T \simeq m_S$ gives the condition

$$g_S^2 g_X^2 > \frac{m_H^4}{m_S^3 m_{\rm pl}}\,. \tag{3.22}$$

Note that one still needs to have a decay rate that is small enough such that ψ_S decays after it dominates the universe. The point of this second example is to show that in non-renormalizable models the heavy species can either decouple while non-relativistic or while relativistic, depending on whether Eq. (3.22) is satisfied or not.

3.4 Models for Producing the Fluctuations

The density perturbations in the DGZK mechanism and our generalization originate in fluctuations in a light scalar field χ. In this section we write down explicit models for couplings between the fields S and χ. The reason for doing this is that these interactions can give rise to back reactions which can constrain the magnitude of the produced density perturbations. Similar results hold for couplings between ψ_S and χ.

The fluctuations in χ are created during the inflationary era when $\delta\chi \sim H_{\rm inf}$. We find it convenient to split χ into homogeneous and inhomogeneous parts, $\chi = \langle\chi\rangle + \delta\chi$. Note that this does not correspond to a perturbative expansion. Then the leading order equation of motion for χ can be split into homogeneous and inhomogeneous parts,

$$\begin{aligned} \langle\ddot{\chi}\rangle &= -3H\langle\dot{\chi}\rangle - \langle V'\rangle\,, \\ \delta\ddot{\chi} &= -3H\delta\dot{\chi} + 4\dot{\phi}\langle\dot{\chi}\rangle - \delta V' - 2\phi\langle V'\rangle\,. \end{aligned} \tag{3.23}$$

Here $V' = \langle V'\rangle + \delta V'$ is the derivative of the χ potential with respect to χ, separated into homogeneous and inhomogeneous parts. Also, ϕ is the time perturbation in conformal Newtonian gauge. The terms proportional to ϕ enter into the leading order equation of motion for $\delta\chi$ because their homogeneous coefficients do not.

To simplify the analysis, we first consider the scenario where $\langle\chi\rangle$ is negligible. From Eqs. (3.23) we see this is the case when $\langle\chi\rangle < \delta\chi$. Thus we require the equation of motion for $\langle\chi\rangle$ to be Hubble friction dominated for $\langle\chi\rangle < \delta\chi$. This gives the condition

$$H^2\delta\chi \;>\; H^2\langle\chi\rangle \;>\; \langle V'\rangle\,. \tag{3.24}$$

The fluctuations $\delta\chi$ persist so long as the equation of motion for $\delta\chi$ is Hubble friction dominated. With $\langle\chi\rangle < \delta\chi$ this translates into the condition

$$H^2\delta\chi \;>\; \delta V' + 2\phi\langle V'\rangle\,. \tag{3.25}$$

Note that we can combine our simplifying condition that the vev $\langle\chi\rangle$ be negligible, Eq. (3.24), with the condition that the fluctuations in $\delta\chi$ be Hubble friction dominated, Eq. (3.25). This gives the inequality

$$H^2\delta\chi \;>\; V'\,. \tag{3.26}$$

We consider the constraints this condition imposes on models for transferring χ fluctuations to the radiation. We first consider the renormalizable interactions

$$\mathcal{L}_\chi = \sqrt{-g}\left[-\frac{\alpha_S}{4}S^2\chi^2 - \frac{\mu_S}{2}S^2\chi\right], \qquad (3.27)$$

and neglect any couplings between χ and X as they are irrelevant to our mechanism. When χ fluctuates these interactions result in S mass fluctuations of

$$\delta_m = \frac{\alpha_S\delta\chi^2}{4m_S^2} + \frac{\mu_S\delta\chi}{2m_S^2} \sim \sqrt{\left(\frac{\alpha_S H_{\rm inf}^2}{m_S^2}\right)^2 + \left(\frac{\mu_S H_{\rm inf}}{m_S^2}\right)^2}, \qquad (3.28)$$

where in the last expression we have estimated the size of the rms fluctuation at two widely separated co-moving points. This mass fluctuation gives rise to fluctuations in the decay rate and the annihilation cross section of S according to the mass dependence of Eqs. (3.13).

As described above, for this fluctuation to persist and for $\langle\chi\rangle$ to remain negligible requires that $H^2\delta\chi > V'$. Although we assume the self-interaction of χ is always negligible, the interactions of $\mathcal{L}_\chi$ contribute to V and provide the constraint

$$H^2\delta\chi > \left(\frac{\alpha_S}{2}\delta\chi + \frac{\mu_S}{2}\right)\langle S^2\rangle, \qquad (3.29)$$

where $\langle S^2\rangle$ is evaluated in the thermal bath. This constraint is tightest at $T = m_S$ when $\langle S^2\rangle/H^2 \sim m_{\rm pl}^2/m_S^2$. Thus we obtain the constraints

$$\alpha_S < \frac{m_S^2}{m_{\rm pl}^2}, \qquad \mu_S < \frac{m_S^2}{m_{\rm pl}^2}H_{\rm inf}. \qquad (3.30)$$

The constraints of Eqs. (3.30) provide the same upper bound to both of the terms in Eq. (3.28). Thus the back reactions of the interactions in $\mathcal{L}_\chi$ limit the level of density perturbations produced via this mechanism to

$$\zeta \sim \delta_m < \frac{H_{\rm inf}^2}{m_{\rm pl}^2} \lesssim 10^{-8}, \qquad (3.31)$$

where the last limit on $H_{\rm inf}/m_{\rm pl}$ is measured by the WMAP collaboration [53].

The fluctuations resulting from the second interaction in $\mathcal{L}_\chi$ are linear in $\delta\chi$ and are therefore predominantly Gaussian in their distribution. Since the observed level of Gaussian fluctuations restricts ζ to be $\zeta \sim 10^{-5}$, this interaction cannot provide a significant fraction of the observed density perturbations. However, the fluctuations resulting from the first term in $\mathcal{L}_\chi$ are quadratic in $\delta\chi$ and are therefore non-Gaussian [59]. Recent analysis [63] limits the amplitude of non-Gaussian perturbations to about 10^{-8}. Thus we see that our model can provide non-Gaussian perturbations right up to the limit of current observation. A lower level of perturbations is obtained by simply reducing α_S or μ_S.

As a variant on the above scenario, we next consider the non-renormalizable couplings

$$\mathcal{L}'_\chi \;=\; \sqrt{-g}\left[-\frac{\lambda}{4}\frac{\chi^2}{M_1^2}S^2X^2 - \frac{\lambda}{4}\frac{\chi}{M_2}S^2X^2\right]. \tag{3.32}$$

When χ fluctuates these interactions result in fluctuations in $\langle\sigma v\rangle$

$$\delta_{\langle\sigma v\rangle} \;=\; \frac{2\delta\chi^2}{M_1^2} + \frac{2\delta\chi}{M_2} \sim \sqrt{\left(\frac{H_{\rm inf}^2}{M_1^2}\right)^2 + \left(\frac{H_{\rm inf}}{M_2}\right)^2}. \tag{3.33}$$

As above, we require that Eq. (3.26) be satisfied. For the interactions of $\mathcal{L}'_\chi$ this gives

$$H^2\delta\chi \;>\; \frac{\lambda}{4}\left(\frac{2\delta\chi}{M_1^2} + \frac{1}{M_2}\right)\langle S^2X^2\rangle. \tag{3.34}$$

As in the previous example, this constraint is tightest at $T = m_S$ when $\langle S^2X^2\rangle/H^2 \sim m_{\rm pl}^2$. Therefore we find

$$\frac{1}{M_1^2} \;<\; \frac{1}{\lambda}\frac{1}{m_{\rm pl}^2}, \qquad \frac{1}{M_2} \;<\; \frac{1}{\lambda}\frac{H_{\rm inf}}{m_{\rm pl}^2}. \tag{3.35}$$

Analogous to the previous example, the constraints of Eqs. (3.35) provide the same upper bound to both terms in Eq. (3.33). Thus the back reactions of $\mathcal{L}'_\chi$ limit the level of density perturbations produced via this mechanism to

$$\zeta \;\sim\; \delta_{\langle\sigma v\rangle} < \frac{1}{\lambda}\frac{H_{\rm inf}^2}{m_{\rm pl}^2}. \tag{3.36}$$

This bound is significantly weaker than the bound of Eq. (3.31) obtained via a fluctuating S mass. For example, the fluctuations resulting from $\mathcal{L}'_\chi$ could form the dominant contribution to the observed density perturbations if λ is sufficiently small. In addition, for a given ζ decreasing λ allows for a lower scale of inflation. Constraints on the smallness of λ are discussed in Section 3.3. Of course, a lower level of Gaussian (non-Gaussian) perturbations is obtained by increasing M_2 (M_1).

Above we have taken $\langle\chi\rangle$ to be negligible, which corresponds to taking $\langle\chi\rangle < \delta\chi$. Although this simplifies the presentation, it also unnecessarily strengthens the constraints on μ_S and M_2. Referring to the second of Eqs. (3.23), we see that for arbitrary $\langle\chi\rangle$ the requirement that $\delta\chi$ remains Hubble friction dominated gives

$$H^2\delta\chi > \dot\phi\langle\dot\chi\rangle, \quad H^2\delta\chi > \delta V', \quad H^2\delta\chi > \phi\langle V'\rangle. \tag{3.37}$$

The first condition provides the constraint $\langle\chi\rangle < \delta\chi/\phi$, with the evolution of $\phi \sim \zeta$ described in

Section 3.5. It is sufficient to take $\phi \sim 10^{-5}$, which also ensures that the homogeneous correction that $\langle\chi\rangle$ provides to m_S does not change m_S by more than order unity.[1] Through an analysis analogous to that above, we find the conditions of Eqs. (3.37) constrain the level of *Gaussian* fluctuations for the respective interactions of $\mathcal{L}_\chi$ and $\mathcal{L}'_\chi$ to

$$\zeta_{\rm g} < \frac{\langle\chi\rangle}{\delta\chi}\frac{H_{\rm inf}^2}{m_{\rm pl}^2}\,, \quad \zeta_{\rm g} < \frac{1}{\lambda}\frac{\langle\chi\rangle}{\delta\chi}\frac{H_{\rm inf}^2}{m_{\rm pl}^2}\,; \quad \frac{\langle\chi\rangle}{\delta\chi} < \frac{1}{\phi}\,. \tag{3.38}$$

The additional factor of $\langle\chi\rangle/\delta\chi$ significantly weakens both bounds on Gaussian perturbations. This allows for greater freedom in choosing μ_S, M_2, λ, and/or $H_{\rm inf}$.

It is important to note that these considerations must be applied with care when allowing for non-Gaussian perturbations ($\alpha_S \neq 0$ or M_1 finite). For example, in this case the mass fluctuation resulting from $\mathcal{L}_\chi$ becomes

$$\delta_m = \frac{\alpha_S\delta\chi^2}{4m_S^2} + \frac{\alpha_S\langle\chi\rangle\delta\chi}{2m_S^2} + \frac{\mu_S\delta\chi}{2m_S^2}\,. \tag{3.39}$$

We see that non-Gaussian perturbations of amplitude $\zeta_{\rm ng}$ are always accompanied by Gaussian perturbations of amplitude $\zeta_{\rm g} \sim (\langle\chi\rangle/\delta\chi)\zeta_{\rm ng}$. Therefore the level of non-Gaussian perturbations is limited to $\zeta_{\rm ng} \sim (\delta\chi/\langle\chi\rangle)\zeta_{\rm g}$.

3.5 Evolution of Density Perturbations

In this section we determine the evolution of density perturbations generated by a fluctuating cross section or mass during freeze-out. Unlike other analytic derivations given elsewhere in this work, the one provided here allows us to easily follow the growth of the non-adiabatic perturbation during the radiation dominated era after freeze-out. We work in conformal Newtonian gauge with negligible anisotropic stress and use the line element

$$ds^2 = a^2\left[-(1+2\phi)d\eta^2 + (1-2\psi)\delta_{ij}dx^i dx^j\right]\,, \tag{3.40}$$

with $\psi = \phi$. We track the evolution of perturbations using the gauge invariant entropy perturbation $\mathcal{S}$ and curvature perturbation ζ.

We will assume that the scattering and annihilation interactions between S and the radiation conserve total particle number. This allows us to obtain a first integral of the Boltzmann equations. Conservation of total particle number in a fixed co-moving volume implies for the entropy

---

[1] In Section 3.5 we find that after freeze-out ϕ evolves as $\phi \sim (\rho_S/\rho)\,\zeta_{\rm f}$, where $\zeta_{\rm f} \sim 10^{-5}$ is the final curvature perturbation. Thus if we consider the scenario where χ fluctuations are transferred at freeze-out and χ subsequently decays, we may take $\langle\chi\rangle$ to be constrained by ϕ^{-1} at freeze-out, which considerable weakens the bounds in Eqs. (3.38). However, in this case $\langle\chi\rangle$ provides a homogeneous adjustment to m_S which may be much larger than m_S. This effect could then significantly alter the constraints calculated in Section 3.3.

perturbation

$$\mathcal{S} \equiv \delta_S - \delta_R = -\left(\delta_R - 3\phi\right)\left(1 + \frac{n_R}{n_S}\right) + \frac{\lambda_0}{n_S a^3}\,, \tag{3.41}$$

where $\delta_S \equiv \delta n_S/n_S$ and $\delta_R \equiv \delta n_R/n_R$ are the perturbations in the number densities of S and radiation, respectively. Here λ_0 is an integration constant which vanishes in the absence of initial adiabatic perturbations.

Eq. (3.41) has a few salient features that we now discuss. First note that it admits an adiabatic solution $\mathcal{S} = 0$ whenever both n_R/n_S and $n_S a^3$ are constant. This solution is the familiar $\delta_S = \delta_R = 3\phi + \text{const}$, with the constant fixed by λ_0. During freeze-out, however, the conditions described above are not satisfied, and entropy perturbations $\mathcal{S}^{\text{f.o.}}$ are generated. Using Eq. (3.1) we find

$$\mathcal{S}^{\text{f.o.}} \simeq -\delta_{\langle\sigma v\rangle} - \delta_m\,. \tag{3.42}$$

After freeze-out the heavy particle no longer interacts with the radiation and therefore both δ_S and δ_R obey the perturbed Einstein field equation $\dot{\delta}_{S,R} = 3\dot{\phi}$. This is trivially integrated to give

$$\delta_{S,R} = 3\phi + \delta_{S,R}^{\text{f.o.}} - 3\phi^{\text{f.o.}}\,. \tag{3.43}$$

Thus, after freeze-out the entropy perturbation remains constant $\mathcal{S} = \mathcal{S}^{\text{f.o.}}$.

To derive the evolution of the curvature perturbation ζ, we use Eq. (3.7) together with $\delta\rho_S = \rho_S(\delta_S + \delta_m)$ and $\delta\rho_R = (4/3)\rho_R\delta_R$ to find, after several lines of algebra,

$$\zeta \simeq \frac{\rho_S(\mathcal{S}^{\text{f.o.}} + \delta_m)}{3\rho_S + 4\rho_R} + \frac{1}{3}\left(\delta_R^{\text{f.o}} - 3\phi^{\text{f.o.}}\right)\,. \tag{3.44}$$

Using Eq. (3.41), the second term on the right is suppressed by $(n_S/n_R)(4\rho_R/\rho_S{+}3)$ compared to the first term and is subsequently neglected. We see that in the radiation dominated era the curvature perturbations produced during freeze-out are suppressed relative to the entropy perturbations.

Once the the heavy particle starts to dominate the energy density of the universe, we find

$$\zeta \simeq -\frac{1}{3}\delta_{\langle\sigma v\rangle}\,, \tag{3.45}$$

in agreement with the results obtained in previous sections of this chapter. If S is the dark matter, such a large ratio of entropy to curvature perturbations is disfavored by data, which requires $|\mathcal{S}/\zeta| \lesssim 1/3$ [53]. If instead S decays, the entropy perturbations in S are transferred to radiation. Including the result of inhomogeneous S decay [18, 19, 54, 55, 56], the resulting curvature perturbation is

$$\zeta \simeq -\frac{1}{3}\delta_{\langle\sigma v\rangle} - \frac{1}{6}\delta_\Gamma\,. \tag{3.46}$$

3.6 Numerical Results

We now solve for the evolution of density perturbations during freeze-out using Boltzmann equations (see for example Refs. [9, 52, 64]). As above, we work in conformal Newtonian gauge and consider only super-horizon perturbations, neglecting all spatial gradients next to conformal time derivatives. Since we assume that S is non-relativistic, the distribution function for S is $f_S = e^{\mu/T} f_{\rm eq}$, where $f_{\rm eq}$ is the Maxwell-Boltzmann equilibrium distribution function. Then integrating the background Boltzmann equations over all of phase space gives

$$\begin{aligned} \dot{n}_S + 3\mathcal{H} n_S &= a\langle\overline{\sigma v}\rangle \left(n_{\rm eq}^2 - n_S^2\right) - a\overline{\Gamma} n_S \,, \\ \dot{n}_R + 3\mathcal{H} n_R &= -\,a\langle\overline{\sigma v}\rangle \left(n_{\rm eq}^2 - n_S^2\right) + a\overline{\Gamma} n_S \,. \end{aligned} \tag{3.47}$$

where dots denote derivatives with respect to conformal time, $\mathcal{H} = \dot{a}/a$ and $n_{\rm eq}$ is the equilibrium number density of S particles. To describe the evolution of the scale factor we use Einstein's field equation,

$$\mathcal{H}^2 = \frac{a^2}{3m_{\rm pl}^2} \left(\rho_R + \rho_S\right) . \tag{3.48}$$

To derive the subleading order Boltzmann equations we allow for fluctuations in m_S, $\langle\sigma v\rangle$ and Γ as defined in Eqs. (3.5). Note that the resulting fluctuations in the number density of the radiation δ_R correspond to temperature fluctuations $\delta_T = \delta_R/3$. Therefore the equilibrium number density of S particles acquires fluctuations

$$\delta_{\rm eq} = \frac{3}{2}\left(\delta_m + \delta_T\right) + \exp\left[-\frac{m_S}{T}\left(\delta_m - \delta_T\right)\right] - 1\,. \tag{3.49}$$

When the subleading Boltzmann equations are integrated over all phase space we obtain

$$\begin{aligned} \dot{\delta}_S - 3\dot{\phi} &= a\langle\overline{\sigma v}\rangle \frac{n_{\rm eq}^2}{n_S} \left(\phi + \delta_{\langle\sigma v\rangle} + 2\delta_{\rm eq} - \delta_S\right) - a\langle\overline{\sigma v}\rangle n_S \left(\phi + \delta_{\langle\sigma v\rangle} + \delta_S\right) - a\overline{\Gamma}\left(\phi + \delta_\Gamma\right) , \\ \dot{\delta}_R - 3\dot{\phi} &= -\,a\langle\overline{\sigma v}\rangle \frac{n_{\rm eq}^2}{n_R} \left(\phi + \delta_{\langle\sigma v\rangle} + 2\delta_{\rm eq} - \delta_R\right) + a\langle\overline{\sigma v}\rangle \frac{n_S^2}{n_R} \left(\phi + \delta_{\langle\sigma v\rangle} + 2\delta_S - \delta_R\right) \\ &\quad + a\overline{\Gamma}\frac{n_S}{n_R}\left(\phi + \delta_\Gamma + \delta_S - \delta_R\right) . \end{aligned} \tag{3.50}$$

Note that in deriving the Boltzmann equations we assumed $\delta_{\rm eq} \ll 1$. This ceases to be valid once $T \lesssim m_S(\delta_m - \delta_T)$. In this case, however, the factor of $n_{\rm eq}^2$ in front of the terms containing $\delta_{\rm eq}$ is exponentially suppressed compared to the remaining terms.

The independent perturbations in Eqs. (3.50) are δ_S, δ_R, and ϕ. To describe the evolution of ϕ

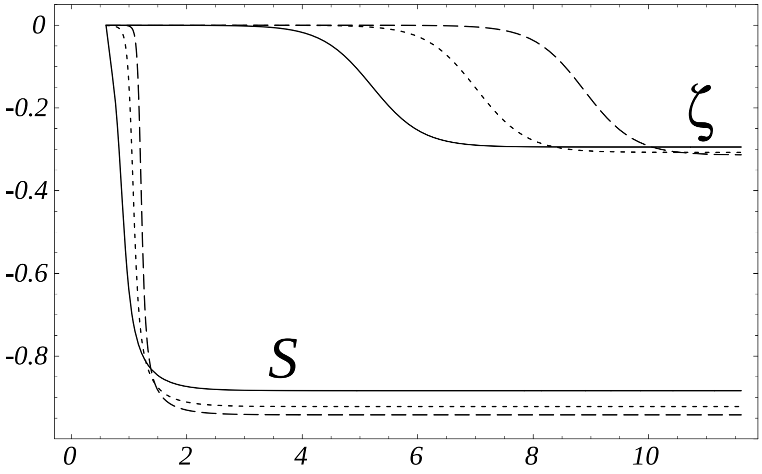


Figure 3.1: Evolution of $\mathcal{S}$ and ζ in units of $\delta_{\langle\sigma v\rangle}$ as a function of $\log(m_S/T)$. The solid, dotted, and dashed curves correspond to $\langle\sigma v\rangle m_S m_{\rm pl} = 10^5$, 10^7, and 10^9, respectively. The large values of $\langle\sigma v\rangle m_S m_{\rm pl}$ are necessary to ensure that the heavy particle is in thermal equilibrium at $T_0 = m_S/4$.

we use the first order perturbation to Einstein's field equation,

$$\mathcal{H}\dot{\phi} + \mathcal{H}^2\phi \;=\; -\frac{a^2}{6m_{\rm pl}^2}\left(\delta\rho_R + \delta\rho_S\right) . \tag{3.51}$$

We solve the above system of equations numerically. The effects of a fluctuating decay rate Γ are well-studied [18, 19, 54, 55, 56] so we set $\Gamma = 0$ to simplify our results. For concreteness we also assume S interacts with one out of one hundred radiative degrees of freedom and we begin integration at $T_0 = m_S/4$ with $n_S(T_0) = n_{\rm eq}(T_0)$. The results for the gauge invariant quantities $\mathcal{S}$ and ζ are shown in Fig. 3.1 for several values of $\langle\sigma v\rangle\, m_S\, m_{\rm pl}$.

The curvature perturbation ζ is negligible compared to $\delta_{\langle\sigma v\rangle}$ until the heavy particle contributes significantly to the energy density. It asymptotes to a value $\zeta \simeq -0.3$. The entropy perturbation $\mathcal{S}$ grows during freeze-out and soon thereafter reaches a constant value of $\mathcal{S} \simeq -0.9$. These results and the other features of Fig. 3.1 are in good agreement with the analytical results given in Eq. (3.42) and Eq. (3.45).

3.7 Conclusions

In Refs. [18, 19, 20] it was shown that fluctuations in the mass and the decay rate of a heavy particle S, which at some point dominates the energy density of the universe, lead to adiabatic density perturbations. In this scenario it was assumed that the heavy particle decouples from radiation while it is still relativistic.

In this work we have shown that if the heavy particle remains in thermal equilibrium until it becomes non-relativistic, fluctuations in the annihilation cross section of this particle with radiation lead to additional sources of perturbations. We have presented two simple toy models illustrating this effect. These additional fluctuations are generic, unless the annihilation cross section is mediated by an additional particle with mass exceeding m_S. If the S particle is stable, for example if S is dark matter, then the resulting perturbations are non-adiabatic.

A simple analytical calculation determines the size of the density perturbations from fluctuations in the mass, decay rate and annihilation cross section. The fluctuations due to variations in the annihilation cross section are shown to be of similar size as the ones generated from the original DGZK mechanism. These results are checked numerically using Boltzmann equations in conformal Newtonian gauge in Section 3.6.

Acknowledgments

We would like to thank Mark Wise for collaboration at an early stage of this work. This work was supported by the Department of Energy under the contract DE-FG03-92ER40701.

Chapter 4

On the Generation of Density Perturbations at the End of Inflation

Recently a mechanism was proposed whereby the primordial density perturbations are generated at the end of inflation. We continue the analysis of the proposed model of this mechanism and calculate the maximum extent to which the density perturbations produced via this model can dominate over those of the standard inflationary paradigm. In addition, we provide a straightforward variation of this model which allows for greater amplification of the density perturbations. Finally, we show that a variation in the implementation of the original model results in significant non-Gaussianities in the resulting spectrum of density perturbations. The level of non-Gaussianities can be made to saturate the current observational bound.

Based on M. P. Salem, *Phys. Rev. D* **72**, 123516 (2005).

4.1 Introduction

Measurements of the cosmic microwave background radiation [49, 50, 51] have revealed a highly uniform background energy density with nearly scale-free perturbations with relative amplitudes typically on the order of a few parts in 10^5. A possible source of these perturbations is found in the quantum fluctuations of one or more light scalar fields during an early epoch of inflation [65, 66] (for reviews see Refs. [9, 10, 11]). This is because in the (quasi-) de Sitter space of inflationary expansion, a quantum fluctuation in a scalar field evolves according to its classical equations of motion after its wavelength exceeds the rapidly decreasing Hubble length [67, 68, 12, 17]. These fluctuations in the scalar field are frozen so long as the effective mass of the scalar is much smaller than the Hubble rate; and they may be converted into energy density perturbations via a variety of proposed mechanisms.

For example in the standard inflationary paradigm [10, 11], inflation is driven by the potential energy of a single slowly-rolling scalar field. In this case the inflaton can be viewed as a unique clock parameterizing the evolution of the early universe. Therefore fluctuations in the inflaton translate into fluctuations in the duration of inflation. Since the energy density of the universe redshifts more rapidly after inflation than during inflation, this results in energy density fluctuations on surfaces of constant scale factor after inflation.

More recently, it was proposed that energy density perturbations could result from fluctuations in light scalar fields that do not contribute significantly toward the inflationary dynamics. For example, in the "curvaton" scenario [22, 23, 24] (see also Refs. [69, 70]) a light scalar field dubbed the curvaton receives fluctuations. After inflation, the curvaton evolves as a massive fluid and therefore redshifts more slowly than the radiative products of reheating. If the curvaton eventually dominates the energy density of the universe and then decays, density perturbations result because the duration of curvaton domination depends on the fluctuating curvaton. The inhomogeneous reheating scenario [18, 20] also achieves density perturbations by varying the duration that a massive fluid dominates the energy density of the universe. In this case the duration of domination is modulated via a decay width that depends on some fluctuating light scalar field. Other ways to modulate the duration of of massive fluid's domination have been proposed in Refs. [19, 59, 26, 62, 71]. A nice review of these ideas can be found in Ref. [72].

Another variant upon these proposals was recently proposed in Ref. [25] (note however that a similar idea was previously applied in Ref. [73]). Whereas in the standard picture the duration of inflation is influenced by the fluctuations in the inflaton as relevant modes leave the Hubble radius, in this picture the duration of inflation is influenced by fluctuations in the inflaton when inflation ends. This can occur, for example, if inflation ends at an inflaton value that depends on some other field that receives fluctuations during inflation. In Ref. [25] a specific model was used to demonstrate that this scenario can lead to significant amplification of the density perturbations that result from slow-roll inflation.

Here we continue the analysis of that model and calculate the maximum extent to which the resulting density perturbations can dominate over those of the standard inflationary paradigm. We also explore the sensitivity of this result to the tuning of model parameters. In addition, we provide a straightforward variation of this model which allows for greater amplification of the density perturbations. Finally, we show that a slight variation in the implementation of the original model allows for significant non-Gaussianities in the spectrum of density perturbations. The level of non-Gaussianities can be made to saturate the current observational bound.

The remainder of this chapter is organized as follows. In Section 4.2 we describe the novel mechanism of Ref. [25] and describe a convenient notation and formalism for this study. The specific model used to illustrate this mechanism in Ref. [25] is then described in Section 4.3, while in Section 4.4 we analyze this model in greater detail. In Section 4.5 we introduce two variations upon this model and study their consequences. One variation greatly amplifies the effect of this mechanism while the other may generate significant non-Gaussianities in the primordial spectrum of density perturbations. Finally, discussion is provided and conclusions are drawn in Section 4.6.

4.2 Background

We parameterize metric perturbations using the curvature perturbation on surfaces of constant scale factor, otherwise known as the Bardeen variable ζ [16, 17]. In this chapter we ignore tensor perturbations, which allows us to write the spatial metric

$$g_{ij} = a^2(t)e^{2\zeta(t,\mathbf{x})}\delta_{ij} \, , \tag{4.1}$$

since scalar field fluctuations carry no anisotropic stress (for a review of cosmological perturbation theory see for example Refs. [74, 75]).

The proposal of Ref. [25] is most easily studied within the so-called δN formalism of Refs. [13, 12, 14, 15] (see Refs. [76, 77] for recent extensions of this formalism). This formalism notes that on super-Hubble scales the number of e-folds of expansion between an initial (time t_0) flat hypersurface and a final (time t) hypersurface of constant density is given by

$$N(t,\mathbf{x}) = \ln\left[\frac{a(t)e^{\zeta(t,\mathbf{x})}}{a(t_0)}\right] \, , \tag{4.2}$$

Rearranging the terms in Eq. (4.2) gives

$$\zeta(t,\mathbf{x}) = N(t,\mathbf{x}) - \ln\left[\frac{a(t)}{a(t_0)}\right] \equiv \delta N(t,\mathbf{x}) \, . \tag{4.3}$$

We limit our attention to the case where inflation is driven by a single inflaton. Then fluctuations in N come only from the value of the inflaton when its wavelength becomes larger than the Hubble length, denoted ϕ_k, and the value of the inflaton at the end of inflation, denoted ϕ_{e}. Thus the number of e-folds of universal expansion is $N(\phi_k, \phi_{\mathrm{e}})$, and

$$\delta N \simeq \frac{\partial N}{\partial \phi_k}\delta\phi_k + \frac{\partial N}{\partial \phi_{\mathrm{e}}}\delta\phi_{\mathrm{e}} \, . \tag{4.4}$$

Both $\delta\phi_k$ and $\delta\phi_{\mathrm{e}}$ will ultimately stem from fluctutions in scalar fields as modes exit the Hubble radius; in the δN formalism both of these are evaluated on the initial flat hypersurface defined at t_0. That the power spectrum for ζ is observed to be predominantly Gaussian allows us to neglect higher-order terms that might appear in the expansion of Eq. (4.4).

In the standard picture, inflation ends when the slow-roll conditions are violated. For single field inflation this happens at a unique value ϕ_{e} and therefore the second term in Eq. (4.4) is zero. For inflation to end at varying values of ϕ_{e} requires the addition of at least one other field. For example, with an additional field σ it is possible that $\phi_{\mathrm{e}}(\sigma)$ depends on σ and inflation ends within a range of $\phi_{\rm e}$ given by

$$\delta\phi_{\rm e} \simeq \frac{\partial\phi_{\rm e}}{\partial\sigma}\delta\sigma\,. \tag{4.5}$$

If the masses of the scalar fields are much less than the Hubble rate the fields acquire a constant power spectrum of fluctuations $\mathcal{P}_{\delta\phi_k} = \mathcal{P}_{\delta\sigma_k} = (H_k/2\pi)^2$, where H_k is the Hubble rate at the time the mode k exits the Hubble radius (for a pedigogical review of such calculations see for example Ref. [78]). Thus the power spectrum for ζ is

$$\mathcal{P}_\zeta \simeq \frac{H_k^2}{4\pi^2}\left[\left(\frac{\partial N}{\partial\phi_k}\right)^2 + \left(\frac{\partial N}{\partial\phi_{\rm e}}\frac{\partial\phi_{\rm e}}{\partial\sigma_k}\right)^2\right]\,. \tag{4.6}$$

The density perturbations generated at the end of inflation will dominate over those produced from the standard picture when

$$\frac{\partial N}{\partial\phi_{\rm e}}\frac{\partial\phi_{\rm e}}{\partial\sigma} > \frac{\partial N}{\partial\phi_k}\,. \tag{4.7}$$

Note that $\partial N/\partial\phi_k \neq \partial N/\partial\phi_{\rm e}$ since the former probes the dependence of N on ϕ deep in the inflationary epoch while the latter probes the dependence of N on ϕ near the end of inflation.

4.3 The Specific Model

The above ideas were demonstrated in Ref. [25] using a specific model described by the potential

$$V(\phi,\chi,\sigma) \;=\; \frac{1}{4g}\left(m_\chi^2 - g\chi^2\right)^2 + \frac{1}{2}m_\phi^2\phi^2 + \frac{1}{2}\lambda_\phi\phi^2\chi^2 + \frac{1}{2}\lambda_\sigma\sigma^2\chi^2 + V_\sigma(\sigma)\,. \tag{4.8}$$

As was noted in Ref. [25], this is the original hybrid inflation model [79, 80] but with interactions involving an additional light scalar field σ added with the last two terms.

Hybrid inflation assumes the the same initial conditions that are assumed in chaotic inflation [81, 79, 80]. However, during the early stages of hybrid inflation the field χ is pushed to zero much faster than the field ϕ (and likewise faster than σ in the above model). Then ϕ becomes a slowly-rolling inflaton with inflation assisted by the vacuum energy $m_\chi^4/4g$. As in Ref. [25], we assume that the self interaction of σ represented by the term V_σ does not contribute significantly toward the inflationary dynamics.

The field χ is pinned to the origin until the curvature in its potential, $\partial^2V/\partial\chi^2$, becomes negative. In Ref. [25] a scenario is described in which inflation ends abruptly as χ rolls away from the origin.

This happens when

$$\frac{\partial^2 V}{\partial \chi^2} = \lambda_\phi \phi_{\rm e}^2 + \lambda_\sigma \sigma_{\rm e}^2 - m_\chi^2 = 0\,. \tag{4.9}$$

Therefore if the field σ receives fluctuations $\delta\sigma$ as modes exit the Hubble radius, then inflation ends at field values $\phi_{\rm e}$ that vary according to

$$\delta\phi_{\rm e} \simeq -\frac{\lambda_\sigma \sigma_{\rm e}}{\lambda_\phi \phi_{\rm e}} \delta\sigma\,. \tag{4.10}$$

Referring back to Eq. (4.6), we see the density perturbations resulting from fluctuations $\delta\phi_{\rm e}$ dominate over those resulting from $\delta\phi_k$ when

$$R \equiv \frac{\lambda_\sigma^2 \sigma_{\rm e}^2}{\lambda_\phi^2 \phi_{\rm e}^2} \left(\frac{\partial N / \partial \phi_{\rm e}}{\partial N / \partial \phi_k} \right)^2 = \frac{\lambda_\sigma^2 \sigma_{\rm e}^2}{\lambda_\phi^2 \phi_{\rm e}^2} \frac{\epsilon_k}{\epsilon_{\rm e}} > 1\,. \tag{4.11}$$

In the second equation we have used that the first slow-roll parameter can be written

$$\epsilon \equiv \frac{m_{\rm pl}^2}{2} \left(\frac{1}{V} \frac{\partial V}{\partial \phi} \right)^2 = \frac{1}{2 m_{\rm pl}^2} \left(\frac{\partial N}{\partial \phi} \right)^{-2}\,. \tag{4.12}$$

An important consequence of this mechanism is that the spectral tilt is given by the tilt in the spectrum of fluctuations $\delta\sigma$, as opposed to $\delta\phi$. This gives a tilt which is independent of the second slow-roll parameter η [25, 78]. Therefore η is not directly constrained by observation. In addition, since $\dot\epsilon \simeq 2H\epsilon\,(2\epsilon - \eta)$ it appears as if ϵ may decrease significantly during the course of inflation if $\eta \gg \epsilon$. Thus it was suggested in Ref. [25] that the condition of Eq. (4.11) is easily satisfied.

However, one might note that R is proportional to ϵ_k which according to this mechanism is constrained by observational bounds on the spectral tilt. These give $\epsilon_k \lesssim 0.02$ [51] in the absence of cancellations [25]. In addition, we do not expect $\epsilon_{\rm e}$ and $\phi_{\rm e}$ to be independent of each other. Thus it is worthwhile to investigate Eq. (4.11) in greater detail in order to determine precisely what limits the extent to which the density perturbations of this model can dominate over those produced via the standard inflationary paradigm.

4.4 A More Detailed Analysis

We now consider the model of Eq. (4.8) in greater detail in order to clarify the constraints on R as given by Eq. (4.11). Consistent with the analysis of Ref. [25] we ignore the contribution of V_σ toward the vacuum energy. As described above, the χ field rolls away from the origin when ϕ takes the value

$$\phi_{\rm e} = \sqrt{\frac{m_\chi^2 - \lambda_\sigma \sigma_{\rm e}^2}{\lambda_\phi}} \approx \frac{m_\chi}{\sqrt{\lambda_\phi}} \,. \tag{4.13}$$

To explain the second approximation, first note that we require $\lambda_\sigma \sigma_{\rm e}^2 < m_\chi^2$ in order to avoid χ remaining trapped at the origin after ϕ reaches zero. Meanwhile, the precise value of $\sigma_{\rm e}$ is a stochastic variable constrained by conditions independent of λ_σ and m_χ. Thus to ensure that the desired dynamics are typical of this model requires we set $m_\chi^2/\lambda_\sigma \gg \sigma_{\rm e}^2$ for typical values of $\sigma_{\rm e}$.

For inflation to last until $\phi = \phi_{\rm e}$ but then end abruptly when χ rolls away from the origin requires one of two possibilities. One possibility is to require that $\phi_{\rm e} = m_\chi/\sqrt{\lambda_\phi}$ be sufficiently large that inflation ends while the potential energy of ϕ dominates the vacuum energy. Since for the potential energy of ϕ to drive inflation requires that ϕ be of the order of the Planck mass [81], this requires a Planck scale m_χ unless $\lambda_\phi \gg 1$. Nevertheless, it can be shown that this scenario gives

$$R \simeq \frac{\phi_k^2}{\sigma_{\rm e}^2} \left(\frac{\lambda_\sigma \sigma_{\rm e}^2}{m_\chi^2} \right)^2 , \tag{4.14}$$

where ϕ_k is ϕ evaluated 60 e-folds before the end of inflation. This value of R is never significantly greater than that calculated for the second possibility, which we now study in greater detail.

The second possibility for inflation to last until $\phi = \phi_{\rm e}$ but then to end abruptly when χ rolls away from the origin requires that $m_\chi^4/4g$ dominate the energy density at the end of inflation and that m_χ^2 be much greater than the Hubble rate at the end of inflation. These conditions require:

$$\frac{2g}{\lambda_\phi} \frac{m_\phi^2}{m_\chi^2} \ll 1 \,, \quad \frac{1}{12g} \frac{m_\chi^2}{m_{\rm pl}^2} \ll 1 \,, \tag{4.15}$$

where $m_{\rm pl}$ denotes the reduced Planck mass. The first constraint of Eqs. (4.15) allows us to write

$$\frac{\partial N}{\partial \phi_{\rm e}} = \frac{V}{m_{\rm pl}^2} \left(\frac{\partial V}{\partial \phi_{\rm e}} \right)^{-1} \simeq \frac{m_\chi^4}{4g\phi_{\rm e} m_\phi^2 m_{\rm pl}^2} \approx \frac{\sqrt{\lambda_\phi} m_\chi^3}{4g m_\phi^2 m_{\rm pl}^2} \,. \tag{4.16}$$

Meanwhile

$$\frac{\partial N}{\partial \phi_k} = \frac{1}{\phi_k m_\phi^2 m_{\rm pl}^2} \left(\frac{m_\chi^4}{4g} + \frac{1}{2} m_\phi^2 \phi_k^2 \right) . \tag{4.17}$$

Plugging these values into Eq. (4.11) gives

$$R \approx \frac{\lambda_\sigma}{4g^2} \left(\frac{\lambda_\sigma \sigma_{\rm e}^2}{m_\chi^2} \right) m_\chi^6 \phi_k^2 \left(\frac{m_\chi^4}{2g} + m_\phi^2 \phi_k^2 \right)^{-2} . \tag{4.18}$$

Again, ϕ_k is the value of the inflaton field when the mode k exits the Hubble radius. For cosmological scales of interest this happens about 60 e-folds of inflation prior to the end of inflation.

In order to find the maximum value for R, we find the value of ϕ_k that maximizes R and set the parameters g, m_χ, and m_ϕ such that the mode k exits the Hubble radius 60 e-folds prior to inflation. This is equivalent to finding the balance between the vacuum energy $m_\chi^4/4g$ and the potential energy $m_\phi^2\phi^2$ that maximizes R. The result is

$$\phi_k = \frac{1}{\sqrt{2g}}\frac{m_\chi^2}{m_\phi}\,, \tag{4.19}$$

with the constraint that the mode k leaves the Hubble radius at $N_k \approx 60$, with

$$N_k = \frac{1}{4m_\phi^2 m_{\rm pl}^2}\left[\frac{m_\chi^4}{g}\ln\left(\frac{\sqrt{\lambda_\phi}\phi_k}{m_\chi}\right) + m_\phi^2\phi_k^2 - \frac{m_\phi^2 m_\chi^2}{\lambda_\phi}\right] \approx \frac{m_\chi^4}{8gm_\phi^2 m_{\rm pl}^2}\ln\left(\frac{\lambda_\phi}{2g}\frac{m_\chi^2}{m_\phi^2}\right)\,. \tag{4.20}$$

In the last expression we have used the constraints of Eqs. (4.15) to identify the most significant term. Putting this all together we find the maximum value of R to be

$$R \approx N_k\left(\frac{\lambda_\sigma\sigma_{\rm e}^2}{m_\chi^2}\right)^2\frac{m_{\rm pl}^2}{\sigma_{\rm e}^2}\left[\ln\left(\frac{\lambda_\phi}{2g}\frac{m_\chi^2}{m_\phi^2}\right)\right]^{-1}\,. \tag{4.21}$$

Before proceeding to study Eq. (4.21) we should check that producing the correct power spectrum normalization does not introduce any constraints that conflict with our present assumptions. Applying Eq. (4.6) to the scenario considered above we find that when density perturbations generated at the end of inflation dominate we have

$$\mathcal{P}_\zeta \approx \frac{2}{\pi^2}N_k R\left(\frac{1}{12g}\frac{m_\chi^2}{m_{\rm pl}^2}\right)\frac{m_\chi^2}{m_{\rm pl}^2}\left[\ln\left(\frac{\lambda_\phi}{2g}\frac{m_\chi^2}{m_\phi^2}\right)\right]^{-1}\,. \tag{4.22}$$

To match observation we require $\mathcal{P}_\zeta$ to be very small, $\mathcal{P}_\zeta \approx (5\times10^{-5})^2$ [49]. According to the second of Eqs. (4.15) the first term in parentheses is already constrained to be much less than order unity and in fact can be set as small as necessary to match observation. In addition we expect $m_\chi^2/m_{\rm pl}^2$ to be very small. Thus to set $\mathcal{P}_\zeta$ to match observation does not introduce any constraints in conflict with those of the above analysis.

We have written the ratio R in the form of Eq. (4.21) in order to emphasize the maximum extent to which the perturbations produced at the end of inflation may dominate over those produced as cosmological scales exit the Hubble radius. The first term in this expression is weakly constrained by the energy scale of inflation and is here taken to be $N_k \approx 60$. The first term in parentheses is constrained to be significantly less than unity as explained in the discussion below Eq. (4.13). Finally, the argument of the logarithm must be much greater than unity in order to satisfy the first constraint of Eqs. (4.15). Therefore the last term must be less than unity. Thus for appropriately tuned parameters we might expect the product of these three factors to be a couple orders of magnitude below unity.

However, Eq. (4.21) also contains a factor $m_{\rm pl}^2/\sigma_{\rm e}^2$. In (quasi-) de Sitter space a scalar field such as σ evolves both according to its classical equation of motion and due to quantum fluctuations as modes leave the Hubble radius. The net effect of this evolution is that the correlation function $\langle\sigma^2\rangle$ migrates toward a fixed value depending on V_σ and the Hubble rate H [82, 83, 84]. For example, if inflation lasts long enough and if $V_\sigma = \frac{1}{2}m_\sigma^2\sigma^2$ then $\langle\sigma^2\rangle \sim H^4/m_\sigma^2$ [82, 83, 84]. Taking a typical value of $\sigma_{\rm e}$ to be $\sigma_{\rm e} \sim \sqrt{\langle\sigma^2\rangle}$ gives

$$\frac{m_{\rm pl}^2}{\sigma_{\rm e}^2} \sim \frac{m_\sigma^2}{H^2}\frac{m_{\rm pl}^2}{H^2}\,. \tag{4.23}$$

The dynamics described above require $m_\sigma^2 \ll m_\phi^2 \ll H^2$, so the first term in Eq. (4.23) must be at least a few orders of magnitude below unity. However, the second term can be very large. Current observation gives $m_{\rm pl}^2/H^2 \gtrsim 10^8$ [53], which is more than sufficient to compensate for all the small factors in R if parameters are tuned appropriately. Reducing the scale of inflation allows for greater values of R. Of course, our Hubble volume could also be a region of atypically small $\sigma_{\rm e}$. Finally, we may reduce $\sigma_{\rm e}$ to an arbitrarily small scale by considering σ to be a pseudo-Nambu-Goldstone boson (see for example Refs. [22, 23, 24, 85]).[1] Thus in a number of circumstances we expect the level of density perturbations generated at the end of inflation to be significantly larger than those produced when cosmological scales exit the Hubble radius.

In Eq. (4.21) the parameters g, m_ϕ, and m_χ are tuned such that scales of cosmological interest leave the Hubble radius when ϕ_k is given by Eq. (4.19). To study the result of relaxing this tuning requires to invert Eq. (4.20) to obtain $\phi_k(N_k)$ and insert the result into Eq. (4.18) to obtain $R\,(N_k)$. This allows for the remaining parameters in R to be freely varied while R retains its original meaning; that is, that R compares density perturbations produced at the end of inflation to those produced when cosmological scales exit the Hubble radius.

Note that the first two terms in the brackets of Eq. (4.20) always dominate over the third term and that they are comparable to each other when ϕ_k is given by Eq. (4.19). Remember that ϕ_k is defined as the value of ϕ $N_k \approx 60$ e-folds before the end of inflation. Decreasing m_ϕ slows the evolution of ϕ which therefore decreases ϕ_k. In this case the first term in brackets becomes more

---

[1] An interesting scenario involves a pseudo-Nambu-Goldstone boson that ranges over a scale $\sim m_\chi$. Then λ_σ may be a coupling of order unity and we obtain

$$R \approx N_k\frac{m_{\rm pl}^2}{m_\chi^2}\,, \tag{4.24}$$

where we have dropped the logarithm and other factors of order unity. Clearly R is much greater than unity in this case.

important and inverting Eq. (4.20) gives

$$\phi_k^2 \approx \phi_{\mathrm{e}}^2 \exp\left(8N_k \frac{g m_\phi^2 m_{\mathrm{pl}}^2}{m_\chi^4}\right). \tag{4.25}$$

For ϕ_k less than in Eq. (4.19) the important functional dependence of R is $R(\phi_k) \propto \phi_k^2$. Thus we see that R decreases exponentially when the ratio $g m_\phi^2 m_{\mathrm{pl}}^2/m_\chi^2$ is decreased from its optimal value.

On the other hand, increasing m_ϕ quickens the evolution of ϕ and therefore increases ϕ_k. In this case the second term in the brackets of Eq. (4.20) becomes more important. Inverting N_k in this case gives $\phi_k^2 \approx N_k m_{\mathrm{pl}}^2$ which is relatively independent of the model parameters. Therefore the magnitude of R changes predominantly through its dependence on $m_\phi^2 \phi_k^2$ in the denominator of Eq. (4.18). Thus we see that significantly increasing the ratio $g m_\phi^2 m_{\mathrm{pl}}^2/m_\chi^4$ from its optimal value results in a roughly proportional decrease in the in size of R.

4.5 Generalizing the Model

4.5.1 Varying the Potential for ϕ

According to Eq. (4.7), the generation of density perturbations at the end of inflation is most effective when the slow-roll parameter near the end of inflation is much less than when cosmological scales of interest exit the Hubble radius. Thus to enhance the resulting perturbations we desire a potential for ϕ that decreases more steeply for ϕ deep in the inflationary epoch and decreases more gently for ϕ near the end of inflation. This can be accomplished by replacing the $\frac{1}{2}m_\phi^2\phi^2$ term in Eq. (4.8) with $\frac{1}{4}\lambda\phi^4$.

We analyze this scenario in exact analogy to the analysis in Section 4.4. Again we assume that the vacuum energy at the end of inflation is dominated by the $m_\chi^4/4g$ term. The calculations proceed just as in Section 4.4, and in the end we find

$$R \simeq 108 N_k^3 \left(\frac{\lambda_\sigma \sigma_{\mathrm{e}}^2}{m_\chi^2}\right)^2 \left(\frac{m_{\mathrm{pl}}^2}{\sigma_{\mathrm{e}}^2}\right) \left(\frac{\lambda_\phi m_{\mathrm{pl}}^2}{m_\chi^2}\right)^2. \tag{4.26}$$

As described in Section 4.4, the term in the first set of parentheses is expected to be significantly less than unity. However the last term is expected to be greater than unity in order that χ becomes pinned to the origin in the early stages of inflation. In addition, the numerical prefactor $108N_k^3 \sim 10^7$ for $N_k \approx 60$. The discussion about the factor $m_{\mathrm{pl}}^2/\sigma_{\mathrm{e}}^2$ near the end of Section 4.4 applies here. Thus we see the maximum level of density perturbations produced at the end of inflation can be greatly amplified by simply introducing a ϕ^4 potential.

4.5.2 Relaxing the Constraint on λ_σ

In Section 4.3 we followed the assumption presented in Ref. [25] that near the end of inflation $\lambda_\sigma \sigma^2 < m_\chi^2$ so that inflation ends before ϕ reaches zero. It is interesting to explore the consequences of lifting this assumption. If we do not demand that $\lambda_\sigma \sigma^2 < m_\chi^2$, then for large enough λ_σ or σ the χ field is still pinned to the origin when ϕ reaches zero. We then expect inflation to continue with the σ field rolling down its potential until the condition of Eq. (4.9) is met. At this point χ rolls away from the origin and abruptly initiates the end of inflation.

Nevertheless, if $m_\phi \ll H$ during inflation, then ϕ will retain a power spectrum of fluctuations given by $\mathcal{P}_{\delta\phi_k} = (H_k/2\pi)^2$. Thus the fields σ and ϕ have essentially changed places, with σ playing the part of the inflaton and ϕ the fluctuating field. However, in this scenario ϕ has no homogeneous component and

$$\delta\sigma_{\rm e} = -\frac{\lambda_\phi}{\lambda_\sigma}\frac{\delta\phi_k^2}{\sigma_{\rm e}}\,. \tag{4.27}$$

The density perturbations produced at the end of inflation are now entirely non-Gaussian!

It should be noted here that the original model described in Section 4.3 also results in non-Gaussianities. As described in Ref. [25], in that model the non-Gaussianities result from higher order terms that appear in the Taylor expansion of δN in Eq. (4.4). As such these non-Gaussianities are always suppressed relative the Gaussian component of the spectrum of fluctuations. The situation we consider here is different in that the non-Gaussianities arise at leading order in the expansion of δN. In fact, it is possible for these non-Gaussians to dominate the spectrum of fluctuations. We are interested in the possibility for Gaussian fluctuations to be generated by the standard mechanism (as modes exit the Hubble radius) while non-Gaussianities are generated at the end of inflation. This would allow for a spectrum like that which is measured to be generated but with non-Gaussianities at a level that could be measured in the foreseeable future.

It is standard to parameterize the level of non-Gaussianities using the quantity $f_{\rm NL}$ which for our purposes can be defined according to the equation

$$\zeta = \zeta_{\rm g} - \frac{3}{5} f_{\rm NL} \zeta_{\rm g}^2\,, \tag{4.28}$$

where $\zeta_{\rm g}$ symbolizes a variable with a Gaussian spectrum [86]. In this case $\zeta_{\rm g}$ is the curvature perturbation produced via the standard inflationary paradigm,

$$\zeta_{\rm g} = \frac{\partial N}{\partial \sigma_k}\delta\sigma_k\,. \tag{4.29}$$

Here we have assumed for simplicity that scales of cosmological interest exit the Hubble radius after ϕ has rolled to zero. Thus we treat the relevant stages of inflation as driven by the σ field, assisted by the vacuum energy $m_\chi^4/4g$, with ϕ being a heavier field fluctuating about the origin. We neglect the non-Gaussian component of ζ produced when relevant scales exit the Hubble radius as this is in general relatively small [21]. For comparison to observation it does not matter whether the non-Gaussian fluctuations are sourced by the same field as the Gaussian fluctuations [63]. Therefore we can combine Eq. (4.4), Eq. (4.28) and Eq. (4.29) to obtain

$$f_{\rm NL} = -\frac{5}{3}\left(\frac{\partial N}{\partial \sigma_{\rm e}}\delta\sigma_{\rm e}\right)\left(\frac{\partial N}{\partial \sigma_k}\delta\sigma_k\right)^{-2} = \frac{5}{3}\frac{\lambda_\phi}{\lambda_\sigma}\frac{1}{\sigma_{\rm e}}\frac{\partial N}{\partial \sigma_{\rm e}}\left(\frac{\partial N}{\partial \sigma_k}\right)^{-2}. \tag{4.30}$$

If we take $V_\sigma = \frac{1}{2}m_\sigma^2\sigma^2$, it is straightforward to calculate the largest possible $f_{\rm NL}$ by translating the arguments of Section 4.4. In particular, we note that $\sigma_{\rm e} = m_\chi/\sqrt{\lambda_\sigma}$ and

$$\frac{\partial N}{\partial \sigma_{\rm e}} \simeq \frac{1}{4g}\frac{m_\chi^4}{\sigma_{\rm e} m_\sigma^2 m_{\rm pl}^2} = \frac{\sqrt{\lambda_\sigma}}{4g}\frac{m_\chi^3}{m_\sigma^2 m_{\rm pl}^2}. \tag{4.31}$$

Likewise, minimizing the factor $(\partial N/\partial\sigma_k)^{-2}$ gives $\sigma_k = m_\chi^2/\sqrt{2g}m_\sigma$ and

$$\frac{\partial N}{\partial \sigma_k} = \frac{1}{\sqrt{2g}}\frac{m_\chi^2}{m_\sigma m_{\rm pl}^2}. \tag{4.32}$$

Finally, putting all this together we find

$$f_{\rm NL} \simeq \frac{5}{6}\frac{\lambda_\phi m_{\rm pl}^2}{m_\chi^2}. \tag{4.33}$$

It must be emphasized this is an upper limit; smaller $f_{\rm NL}$ are easily achieved by choosing parameters that do not minimize $(\partial N/\partial\sigma_k)^{-2}$. Note also that in this implementation of the model the only constraint on λ_ϕ is that $\lambda_\phi\delta\phi^2 = \lambda_\phi(H_k/\sqrt{2})^2 \ll m_\chi^2$. This constraint assures that χ always rolls away from the origin before σ reaches zero, and gives

$$\lambda_\phi\left(\frac{1}{12g}\frac{m_\chi^2}{m_{\rm pl}^2}\right) \ll 1. \tag{4.34}$$

Referring to second constraint of Eqs. (4.15), we see the term in parenthesis is already constrained to be much less than unity. Therefore this mechanism permits a non-Gaussian component to the density perturbations up to the observational limit of $f_{\rm NL} \lesssim 135$ [63].

For the above calculation to be appropriate requires that the $\zeta_{\rm g}$ of Eq. (4.29) actually be the dominant contribution to the curvature perturbation. For this to be the case its power spectrum must have a magnitude $\mathcal{P}_{\zeta,{\rm g}} \approx (5\times 10^{-5})^2$ in order to match observation [49]. From Eq. (4.6) and Eq. (4.32) we have

$$\mathcal{P}_{\zeta,{\rm g}} = \left(\frac{\partial N}{\partial \sigma_k}\right)^2\frac{H_k^2}{4\pi^2} = \frac{1}{8\pi^2\epsilon_k}\frac{H_k^2}{m_{\rm pl}^2}, \tag{4.35}$$

where we have used that in this scenario the first slow-roll parameter as the mode k exits the Hubble radius is

$$\epsilon_k = \frac{1}{2m_{\rm pl}^2}\left(\frac{\partial N}{\partial \sigma_k}\right)^{-2} = \frac{g\, m_\sigma^2 m_{\rm pl}^2}{m_\chi^4}\,. \tag{4.36}$$

We have written the power spectrum in this way in order to employ the observational constraints on the spectral tilt and on the level of gravity waves.

Since the primary, Gaussian density perturbations are now sourced during inflation, the spectral tilt is $n-1 = 2\eta - 6\epsilon_k \approx 2\epsilon_k \lesssim 0.04$ [51], where we have used that the second slow-roll parameter in the scenario we are considering is given by

$$\eta \equiv \frac{m_{\rm pl}^2}{V}\frac{\partial^2 V}{\partial \sigma^2} \approx \frac{4g\, m_\sigma^2 m_{\rm pl}^2}{m_\chi^4} \approx 4\epsilon_k\,. \tag{4.37}$$

Meanwhile, it is observed that $H_k^2/m_{\rm pl}^2 \lesssim 10^{-8}$ [53]. In this model we can decrease the spectral tilt independently of $H_k^2/m_{\rm pl}^2$ by simply decreasing m_σ. Therefore we simply require to set $H_k^2/m_{\rm pl}^2 \lesssim 10^{-9}$ to match observation. Here

$$\frac{H_k^2}{m_{\rm pl}^2} = \frac{1}{6g}\frac{m_\chi^4}{m_{\rm pl}^4} = \frac{5}{3}\lambda_\phi f_{\rm NL}^{-1}\left(\frac{1}{12g}\frac{m_\chi^2}{m_{\rm pl}^2}\right), \tag{4.38}$$

where we have expressed the result in terms of $f_{\rm NL}$ to clarify which are the remaining free parameters. According to the second constraint in Eq. (4.15), the term in parentheses is already constrained to be much less than unity. In fact, it can be set as small as necessary to satisfy observation. Moreover, since the three terms in Eq. (4.38) depend on three independent parameters (λ_ϕ, g, and m_χ), we have considerable freedom in exactly how we satisfy the observational bound. Thus we conclude that this model allows for significant non-Gaussianities for a range of model parameters.

4.6 Discussion and Conclusions

In this work we continue the analysis of the model proposed in Ref. [25] to generate density perturbations at the end of inflation. We confirm that these density perturbations can easily dominate over those produced via the standard inflationary paradigm, and explore the sensitivity of this result to the tuning of model parameters. In addition, we provide a straightforward variation of this model which allows for even greater amplification of the density perturbations.

It is worthwhile to consider how general is this analysis. According to Eq. (4.7) the production of density perturbations at the end of inflation is most effective when the slow-roll parameter near the end of inflation is much less than that when cosmological scales of interest exit the Hubble radius. Since inflation can only end with the slow-roll parameter rising to unity, this suggests the mechanism is most effective only when the inflationary potential contains large derivatives near the point where it dips toward its minimum. These large derivatives are most naturally accomplished by inserting a second field direction for the vacuum energy to fall to zero and initiate a reheating phase. This is precisely the scenario implemented in hybrid inflation and generalized in the models considered here.

We also study a variation in the implementation of the model proposed in Ref. [25] that results in modified inflationary dynamics. We show that this case results in a spectrum of density perturbations with significant non-Gaussianities for a range of model parameters. In particular it is shown that these non-Gaussianities are capable of saturating the current observational bound.

Acknowledgments

We would like to thank Michael Graesser and Mark Wise for helpful discussions and Michael Graesser for useful comments upon his review of the manuscript. We also thank David Lyth for correcting an assumption made in an earlier report of this work and for other helpful comments. This work was supported by the Department of Energy under the contract DE-FG03-92ER40701.

Chapter 5

The Scale of Gravity and the Cosmological Constant within a Landscape

It is possible that the scale of gravity, parameterized by the apparent Planck mass, may obtain different values within different universes in an encompassing multiverse. We investigate the range over which the Planck mass may scan while still satisfying anthropic constraints. The window for anthropically allowed values of the Planck mass may have important consequences for landscape predictions. For example, if the likelihood to observe some value of the Planck mass is weighted by the inflationary expansion factors of the universes that contain that value, then it appears extremely unlikely to observe the value of the Planck mass that is measured within our universe. This is another example of the runaway inflation problem discussed in recent literature. We also show that the window for the Planck mass significantly weakens the anthropic constraint on the cosmological constant when both are allowed to vary over a landscape.

Based on M. L. Graesser and M. P. Salem, eprint astro-ph/0611694.

5.1 Introduction

Theoretical results from inflationary cosmology [29, 30, 31, 87] and from string theory [88, 89, 90, 91, 92, 28] motivate the possibility of an eternally inflating multiverse that is populated by an infinite number of sub-universes, each obtained via local tunneling, diffusion, and/or classical slow-roll into one of a myriad of allowed metastable states. (For other motivations to consider such a landscape see for example Refs. [93, 94, 95, 96, 97].) In this landscape picture, each of these universes may contain different values for physical parameters, or even different particles and interactions, from those that are observed within our local universe. However, the anthropic principle [34, 35, 36] asserts that the physical laws that may be observed within any universe must be restricted to those that permit the evolution of observers in the first place.

Combined with the anthropic principle, the landscape picture has emerged as a plausible explanation for many striking features of our universe. In particular, it has been used to justify the "unnatural" smallness of the Higgs mass [98, 99, 100], the cosmological constant [32, 33] (for earlier work see [101, 102, 35, 103]), and the neutrino masses [104]; to predict the size of supersymmetry breaking [105, 106, 107, 108, 109, 110, 111]; to describe the tilt in the spectrum of density perturbations [112] and other inflationary parameters [113]; to constrain the baryon to photon ratio [114] and the ratio of baryons to dark matter [115, 116, 117], as well as to explain some seemingly fine-tuned relationships between parameters describing the theories of quantum electrodynamics and quantum chromodynamics [118]. Nevertheless, generating precise predictions from a landscape picture faces several major challenges.

One of these challenges is to identify an anthropic criteria that is both specific and compelling. Yet even after such an anthropic condition has been defined, it is a daunting task to discern its environmental requirements, to deduce their implications for physical parameters, and then to derive the associated anthropic constraints. One may proceed by considering the variation in only one physical parameter, starting from its value within our universe. However, apparently tight constraints on any single parameter may be significantly weakened when more than one parameter is allowed to vary. This seems to be the case with both the Higgs mass [119] (see however Ref. [120]) and the cosmological constant [121, 122, 123]. Moreover, all anthropically allowed universes may not be connected by the continuous variation of physical parameters. For example, the seemingly viable "cold big bang" universe [124] results from independently varying several cosmological parameters to values very far from those obtained within our universe.

In addition, to calculate the expectation values of physical parameters within a landscape requires determining an appropriate measure to weight among the various possible universes [125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140]. That is, a precise anthropic criteria does not account for all of the selection effects that contribute to the probability for a particular universe to be observed. Universes may be more or less likely based on how readily they are obtained via the physical dynamics that govern the multiverse. To account for this requires a complete understanding of the multiverse, its landscape, and the governing theory. Indeed, there is a more subtle challenge underlying this program, which is to develop an appropriate and self-consistent calculus to regulate calculations involving the infinite number of infinitely expansive universes that may be contained within the multiverse [125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140].

Nevertheless, a set of hypotheses to resolve these challenges may be excluded if it predicts a very low likelihood to observe a universe with some physical characteristic that our universe possesses. For example, consider a proposal that includes a specific notion of observer, a consistent calculus to determine expectation values over the multiverse, and a theory to describe the landscape including how metastable states are mapped onto universes within the multiverse. If this proposal then predicts that an exponentially small number of observers measure a cosmological constant at or below the value obtained in our universe, then the proposal and the specific landscape in question are probably not both correct.

We investigate the possibility that the scale of gravity may scan over the landscape. This is consistent with the results of Ref. [141], where only parameters with mass dimension were found to vary over a model of the landscape. We everywhere parameterize the scale of gravity using the (reduced) Planck mass $m_{\rm P}$. Although $m_{\rm P}$ is commonly taken to be a fixed fundamental scale, this need not be the case. For example, the multiverse may be governed by a low-energy effective theory with the Lagrangian,

$$\mathcal{L} = \frac{1}{2}\sqrt{-g}M^2 F[\phi]R + \mathcal{L}_\phi + \mathcal{L}' \,, \tag{5.1}$$

where the fundamental mass scale is M, R is the Ricci scalar, ϕ is the collection of fields that specify the metastable state of a universe, $\mathcal{L}_\phi$ is the effective Lagrangian for these fields, and $\mathcal{L}'$ is the effective Lagrangian for matter. We assume that within each metastable state the fields ϕ are very massive and fixed to values $\phi \to \phi_*$ and are therefore non-dynamical. The strength of gravity will therefore be a constant within each metastable state, determined by the effective Planck mass,

$$m_{\rm P} \equiv \sqrt{F[\phi_*]}M \,. \tag{5.2}$$

We also assume that within each universe $m_{\rm P}$ is fixed prior to slow-roll inflation.

Alternatively, our analysis may be viewed in the so-called Einstein frame where the scale of gravity is everywhere fixed. To accomplish this, one simply performs the conformal transformation $g_{\mu\nu} \to F^{-1}[\phi_*]\tilde{g}_{\mu\nu}$. Then the scale of gravity is everywhere M, but all other parameters with dimension mass are scaled by the factor $F[\phi_*]^{-1/2}$. Thus our analysis is equivalent to fixing the Planck mass to be the fundamental scale of physics but varying all other mass scales uniformly. Stated another way, in our analysis the frame-independent ratio of masses $m/m_{\rm P}$ scales as $F[\phi_*]^{-1/2}$, where for example m may be the cutoff of the theory, the Higgs mass, or the scale of strong dynamics. This picture was previously suggested in the penultimate section of Ref. [119]. The idea that the effective value of $m_{\rm P}$ may vary across the multiverse within the context of Brans-Dicke theory was studied in Refs. [128, 129, 130, 142]. In addition, Ref. [143] studied a model of the form of Eq. (5.1) to show how inflationary dynamics can explain the hierarchy between the apparent Planck scale and the electroweak scale. As described below, our focus is different from the focus of this work.

We calculate the range over which $m_{\rm P}$ may scan while still satisfying anthropic constraints. We restrict our attention to universes that possess the same particles, interactions, couplings, and physical scales that are observed within our universe. Note that this means that the cut-off to the low-energy effective theory depends on M, not $m_{\rm P}$. Within the context of the model described by

Eq. (5.1), these assumptions require the existence of a large number of states that have approximately equal particle physics parameters yet different values of $m_{\rm P}$. We do not explore the interesting case of a landscape model that permits only correlated changes in $m_{\rm P}$ and the other particle physics parameters. Note that the above restrictions are conservative in the sense that lifting them can only expand the range of allowed $m_{\rm P}$.

Many anthropic constraints relate to the formation of galaxies and depend on the spectrum of energy density fluctuations evaluated at matter-radiation equality. On any given distance scale, this spectrum has an approximately Gaussian distribution about some root-mean-square (rms) amplitude, and we calculate constraints as if all fluctuations have the rms amplitude. Perhaps not surprisingly, the range of $m_{\rm P}$ that is consistent with all of the anthropic constraints is rather narrow. It also depends on what models are chosen for inflation, baryogenesis, and the dark matter. As an example, if we assume that inflation is chaotic with potential $V(\varphi) = \frac{1}{2}m_\varphi^2\varphi^2$, that baryogenesis results from efficient leptogenesis, and assume a weakly interacting massive particle (WIMP) to be the dark matter, then anthropic considerations combine to constrain $m_{\rm P}$ to be $0.1 \lesssim \hat{m}_{\rm P} \lesssim 1.5$, where $\hat{m}_{\rm P}$ is the ratio between $m_{\rm P}$ and the value obtained within our universe: 2.4×10^{18} GeV.

Even a very narrow anthropic range for $m_{\rm P}$ may have significant consequences for proposals to calculate its expectation value within a landscape. In particular, it is plausible that some proposals will ultimately weight universes in part according to their inflationary expansion factor. This expansion factor depends exponentially on the number of e-folds of inflation that the universe undergoes, which in turn depends on $m_{\rm P}$. In this case the probability distribution for $m_{\rm P}$ will be peaked only where some other selection effect cancels this strong exponential dependence. This other selection effect could be a very sharp peak or boundary to the underlying landscape distribution; otherwise the effect must come from an exponentially strong anthropic dependence on $m_{\rm P}$. Yet such a strong anthropic dependence on $m_{\rm P}$ would be in conflict with the observation that $m_{\rm P}$ has even a narrow anthropic window in our universe. Thus we are forced to conclude that under these weighting schemes the observation of our universe is either extremely atypical or our value of $m_{\rm P}$ sits at some sharp peak or boundary in the underlying landscape distribution. This point is completely analogous to the "σ-problem" and "Q catastrophe" identified in Refs. [144, 145].

We here note that a runaway problem associated with varying the effective Planck mass during eternal inflation has already been discussed in Refs. [142, 128, 129, 130]. These papers studied the evolution of $m_{\rm P}$ in Brans-Dicke theory when the Brans-Dicke field is allowed to be dynamical during inflation. On the other hand, we study the case where the fields ϕ in Eq. (5.1) are very massive and therefore non-dynamical. The difference between these scenarios is subtle because at some level ϕ must be dynamical in order for the landscape to be populated within the multiverse. Our approach is to assume that the fields ϕ are only dynamical at the very high energies that dominate the dynamics of the multiverse. At these energies the dynamics of ϕ could be described as in Refs. [142, 128, 129, 130] or they could be described by different effects. We simply treat these dynamics as unknown except to assume that the ϕ are fixed prior to the slow-roll inflation that eventually reheats into each of the anthropically favorable low-energy universes such as our own.

We also consider the anthropic window for the cosmological constant Λ when both Λ and $m_{\rm P}$ are allowed to (independently) scan over the landscape. Even when the allowed range for $m_{\rm P}$ is relatively narrow, it still allows for a significant broadening of the allowed range for Λ. To see this, note that Λ is constrained only by Weinberg's anthropic bound [32, 33],

$$\rho_\Lambda \lesssim \rho_{\rm eq}\sigma_{\rm eq}^3 \,. \tag{5.3}$$

Here ρ_Λ is the energy density in cosmological constant, $\rho_{\rm eq}$ is the matter density at matter-radiation equality, and $\sigma_{\rm eq}$ is the typical fluctuation in matter density at equality. The broadening occurs because for WIMP dark matter, decreasing $m_{\rm P}$ significantly increases $\rho_{\rm eq}$ and for most models also significantly increases $\sigma_{\rm eq}$. For example, if we again assume chaotic inflation with potential $V(\varphi) = \frac{1}{2}m_\varphi^2\varphi^2$ and that baryogenesis results from efficient leptogenesis, then ρ_Λ may be over a million times the value observed within our universe when $\hat{m}_{\rm P} \gtrsim 0.1$. Of course, a larger anthropic window for ρ_Λ does not necessarily imply that our value of ρ_Λ is less likely to be observed. We illustrate the distribution of observed values of ρ_Λ with a very simplified calculation. The results of this calculation suggest that to observe the cosmological constant at or below the level obtained within our universe is very unlikely unless the landscape distribution of $m_{\rm P}$ is dominated by values very near to or larger than the value obtained within our universe.

The remainder of this chapter is organized as follows. In Section 5.2 we calculate the range of $m_{\rm P}$ allowed by anthropic constraints in universes otherwise like ours. Constraints come from a variety of cosmological processes and we summarize our results in Section 5.2.8. Then in Section 5.3 we argue that the value of $m_{\rm P}$ that we observe is extremely unlikely if universes within the landscape are weighted by their inflationary expansion factor. In this section we also discuss some caveats to this argument. The scenario where both $m_{\rm P}$ and Λ may vary across the landscape is discussed in Section 5.4. The next two sections compliment work in the main body of this chapter. Section 5.5 discusses some anthropic constraints for cosmological histories that differ signficantly from ours. Section 5.6 demonstrates how runaway inflation problem dominates what might otherwise seem apparently tight selection effects. Finally, we draw our conclusions in Section 5.7.

5.2 Anthropic Constraints on the Scale of Gravity

It is straightforward to organize the immediate effects of changing the scale of gravity when all other mass scales and couplings are kept fixed (this implies that the cut-off of the theory is also fixed).

Then scanning the Planck mass corresponds to changing the proportionality constant between the Einstein and the stress-energy tensors,

$$G_{\mu\nu} = \frac{1}{m_{\rm P}^2} \, T_{\mu\nu} \, . \tag{5.4}$$

In a homogeneous universe this simply changes the relationship between the Hubble rate H, its time rate of change $\dot{H}$, and the energy (ρ) and pressure (P) densities,

$$H^2 = \frac{\rho}{3m_{\rm P}^2} \, , \quad \dot{H} = -\frac{\rho + P}{2m_{\rm P}^2} \, . \tag{5.5}$$

This bears upon anthropic conditions because the Hubble rate determines when particle interactions freeze out of equilibrium. This affects the relative densities of, for example, matter to radiation and protons to neutrons.

Of course the universe is only approximately homogeneous. According to the present understanding, inhomogeneities are generated by quantum fluctuations in at least one scalar field as it exits the Hubble radius during (nearly) de Sitter inflationary expansion in the early universe. If this is the case, then the Hubble rate also effects the size of the initial inhomogeneities. At late times, these inhomogeneities re-enter the Hubble radius and the scale of gravity takes on a new role. Then gravity provides a self-interaction to over-densities that may cause them to grow. Over-densities that grow too large become gravitationally bound and separate from the cosmic expansion. Within these structures, the expansion of the universe is inconsequential but the scale of gravity still determines the internal dynamics.

We analyze the anthropic significance of these effects in chronological order, beginning with the effects on inflation. We then discuss baryogenesis, big bang nucleosynthesis, matter domination, structure formation, stellar dynamics, and finally the stability of stellar systems. The anthropic constraints are displayed in Figs. 5.1–5.4 and summarized in Section 5.2.8. The complexity of this analysis, along with the many uncertainties in our understanding of various cosmological processes, make a precise determination of anthropic constraints impractical. Therefore we strive for approximations that capture the key effects of scanning $m_{\rm P}$. Usually, we determine the dominant scaling behavior of a quantity with $m_{\rm P}$ and cite a precisely determined value from our universe to determine the value in another universe. Unless otherwise stated, the values for cosmological parameters within our universe are taken from the tables in Ref. [117] (note however that we work in terms of the reduced Planck mass). Throughout this chapter we denote the ratio of a quantity to the value that it obtains within our universe using a hat, for example

$$\hat{m}_{\rm P} \equiv \frac{m_{\rm P}}{2.4 \times 10^{18} {\rm GeV}} \, . \tag{5.6}$$

Finally, we use units where $\hbar = c = k_B = 1$.

5.2.1 Inflation

An early period of inflation is believed to have homogenized our universe and yet provided the seeds of cosmic structure through the generation of small density perturbations (for reviews of inflation see for example [10, 11]). We parameterize these effects using the total number of e-folds of inflation N and the Bardeen curvature perturbation ζ. In principle, both N and ζ are constrained by anthropic considerations. Meanwhile, for inflation to occur in the first place requires that the Hubble radius expand at a rate slower than the speed of light. This effect is parameterized by requiring that the first slow-roll parameter, $\epsilon_I \equiv -\dot{H}/H^2$, is smaller than unity. Finally, at some point inflation must end and the universe must reheat to establish the initial conditions for the subsequent big bang evolution. We discuss the $m_{\rm P}$ dependence of each of these below.

5.2.1.1 Satisfying Slow-Roll for N e-folds of Inflation

For inflation to occur in the first place requires that at some time $\epsilon_I < 1$. When inflation is driven by the potential energy of a canonical scalar field φ, the first slow-roll parameter can be written

$$\epsilon_I \simeq \frac{m_{\rm P}^2}{2}\left(\frac{V_\varphi}{V}\right)^2 , \tag{5.7}$$

where V is the inflaton potential and the subscript on V_φ denotes differentiation with respect to φ. Although at first glance ϵ_I appears to increase with increasing $m_{\rm P}$, this can usually be compensated for by starting the inflaton φ further up the potential. This is the case with each of the canonical inflationary models presented below. Therefore we assume that the occurrence of inflation in the first place does not significantly constrain $m_{\rm P}$.

Meanwhile, the total duration of inflation is constrained by the need to homogenize a universe large enough to allow for the formation of structure. This constraint, however, is very weak, since our observable universe appears immensely larger than is necessary to form a galaxy. Moreover, inflationary scenarios that predict the observed value of ζ typically allow for far more e-folds of inflation than are necessary to enclose our universe. Therefore we assume that the anthropic bound on N does not significantly constrain $m_{\rm P}$.

5.2.1.2 The Curvature Perturbation ζ

Anthropic constraints on the amplitude of ζ stem from primordial black hole production, structure formation, and the stability of stellar systems and are described in Sections 5.2.4, 5.2.5, and 5.2.7. Presently we discuss the dependence $\zeta(m_{\rm P})$ for future reference. For inflation driven by a canonical

Mechanism to generate $\zeta_{\rm eq}$	$m_{\rm P}$ dependence of $\zeta_{\rm eq}$
Inflation with $V(\varphi) \propto \varphi^p$	$m_{\rm P}^{p/2-2}$
Natural inflation	$m_{\rm P}^{-3}$
Hybrid inflation	$m_{\rm P}^{-3}$
Ghost inflation	$m_{\rm P}^{-5/2}$
End of inflation scenario	$\chi\, m_{\rm P}^{-3}$
Curvaton scenario	$\chi^{-1} m_{\rm P}^{-1}$
Inhomogeneous reheating	$\chi^{-1} m_{\rm P}^{-1}$ or $\chi\, m_{\rm P}^{-1}$

Table 5.1: The dependence of $\zeta_{\rm eq}$ on $m_{\rm P}$ for a variety of mechanisms to generate the curvature perturbation. The result for inhomogeneous reheating depends on the relative size between χ and other mass scales in the Lagrangian, and can interpolate between the two dependences given above.

scalar φ, the curvature perturbation on a co-moving scale with wave-vector k is

$$\zeta(k) \propto \left. \frac{V^{3/2}}{m_{\rm P}^3 V_\varphi} \right|_{k=aH} , \tag{5.8}$$

where V and V_φ are evaluated when the scale k exits the Hubble radius. Anthropic constraints on ζ apply to scales $k \lesssim k_{\rm eq}$, where $k_{\rm eq}$ is the wave-vector of the Hubble radius at matter-radiation equality. The potential V evaluated when these scales first exit the Hubble radius may depend upon $m_{\rm P}$ even when $V(\varphi)$ does not.

In our universe, ζ does not change appreciably with k, and we assume this holds at least approximately in other universes. Therefore we take $\zeta \approx \zeta(k_{\rm eq}) \equiv \zeta_{\rm eq}$ over all scales of interest. The scale $k_{\rm eq}$ of matter-radiation equality itself depends on $m_{\rm P}$. However this dependence is logarithmic and its effect on $\zeta_{\rm eq}$ is suppressed by the smallness of the slow-roll parameter, so we ignore it. Still, to solve for $\zeta_{\rm eq}$ requires to choose a specific model for inflation. Since there is no standard model of inflation, we must be content with a only a plausible range for the dependence on $m_{\rm P}$. We deduce this range by studying several of the most popular models of inflation. The results for chaotic inflation [81] with $V(\varphi) \propto \varphi^p$, for hybrid inflation [80], for natural inflation [146], and for ghost inflation [147] are listed in Table 5.1.

The curvature perturbation $\zeta_{\rm eq}$ may also be generated at the very end of inflation [25, 27] or even much later as in the curvaton [22, 23, 24] and inhomogeneous reheating [18, 19, 20, 26] scenarios. In each of these models a sub-dominant scalar χ receives fluctuations while the fluctuations to the inflaton are presumed to be negligible. The fluctuations in χ are then transferred to radiation either at the end of inflation, during reheating, or during a phase transition much later. In each of these cases, the amplitude of the curvature perturbation depends on the local vacuum expectation value (vev) of χ. The dependence of $\zeta_{\rm eq}$ on χ and $m_{\rm P}$ when each of these mechanisms operates efficiently (that is, when any reheating occurs far out of equilibrium) is listed in Table 5.1.

We note that these models are very flexible to anthropic selection. Although in principle χ may be set by interactions such that it is fixed among the set of universes we consider, in most cases χ is a stochastic variable over these universes. When $\zeta_{\rm eq}$ is generated at the end of inflation or via the curvaton or inhomogeneous reheating scenarios, then this implies that $\zeta_{\rm eq}$ is also a stochastic variable over different universes. Therefore in these cases there exist universes with far different $m_{\rm P}$ but the same $\zeta_{\rm eq}$ as in our universe, as well as universes with the same $m_{\rm P}$ but different $\zeta_{\rm eq}$.

For future convenience we write $\zeta_{\rm eq}$ in the form,

$$\hat{\zeta}_{\rm eq} \approx \hat{m}_{\rm P}^{-\alpha}\,, \tag{5.9}$$

where according to our notation $\hat{\zeta}_{\rm eq}$ is the curvature perturbation relative its value in our universe. The various models of inflation that we studied suggest that we should restrict α to the range $0 \leq \alpha \leq 3$. However, the mechanisms that generate $\zeta_{\rm eq}$ at the end of or well after inflation may generate a wide range of $\zeta_{\rm eq}$ for a wide range of $m_{\rm P}$. Although it is technically possible that keeping inflationary parameters fixed but varying $m_{\rm P}$ will cause the dominant contribution to the curvature perturbation to shift from one mechanism to another, this scenario should still be well approximated by the above guidelines so long as the variation in $m_{\rm P}$ is not too large.

5.2.1.3 Reheating

The reheating of the universe after inflation is achieved by coupling the inflaton to other degrees of freedom. In a typical model, after inflation the inflaton rocks within its potential well and redshifts like matter. The radiative decay products of the inflaton then dominate the energy density of the universe only after the Hubble rate falls below the decay width Γ_I. At this time the energy density in the inflaton is $\rho_I = 3m_{\rm P}^2\Gamma_I^2$. If Γ_I is independent of $m_{\rm P}$, then the reheat temperature scales like

$$\hat{T}_{RH} = \hat{m}_{\rm P}^{1/2}\,. \tag{5.10}$$

We have no empirical knowledge about reheating other than that T_{RH} is above the temperature of big bang nucleosynthesis. However, if our universe is described by a grand unified theory, then T_{RH} must be below the temperature of monopole production. In addition, if net baryon number is not generated during reheating, then T_{RH} must be high enough to support the dominant mechanism of baryogenesis. We comment on constraints like these in Section 5.3.

5.2.2 Baryogenesis

We parameterize the net baryon number of a universe with the ratio between the number density of baryons and the number density of photons: $\eta \equiv n_b/n_\gamma$. Although we see no direct anthropic constraints on the value of η, it will enter into the anthropic constraints described in Sections 5.2.5

and 5.2.7. Presently, we seek to parameterize the dependence of η on m_{P} for future reference. As with inflation, there is no standard model of baryogenesis. Therefore we must again content ourselves with only a range for the dependence on m_{P}, based on the most plausible mechanisms. For a summary of these see, for example, the reviews of Refs. [148, 149, 150].

Perhaps the most plausible mechanism to produce net baryon number is leptogenesis [151, 152]. For example, net lepton number is rather easily obtained by the out-of-equilibrium decay of a right-handed neutrino (RHN). The resulting lepton asymmetry can then be converted into net baryon number by sphaleron transitions within the Standard Model [153]. The value of η that results from leptogenesis depends on how far out of equilibrium the RHN decays. If RHN decay occurs far out of equilibrium then the resulting baryon asymmetry η is independent of m_{P}. Otherwise, it scales roughly according to $\eta \propto m_{\mathrm{P}}^{-1}$. Note that baryogenesis via leptogenesis requires that the RHN, for example, be produced in the first place. We comment on this requirement in Section 5.3.

The Standard Model of particle physics itself generates appropriate conditions for baryogenesis, when it is augmented by relatively light supersymmetric (SUSY) scalars to strengthen the electroweak phase transition. The process of electroweak baryogenesis is complex; yet interestingly it operates independently the scale of gravity. That is, although universal expansion is necessary to decrease the temperature of the universe and thus spur the electroweak phase transition, this process is relatively independent the rate of temperature change. Therefore electroweak baryogenesis gives a baryon asymmetry that is independent of m_{P} [148, 149, 150].

Finally, we look at Affleck Dine (AD) baryogenesis [154]. This mechanism takes advantage of scalar fields that possess baryon or lepton number, as would exist in SUSY or a grand unified theory (GUT). During inflation these fields may acquire large vevs, and then the influence of baryon non-conserving interactions on their subsequent evolution may generate significant baryon number. There are many models to implement AD baryogenesis and each may give a different dependence on m_{P}. We simply give the result for a set of scenarios described in Ref. [155], where the scalar fields overlap a SUSY flat direction that is lifted by non-renormalizable interactions and a negative induced mass term during inflation. To fit the resulting baryon asymmetry to observation requires different parameterizations when different non-renormalizable interactions dominate; however in each case the dependence on m_{P} is given by $\eta \propto m_{\mathrm{P}}^{-3/2}$ [155].

For future reference, it is convenient to write the baryon to photon ratio in the form

$$\hat{\eta} \approx \hat{m}_{\mathrm{P}}^{-\beta} \,. \tag{5.11}$$

From the above discussion, we expect $0 \leq \beta \leq 3/2$, with perhaps the most plausible values being $\beta = 0$ or $3/2$. As was the case with inflation, if more than one mechanism contributes to the net baryon number then we still expect the η that results to be well approximated by the above guidelines.

5.2.3 Big Bang Nucleosynthesis

The process of big bang nucleosynthesis (BBN) populates the universe with light elements. In particular, the mass fractions of hydrogen (X) and helium (Y) are important for anthropic considerations described in Sections 5.2.5 and 5.2.6. Since we are only interested in X and Y, we may take a very simplified view of BBN. Specifically, we assume that BBN generates appreciable concentrations of only hydrogen and helium-4. This is clearly appropriate within our universe, where the other products of BBN account for only about 0.01% of the mass fraction of the universe. Although this fraction may change significantly for differing values of m_{P}, it would take a very large variation in m_{P} for this change to become significant next to X or Y. A basic description of BBN can be found in Ref. [9].

When we approximate BBN to result in only hydrogen and helium-4, we require only the ratio of neutrons (n) to protons (p) to deduce X and Y. Specifically,

$$X \approx \frac{1-n/p}{1+n/p}\,, \qquad Y \approx 1-X\,. \tag{5.12}$$

A free neutron has $\Delta E \approx 1.3$ MeV more energy than a free proton. Thus if neutrons and protons are kept in thermal equilibrium by interactions that convert each into the other, then n/p is given by the Boltzmann factor,

$$n/p = \exp\left(-\Delta E/T\right)\,, \tag{5.13}$$

where T is the temperature. Neutrons and protons are converted into each other via interactions such as $n+\nu \longleftrightarrow p+e$, where the ν denotes an electron neutrino and the e an electron. In fact, the rates of these interactions decrease faster with temperature than does the Hubble rate. Therefore, below some temperature T_F the universe expands too rapidly for, for example, an n and a ν to find each other and convert into a p and an e. Below this temperature the relative concentrations of n and p are fixed and the interaction is said to freeze-out.[1]

The freeze-out temperature T_F is obtained by equating the total rate of interactions converting neutrons to protons, Γ_{np}, with the Hubble rate H. Then the neutron to proton ratio is $n/p \approx \exp\left(-\Delta E/T_F\right)$. Our universe contains $n/p \approx 1/7$ such that $X \approx 3/4$. Note that varying m_{P} so as to decrease T_F works to decrease n/p and therefore push X closer to unity. Since in our universe

---

[1] The decay of neutrons decreases n/p from its value at T_F by about 14% within our universe. This loss is determined by the time at which neutrons are efficiently captured into Helium, which depends on m_{P} only via a logarithmic dependence on $\eta(m_{\mathrm{P}})$ [9]. Inspecting Fig. 4.4 of Ref. [9], it can be shown that in the most extreme case of $\beta = 3/2$, less than half of the neutrons decay for $\hat{m}_{\mathrm{P}} \lesssim 6$. On the other hand, for $\hat{m}_{\mathrm{P}} \gtrsim 6$ the helium fraction is less than a tenth of the hydrogen fraction. Therefore this effect is always negligible at our level of analysis.

$X \approx 3/4$, this effect is negligible at our level of analysis. On the other hand, for temperatures $T \geq T_F$ the rate of conversion between neutrons and protons is $\Gamma_{\mathrm{np}} \propto T^5$ and is independent of η. Since $H \propto T^2/m_{\mathrm{P}}$, the freeze-out temperature scales like $T_F \propto m_{\mathrm{P}}^{-1/3}$ when T_F is larger than in our universe. Therefore we find,

$$n/p \approx \exp\left[-\ln(7)\,\hat{m}_{\mathrm{P}}^{1/3}\right]\,. \tag{5.14}$$

The fractions X and Y are given by Eqs. (5.12). For example, when $\hat{m}_{\mathrm{P}} = 10$, 5, 0.2, and 0.1 we have $X \approx 1$, 0.9, 0.5, and 0.4, respectively.

5.2.4 Matter Domination

We have assumed that the early universe is dominated by relativistic degrees of freedom, at least since BBN. However as the universe cools, massive degrees of freedom eventually become non-relativistic. As it becomes non-relativistic, the number density of this matter becomes exponentially suppressed relative that of radiation. However, this dilution eventually causes matter to freeze out of equilibrium with the remaining radiation. Subsequently, the energy density of a massive species i will redshift as $\rho_i \propto m_i n_i$ while the energy density in radiation scales as $\rho_{\mathrm{rad}} \propto T n_{\mathrm{rad}}$. Thus it is inevitable that matter should ultimately come to dominate the energy density of the universe (structure formation constraints ensure that the cosmological constant does not become significant before matter domination).

Nevertheless, the energy density at matter-radiation equality and the fraction of matter in baryons are relevant to anthropic constraints described in Sections 5.2.5 and 5.2.7. In addition, our assumption that the early universe is radiation dominated does not hold if ζ is too large. In this case, primordial black holes may dominate the energy density of the universe while baryons are still relativistic. Then all of the baryons would be redshifted away or swallowed into black holes. This possibility is studied at the end of this section. In the following we neglect the neutrino content of the universe. Their influence upon cosmology is commonly viewed as insignificant and we do not expect this to change since as a hot relic their density relative baryons is fixed. In addition, we assume the dark matter to be a WIMP. This allows for relatively precise predictions, as opposed to, for example, axion dark matter where the density is set by a stochastic variable [156]. (Note however that the stochastic nature of axion dark matter makes this possibility more flexible to anthropic selection, see for example Refs. [115, 117].)

The energy density in a WIMP dark matter candidate is set by the relic abundance that results from the freeze-out of annihilation interactions when the temperature drops below the mass of the

WIMP. After matter-radiation equality this gives the scaling,

$$\rho_{\rm cdm} \propto m_{\rm P}^{-1}\, T^3 \,. \tag{5.15}$$

Meanwhile, after baryogenesis the relative abundance of baryons η is conserved. Therefore at temperatures below the nucleon mass the energy density in baryons scales like

$$\rho_b \propto \eta T^3 \,. \tag{5.16}$$

Using that in our universe $\rho_b/\rho_{\rm cdm} \approx 1/5$ and that the energy density in radiation scales as $\rho_{\rm rad} \propto T^4$, we find the energy density at matter-radiation equality to be

$$\hat{\rho}_{\rm eq} \approx \left(\frac{1}{6}\hat{\eta} + \frac{5}{6}\hat{m}_{\rm P}^{-1}\right)^4 \,. \tag{5.17}$$

Finally, we note the baryon fraction within matter,

$$\hat{f}_b \approx \left(\frac{1}{6} + \frac{5}{6}\hat{\eta}^{-1}\hat{m}_{\rm P}^{-1}\right)^{-1} \,. \tag{5.18}$$

As described above, these results do not hold if ζ is so large as to produce an abundance of primordial black holes (PBHs) [157, 158]. Numerical analysis reveals that a PBH is formed when an energy density fluctuation $\sigma \gtrsim 0.7$ enters the Hubble radius [159]. Meanwhile, during radiation domination $\sigma = \frac{4}{3}\zeta$ at Hubble radius crossing [74]. Therefore we require $\zeta(k) \lesssim 0.5$ in order to prevent the formation of a PBH when the scale k enters the Hubble radius. In fact, this implies a somewhat stronger constraint on ζ. This is because ζ is a stochastic variable with a Gaussian tail and because a PBH need not be formed each Hubble time in order for PBHs to dominate the energy density of the universe. This constraint is worked out in Ref. [121] and we follow that analysis.

The likelihood that a curvature perturbation with root-mean-square ζ is greater than or equal to $1/2$ is $f(\zeta) = \mathrm{erfc}(2^{-3/2}\zeta^{-1})$, where the complementary error function is defined $\mathrm{erfc}(x) \equiv 2\pi^{-1/2}\int_x^\infty e^{-z^2}dz$. PBHs redshift like matter while the other degrees of freedom redshift like radiation. Therefore by matter-radiation equality PBHs will compose roughly $(a_{\rm eq}/a_{\rm pbh})f$ of the energy density of the universe. Here $a_{\rm pbh}$ is the scale factor at which newly formed PBHs have sufficient mass to persist until equality. Then PBHs do not dominate the energy density of the universe when

$$\mathrm{erfc}(2^{-3/2}\zeta_{\rm eq}^{-1}) \lesssim \frac{a_{\rm pbh}}{a_{\rm eq}} \,, \tag{5.19}$$

where we have neglected any tilt in ζ. This approximation underestimates the largest $\zeta_{\rm eq}$ since recent observations suggest a negative tilt [53, 160].

The complementary error function depends very strongly upon its argument; therefore the $m_{\rm P}$ dependence of the ratio $a_{\rm pbh}/a_{\rm eq}$ is inconsequential for our analysis. As one looks to earlier times, the lifetimes of PBHs decreases more rapidly than cosmic time decreases. Therefore Eq. (5.19) ensures that PBHs dominate at no time prior to equality. Solving for when PBH lifetimes equal about 70,000 years gives $a_{\rm pbh}/a_{\rm eq} \sim 10^{-20}$ which gives $\zeta_{\rm eq} \lesssim 6 \times 10^{-2}$. Translating this into a constraint on $m_{\rm P}$, we find

$$\hat{m}_{\rm P}^{\alpha} \gtrsim 7 \times 10^{-4}\,, \tag{5.20}$$

where we have used that the density fluctuation at Hubble radius crossing is $\sigma \approx 5 \times 10^{-5}$ [53, 160]. Eq. (5.20) is always weaker than the stellar lifetime constraint of Section 5.2.6.

5.2.5 Structure Formation

The formation of structure within our universe occurs in several stages. First, over-densities in the nearly pressureless dark matter begin to grow upon entry into the Hubble radius. During radiation domination, this growth is logarithmic with time, while after the dark matter comes to dominate the energy density of the universe over-densities grow in proportion to the growth in the cosmic scale factor. On the other hand, over-densities in the baryons cannot grow until after recombination. However, within an e-fold or so after recombination they have grown to match the over-density in dark matter, and subsequently grow in proportion to the growth in the cosmic scale factor. When these over-densities have grown sufficiently they separate from the Hubble flow and virialize to form what are termed halos.

After virialization, the cold dark matter within halos is stabilized against gravitational collapse by its inability to release its kinetic energy. However, the baryons within the halo must collapse beyond their initial virialization radius if they are to fragment and condense into galaxies and ultimately into stars. This requires that the baryons have a means to dissipate their thermal energy. The constraints on $m_{\rm P}$ that are implied by these stages of structure formation are discussed in the sections below. Presently, we describe the initial growth in over-densities for future reference. The subject of galaxy formation and in particular star formation is complex and not yet fully understood. We rely heavily on the simplifying assumptions and models of Refs. [117, 121].

We find it convenient to track the evolution of over-densities in position space, as opposed to Fourier space. At matter-radiation equality, the variance of energy density fluctuations over scales with co-moving radius R is,

$$\sigma^2_{\rm eq}(R) = \int_0^\infty \frac{dk}{2\pi^2}\, k^2\, W(kR) T(k) P(k)\,, \tag{5.21}$$

where W is a window function that may be chosen to be a "top hat" with radius R, T is a transfer function to account for the evolution of perturbations between when they enter the Hubble radius and equality, and P is the primordial power spectrum of fluctuations, $P(k) \propto \langle \zeta(k)^2 \rangle$. We parameterize a co-moving scale with radius R according to the total mass μ that is enclosed within a sphere of radius R. In addition, we measure μ relative to the mass of our galaxy (more precisely the mass of our galaxy plus its dark matter halo); thus $\mu = 1$ corresponds to $10^{12} M_\odot$, where $M_\odot$ is the mass of the sun.

A numerical curve fit to Eq. (5.21) gives [117],

$$\sigma_{\rm eq} \simeq 1.45 \times 10^{-3}\, s(\mu) \hat{\zeta}_{\rm eq}(m_{\rm P})\,, \tag{5.22}$$

where the function $s(\mu)$ carries the scale dependence of $\sigma_{\rm eq}$. This scale dependence occurs because at the time of equality smaller scales have been within the Hubble radius for a longer time than larger scales. The function $s(\mu)$ is equivalent to Eq. (A13) in Ref. [117]. However, we have normalized $s(\mu)$ such that $s(1) = 1$. In addition, we define the variable μ with respect to a different scale than the authors of Ref. [117]. Therefore within this chapter s is given by,

$$s(\mu) = \Big[\Big(0.76 \ln[17 + \mu^{-1/3}] - 0.22\Big)^{-0.27} + 0.17 \mu^{0.18} \Big]^{-3.7}\,. \tag{5.23}$$

Note that s is a decreasing function of μ. In addition, note that $\sigma_{\rm eq}$ is the root-mean-square (rms) value of a Gaussian random field. Therefore constraints involving $\sigma_{\rm eq}$ (or $\hat{\zeta}_{\rm eq}$) are never sharp in the sense that they may be overcome by fluctuations that happen to be larger or smaller than is typical.

After recombination but before the domination of cosmological constant, a linear over-density is given by [104]

$$\sigma \approx \left(\frac{2}{5} + \frac{3}{5} \frac{a}{a_{\rm eq}} \right) \sigma_{\rm eq}\,. \tag{5.24}$$

Soon after recombination the first term is negligible. Eq. (5.24) is accurate until cosmological constant domination, after which σ grows by another factor of about 1.44 and then stops. An over-density separates from the Hubble flow and virializes when a linear analysis gives $\sigma = 1.69$ [37, 161]. Thus the cosmic mean energy density at virialization is,

$$\rho_* \approx \left(\frac{3}{5} \frac{\sigma_{\rm eq}}{1.69} \right)^3 \rho_{\rm eq} \approx 1.4 \times 10^{-10} \rho_{\rm eq} \hat{\zeta}_{\rm eq}^{\,3} s^3\,. \tag{5.25}$$

The energy density within the condensed halo is larger by roughly a factor of $18\pi^2$. We denote this,

$$\rho_{\rm vir} \approx 18\pi^2 \rho_* \approx 2.4 \times 10^{-8} \rho_{\rm eq} \hat{\zeta}_{\rm eq}^{\,3} s^3\,. \tag{5.26}$$

Note that these quantities depend on both $m_{\rm P}$ and the mass scale μ of the virialized halo.

The above description of halo formation relies on three important aspects of the standard cosmology: we assume that the dark matter density dominates over the baryon density, that recombination occurs before the virialization of the dark matter halo, and that virialization occurs before the domination of cosmological constant. Enforcing the above conditions implies constraints on $m_{\rm P}$. These are discussed in the next section. We discuss the possibility for a non-standard path toward structure formation in Section 5.5.

5.2.5.1 Halo Virialization

Before proceeding to galaxy formation, we must ensure that over-densities separate from the cosmic expansion and virialize before the domination of the cosmological constant halts their growth. As mentioned above, an over-density σ has separated from the Hubble flow when a linear analysis gives $\sigma \geq 1.69$. On the other hand, the maximum size that is reached by an rms linear over-density is given by,

$$\sigma_\infty \approx 1.44 \times \frac{3}{5}\frac{a_\Lambda}{a_{\rm eq}}\sigma_{\rm eq} \approx 3.20\,\hat{\rho}_{\rm eq}^{1/3}\hat{\zeta}_{\rm eq}s\,, \tag{5.27}$$

where a_Λ is the scale factor at ρ_Λ domination. Therefore the requirement that $\sigma_\infty \geq 1.69$ gives

$$\hat{\rho}_{\rm eq}\hat{\zeta}_{\rm eq}^{\,3}s^3 \gtrsim 0.1\,. \tag{5.28}$$

Since with a larger value for $m_{\rm P}$ structures form later, this is a constraint against increasing $m_{\rm P}$. Substituting previous results into Eq. (5.28) gives

$$\left(\frac{1}{6}\hat{m}_{\rm P}^{-\beta} + \frac{5}{6}\hat{m}_{\rm P}^{-1}\right)^4 \hat{m}_{\rm P}^{-3\alpha}s(\mu)^3 \gtrsim 0.1\,. \tag{5.29}$$

The curves that saturate this inequality for various choices of α and β are displayed in Figs. 5.1–5.4 under the label "$\sigma_\infty = 1.69$." In light of these plots, our value of $m_{\rm P}$ may be construed as nearly saturating this constraint. However, this perception derives from the strong $m_{\rm P}$ dependence of $\rho_{\rm vir}$, and not from our galaxy being at the edge of saturating the Weinberg bound. In addition this constraint, analogous to many others below, holds for an rms fluctuation $\sigma_{\rm eq}$ but is weaker for larger fluctuations. For these reasons we emphasize that although Eq. (5.29) does not allow for $m_{\rm P}$ to be increased significantly, it is still true that our value of $m_{\rm P}$ is not at the edge of the anthropic range. Since the existence of an anthropically allowed window surrounding our value of $m_{\rm P}$ is essential to the arguments of Section 5.3, we provide a more elaborate discussion of this boundary in Section 5.6. This may serve as an illustration of how 'soft' are other constraints that depend on $\sigma_{\rm eq}$.

The above analysis assumes that halos virialize at least an e-fold or so after recombination. This is to ensure that baryons may collapse into the dark matter potential wells and participate in the virialization. Thus we require $\rho_*/\rho_{\rm rec} \lesssim e^{-3}$. The energy density at recombination is set by the temperature of recombination, which depends only logarithmically on $m_{\rm P}$ and η. We ignore this logarithmic dependence and take $T_{\rm rec} \approx 3000$ K in every universe that we consider. Using that at any time after equality $\rho_* \propto \rho_{\rm rec} \propto T^3$, we find for this constraint,

$$\hat{\rho}_{\rm eq}^{-3/4}\,\hat{\zeta}_{\rm eq}^{-3} s^{-3} \gtrsim 8 \times 10^{-8}\,. \tag{5.30}$$

Since with a smaller value of $m_{\rm P}$ the matter energy density at equality and the amplitude of density perturbations are both larger, this is a constraint against decreasing $m_{\rm P}$. In terms of $\hat{m}_{\rm P}$ and μ, this gives

$$\left(\frac{1}{6}\hat{m}_{\rm P}^{-\beta} + \frac{5}{6}\hat{m}_{\rm P}^{-1}\right)^{-3} \hat{m}_{\rm P}^{3\alpha} s(\mu)^{-3} \gtrsim 8 \times 10^{-8}\,. \tag{5.31}$$

The curves that saturate this constraint for various α and β are displayed in Figs. 5.1–5.4 with the label "$t_{\rm rec} = e^{-2/3} t_{\rm vir}$."

Finally, we require that dark matter dominate over baryonic matter so that dark matter potential wells are deep enough to condense baryon over-densities after recombination. This simply translates into the constraint $f_b \lesssim 1/2$, which gives

$$\begin{aligned} \hat{m}_{\rm P} &\lesssim 5^{1/(1-\beta)} \quad \text{for } \beta < 1\,, \\ \hat{m}_{\rm P} &\gtrsim 5^{1/(1-\beta)} \quad \text{for } \beta > 1\,. \end{aligned} \tag{5.32}$$

For $\beta = 1$ this argument provides no constraint on $m_{\rm P}$ since in that case f_b is independent $m_{\rm P}$.

5.2.5.2 Galaxy Formation

Although the dark matter within a halo cannot dissipate its kinetic energy to further collapse, the baryons may do so via electromagnetic interactions. If the cooling timescale $\tau_{\rm cool}$ is less than the timescale of gravitational dynamics $\tau_{\rm grav}$, then not only do the baryons collapse, but perturbations in the baryon density fragment into smaller structures. These structures ultimately fall into a rotationally supported disk. Perturbations may further fragment if the disk satisfies the Jeans instability criteria, which is that $\tau_{\rm grav}$ be less than the time it takes for a pressure wave to traverse the perturbation. Fragmentation continues until perturbations become Jeans-stable and over-densities relax adiabatically into hot balls of gas. This appears to be the path by which halos within our universe ultimately condense into galaxies of stars (for background see for example Ref. [162]).

In order to ensure galactic dynamics similar to those within our universe, one might therefore first impose that for typical halos, $\tau_{\rm cool} \lesssim \tau_{\rm grav}$.[2] We take the dynamical timescale of the halo to be the time it would take for a test particle to free-fall to the center of the halo. For a spherical halo of constant density, this is

$$\tau_{\rm grav} \approx \sqrt{\tfrac{3}{2}\pi^2}\, m_{\rm P} \rho_{\rm vir}^{-1/2} \,. \tag{5.33}$$

For the cooling timescale we use the total thermal energy divided by the rate of energy loss, per unit volume:

$$\tau_{\rm cool} \approx \frac{3}{2} \frac{f_b\, \rho_{\rm vir}}{m_N \mu_b} \frac{T_{\rm vir}}{\Lambda_c} \,, \tag{5.34}$$

where μ_b is the mean molecular weight of the baryons in the halo (in units of the nucleon mass m_N), $T_{\rm vir}$ is the mean temperature of the halo, and Λ_c is the rate of energy loss per unit volume. The quantity $f_b\, \rho_{\rm vir}/m_N \mu_b$ is the baryon number density, including electrons.

The mean molecular weight depends on the ionization fraction and the hydrogen mass fraction of the halo. For example, for a fully ionized halo we have,

$$\mu_b \approx \frac{n_{\rm H} + 4 n_{\rm He}}{n_e + n_{\rm H} + n_{\rm He}} \approx \frac{4}{3 + 5X} \,, \tag{5.35}$$

where the subscripts denote electrons, hydrogen, or helium. Note that μ_b never strays more than a factor of two from unity. To estimate the temperature of the halo, we first note that in a virialized halo the mean kinetic energy equals half the mean gravitational binding energy. Thus for a halo of mass M we write $M v_{\rm vir}^2 \approx \frac{3}{40\pi} m_{\rm P}^{-2} M^2 R^{-1}$, where $v_{\rm vir}$ is a characteristic velocity for virialized particles and R is the radius of the halo. Since $M \approx \frac{4\pi}{3} R^3 \rho_{\rm vir}$ and since $T_{\rm vir} \approx \frac{1}{3} \mu_b m_N v_{\rm vir}^2$, we obtain

$$T_{\rm vir} \approx \frac{1}{78} \mu_b m_N m_{\rm P}^{-2} M^{2/3} \rho_{\rm vir}^{1/3} \,. \tag{5.36}$$

Note that $T_{\rm vir} \propto \sigma_{\rm eq}$ so that $T_{\rm vir}$ is a stochastic variable for halos of a given mass.

The baryons within a halo may cool via Compton scattering, bremsstrahlung, the excitation of hydrogen or helium lines, in addition to other mechanisms. These all contribute to the rate of thermal energy dissipation Λ_c. Thus Λ_c is a complicated function of temperature, which also

---

[2] We note that more careful considerations involving galactic dynamics may suggest a far weaker constraint than the one we pursue. We require $\tau_{\rm cool} \lesssim \tau_{\rm grav}$, where each timescale is evaluated at virialization. However, the baryons within a halo will cool even if this condition is not satisfied. As described in Ref. [163], this cooling pushes the gas of baryons along a curve in the temperature-density phase space that eventually leads to the condition $\tau_{\rm cool} \lesssim \tau_{\rm grav}$ being satisfied, albeit at a much later time. It is then necessary to consider any other factors that might constraint the timescale $\tau_{\rm cool}$. Ref. [117] has pointed out that if $\tau_{\rm cool}$ is not much smaller than the Hubble time, then baryons do not cool significantly before being reheated by halo mergers. Eventually halo merging ceases due to cosmological constant domination. However, even then, one must worry about too large a faction of baryons evaporating from the halo before they cool sufficiently to sink deeper into the gravitational potential well [117]. These considerations are beyond the scope of this work.

depends on the halo composition and therefore the hydrogen and helium fractions X and Y. Rather than attempt an estimate of Λ_c, we use the cooling rates given in Refs. [164, 165]. These include the processes listed above, but we neglect the possibility for molecular cooling, which is insignificant at the temperatures we consider. The galactic cooling constraint $\tau_{\text{cool}} \lesssim \tau_{\text{grav}}$ is now,

$$m_{\text{P}}^3\, \Lambda_c\, f_b^{-1} M^{-2/3} \rho_{\text{vir}}^{-11/6} \gtrsim 5 \times 10^{-3}\,. \tag{5.37}$$

The curves that saturate this inequality are displayed in Figs. 5.1–5.4 with the label "$\tau_{\text{cool}} = \tau_{\text{grav}}$."

As noted in Ref. [117], if a galaxy contains too little mass then early supernovae may blow away a significant fraction of its baryons when they explode. We expect the effects of a supernova to be relatively localized if the gravitational binding energy of the galaxy by far exceeds the energy released in the supernova [117]. This can be ensured by requiring that the energy released in the supernova be less than the halo binding energy,

$$E_{\text{bind}} \approx \frac{3}{40\pi m_{\text{P}}^2} \frac{M^2}{R} \approx \frac{1}{26} m_{\text{P}}^{-2} M^{5/3} \rho_{\text{vir}}^{1/3}\,. \tag{5.38}$$

Note that the baryons within a galaxy are much more tightly bound than when in the original halo (see for example the galactic disk estimates of Sections 5.2.5.3 and 5.2.7). Therefore Eq. (5.38) is a significant underestimate of the binding energy of a galaxy.

We expect the energy released in a supernova to scale roughly as the binding energy of a Chandrasekhar mass at its Schwarzschild radius [117], or as the binding energy of a typical star, both of which are proportional to m_{P}^3 (see Section 5.2.6). Thus we write this energy $E_{\text{sn}}\hat{m}_{\text{P}}^3$, where $E_{\text{sn}} \approx 10^{51}$ erg is the typical supernova energy within our universe. Requiring that $E_{\text{bind}} \gtrsim E_{\text{sn}}$ gives,

$$\hat{m}_{\text{P}}^{-5} \hat{\rho}_{\text{eq}}^{1/3} \hat{\zeta}_{\text{eq}}\, \mu^{5/3} s \gtrsim 4 \times 10^{-9}\,. \tag{5.39}$$

Since halos of a given mass become more weakly gravitationally bound as m_{P} is increased, Eq. (5.39) is a constraint against increasing m_{P}. After inserting previous results this becomes,

$$\left(\frac{1}{6}\hat{m}_{\text{P}}^{-\beta} + \frac{5}{6}\hat{m}_{\text{P}}^{-1}\right)^{4/3} \hat{m}_{\text{P}}^{-5-\alpha} \mu^{5/3} s(\mu) \gtrsim 4 \times 10^{-9}\,. \tag{5.40}$$

The curves that saturate Eq. (5.40) are displayed in Figs. 5.1–5.4 under the label "$E_{\text{bind}} = E_{\text{sn}}$."

5.2.5.3 Star Formation

If the above conditions are met, the baryons in a halo will radiate away energy and settle into a disk supported by its angular momentum. We then require that the disk fragment so that ultimately stars may form [117]. The stability of galactic disks against both radial and vertical perturbations can be studied using a standard Jeans analysis, which compares the dynamical timescale $\tau_{\rm grav}$ to the time it takes a pressure wave to traverse the perturbation. It turns out that the stability criteria for the two modes differ by only an order unity coefficient [117, 162, 166]. In Ref. [166] it is shown that for perturbations in the vertical direction, perturbations are unstable when the total mass of the disk satisfies,

$$M_{\rm disk} \gtrsim 120\, m_{\rm P}^2 v_p v_c R_{\rm disk}\,, \tag{5.41}$$

where v_p is the typical peculiar velocity of particles in the disk, v_c is the circular velocity of these particles, and $R_{\rm disk}$ is the disk radius.

The mass of the disk is simply the mass of the baryons in the halo, $M_{\rm disk} = f_b M$. Meanwhile, the peculiar velocity is related to the temperature of the disk. This temperature will be the lowest temperature to which the baryons can cool as they collapse, which is roughly set by the hydrogen line temperature $T_{\rm H} \approx 10^4$ K.[3] Therefore the typical peculiar velocity may be written,

$$v_p \approx \sqrt{\frac{3T_{\rm H}}{\mu_b m_N}}\,. \tag{5.42}$$

The circular velocity v_c is deduced by conserving angular momentum as the baryonic halo collapses. On the one hand, the disk angular momentum can be roughly written as $R_{\rm disk} M_{\rm disk} v_c$. On the other hand, the baryons in a halo start with angular momentum $\frac{1}{\sqrt{8\pi}} f_b \lambda m_{\rm P}^{-1} M^{3/2} R^{1/2}$, where λ is the dimensionless spin parameter [167, 168, 169, 170],

$$\lambda \equiv \sqrt{8\pi}\frac{J E_{\rm bind}}{m_{\rm P} M^{5/2}}\,, \tag{5.43}$$

where J is the magnitude of the angular momentum, $E_{\rm bind}$ is the gravitational binding energy, and all of the above quantities are evaluated for the original halo. Before the gravitational collapse of the baryons out of the dark matter halo, it is reasonable to assume that the angular momentum of the baryons and dark matter are equally distributed according to mass, such that initially $J_b = M_b J_h / M_h$ where the subscripts b and h refer to baryonic and dark matter halo quantities [166]. Then assuming that angular momentum is conserved as the baryons in the halo collapse, we find

$$v_c \sim \frac{\lambda}{\sqrt{8\pi}}\, m_{\rm P}^{-1} M^{1/2} R^{1/2} R_{\rm disk}^{-1}\,. \tag{5.44}$$

---

[3] Although the rate of galactic cooling is reduced below the temperature of hydrogen line freeze-out, cooling still proceeds via molecular transitions in, for example, H_2. Therefore $T_{\rm min}$ may become very small if one is willing to wait a long time before disk fragmentation. We follow Ref. [117] and study when galactic dynamics are similar to those within our universe.

We do not require to solve for $R_{\rm disk}$ after this expression for v_c is inserted into Eq. (5.41).

Eq. (5.41) can now be written in the simple form [117],

$$f_b \gtrsim 4\lambda \left(\frac{T_{\rm H}}{T_{\rm vir}}\right)^{1/2} . \tag{5.45}$$

The spin parameter λ is different for different halos, but it is roughly independent of $m_{\rm P}$, μ and the amplitude of density perturbations, and it typically lies near $\lambda \approx 0.05$ [167, 168, 169, 170]. Substituting into Eq. (5.45) gives

$$\hat{\mu}_b^{1/2} \hat{f}_b \hat{m}_{\rm P}^{-1} \hat{\rho}_{\rm eq}^{1/6} \hat{\zeta}_{\rm eq}^{1/2} \mu^{1/3} s^{1/2} \gtrsim 0.2\,. \tag{5.46}$$

This results in a constraint against increasing $m_{\rm P}$. We use previous results to convert this into a constraint on $\hat{m}_{\rm P}$ and μ, which gives

$$\left(\frac{1}{6}\hat{m}_{\rm P}^{-\beta} + \frac{5}{6}\hat{m}_{\rm P}^{-1}\right)^{-1/3} \hat{m}_{\rm P}^{-1-\beta-\alpha/2} \mu^{1/3} s^{1/2} \gtrsim 0.2\,, \tag{5.47}$$

where $\hat{\mu}_b(\hat{m}_{\rm P})$ depends relatively weakly on $m_{\rm P}$ and has been ignored. The curves that saturate this constraint for various α and β are displayed in Figs. 5.1–5.4 with the label "disk inst.."

It is essential that at some point fragmentation ceases so that over-densities can smoothly collapse into a star. The process of fragmentation may be seen to terminate when individual fragments become sufficiently opaque so as to trap most of their radiation [171]. In the Jeans picture, this allows the temperature of a perturbation to rise and correspondingly increase the sound speed and thus prevent further fragmentation. As described in Ref. [171], the mass scale at which this occurs is relatively independent the dominant contributions to the cooling rate and opacity and gives a typical stellar mass that scales like $m_{\rm P}^3$. Interestingly, this is the same scaling behavior that restricts the sizes of stars based on their internal temperature being high enough to fuse hydrogen and their radiation pressure being low enough so as to not blow the star apart. We elaborate on this in the next section.

5.2.6 Stellar Dynamics

We have so far ensured that the fragmentation of over-densities persists on all scales greater than a relatively small scale that is roughly proportional to $m_{\rm P}^3$. We now require that the temperature within some of the remaining structures is sufficient to fuse hydrogen to form a star. It is possible that the mere existence of stars is not a sufficient condition for the existence of observers. Therefore we also consider the requirement that some of these stars supernova in order to generate heavy elements. In addition, we consider the requirement that some of these stars have both a surface temperature within a factor of two that of the sun and a main-sequence lifetime of at least a few billion years. Our motivation for selecting these specific criteria is simple. Without knowing what are the necessary conditions for observers to arise within a stellar system, we study what seem at least to be two sufficient conditions.

Ref. [118] has studied the basic requirements that constrain the properties of stars. For decreasing stellar mass, the central temperature must be above some minimum temperature $T_{\rm nuc}$ that is necessary to fuse hydrogen. The central temperature within a low-mass star is estimated by balancing the influences of gravitational pressure and electron degeneracy. This gives for the central temperature of a low-mass star [118],

$$T_c \propto m_{\rm P}^{-4} M^{4/3}\,, \tag{5.48}$$

where M is the stellar mass. The least massive stars have $T_c = T_{\rm nuc}$. Since $T_{\rm nuc}$ is independent of $m_{\rm P}$, in any universe these stars have mass

$$M_{\rm min} \propto m_{\rm P}^3\,. \tag{5.49}$$

On the other hand, there is also an upper limit to the mass of a star. If the radiation pressure within a star well exceeds the gravitational pressure that its mass can provide, then the star itself becomes unstable upon the ignition of its core. This constrains the maximum mass that a star may have, such that it satisfies the scaling [118],

$$M_{\rm max} \propto m_{\rm P}^3\,. \tag{5.50}$$

Note that the minimum and maximum mass of a star both scale as $m_{\rm P}^3$. This is also the scaling of the typical mass that becomes sufficiently opaque to prevent further fragmentation. This means that, given a window of masses for which stars exist in our universe, there will also be such a window within universes with significantly different values of $m_{\rm P}$.

5.2.6.1 Stellar Lifetimes and Spectra

We now seek constraints to ensure that some of the stars produced within a particular universe have surface temperatures and main-sequence lifetimes that appear to be sufficient for the evolution of observers. Our purpose in investigating this condition is to ensure that we do not overlook what might be viewed as an important anthropic constraint. Therefore we adopt a very restrictive perspective and require that some stars have surface temperatures of at least 3500 K and that these stars have main-sequence lifetimes greater than the timescale of biological evolution, $\tau_{\rm evol} \approx 5 \times 10^9$ yrs. This surface temperature is chosen in part to simplify the calculation of stellar lifetimes and in part because a black body at this temperature radiates a significant fraction of its power into the frequency band accessible to chemistry. The evolutionary timescale τ_{evol} should be understood to include the time required for a planet to condense and cool, minus the time it takes the star to reach main-sequence hydrogen burning. This time may be different for different planets, but we do not expect it to form the dominant contribution to τ_{evol}.

The main-sequence lifetime of a star is roughly equal to the available energy of the star divided by the typical rate that it radiates energy away,

$$\tau_\star \propto XML^{-1}\,. \tag{5.51}$$

Here L is the typical luminosity of the star during main-sequence and X is the hydrogen fraction. We assume that differences in composition other than differences in the hydrogen fraction have benign consequences. In addition, we neglect the m_{P} dependence of the mean molecular weight $\mu_\star$, since $\mu_\star$ changes by only roughly a factor of two as X ranges from zero to unity. For an introduction to the major concepts of stellar astrophysics, we have found useful Refs. [172, 173, 174, 175].

Our narrow purpose allows for a simplified analysis of the necessary stellar dynamics. Since we specify stars by their surface temperature, we write the luminosity $L \propto R^2 T_s^4$ for stellar radius R and surface temperature T_s. To eliminate R, we note that the central temperature of a star scales as

$$T_c \propto \frac{1}{m_{\text{P}}^2}\frac{M}{R}\,. \tag{5.52}$$

Thus we can write,

$$\tau_\star \propto \frac{X m_{\text{P}}^4 T_c^2}{M T_s^4}\,. \tag{5.53}$$

The lifetime of a star is maximized by considering the minimum allowed surface temperature, in this case $T_s = 3500$ K. Since both the lower end and the upper end of the window for stellar masses scale as m_{P}^3, as a basic approximation we may take this to be the scaling for all stellar masses at fixed T_s and T_c. Combining this with Eqs. (5.51), (5.52), and (5.53) gives,

$$\hat{\tau}_\star \approx \hat{X}\hat{m}_{\text{P}}\,, \tag{5.54}$$

where $\hat{\tau}_\star$ is measured in units of the main-sequence lifetime of these stars within our universe. This lifetime is roughly 100 billion years [176]. Therefore the constraint $\tau_\star \gtrsim \tau_{\text{evol}}$ becomes

$$\hat{X}\hat{m}_{\text{P}} \gtrsim 5 \times 10^{-2}\,. \tag{5.55}$$

This requires that $\hat{m}_\mathrm{P}$ satisfy $\hat{m}_\mathrm{P} \gtrsim 0.1$.

We now consider this analysis in a little more detail. In particular, we consider the effects of convection and electron degeneracy explicitly in order to motivate that we can keep T_s/T_c fixed while scaling m_P and that the stellar mass scales like m_P^3 at fixed T_c. An ionized star in which radiation pressure can be neglected and in which the energy transport is everywhere dominated by convection is well approximated as a polytrope with polytropic index $i = 3/2$ [172, 174, 175]. This implies certain scalings between stellar properties and in particular that for these stars T_s/T_c is independent of m_P, M and R. Within our universe, stars with surface temperatures at and below 3500 K have masses $M \lesssim 0.35 M_\odot$ and are well approximated by these polytropes. In addition, it can be shown that stars defined by these temperatures remain convection dominated as m_P is decreased [174, 175]. To see this intuitively, note that convection is driven by tidal forces and, all else being equal, one expects the tidal forces within a star to increase as m_P is decreased. Therefore, we expect T_s/T_c to be fixed for stars with $T_s = 3500$ K as m_P is decreased.

Meanwhile, so long as electron degeneracy is significant within the center of the star, the mass required to achieve a fixed central temperature scales like m_P^3 [118]. Stars with a surface temperature of 3500 K are indeed partially degenerate within our universe, but they could become non-degenerate after some amount of scaling $M \propto m_\mathrm{P}^3$. To see that there can exist stars for which the degeneracy remains fixed, consider again the polytrope model. The electron degeneracy at the center of a star is a function of the ratio $n_e T_c^{-3/2}$, where n_e is the number density of electrons. The scaling relations applicable to an $i = 3/2$ polytrope imply that the electron density n_e scales like the average density of the star, which at constant central temperature scales like $m_\mathrm{P}^6 M^{-2}$. Next note that the electron degeneracy at the center of a star is a constant if the mass of the star scales as $M \propto m_\mathrm{P}^3$ for fixed central temperature T_c. This is precisely the scaling that describes a partially degenerate star, which means that a partially degenerate star remains partially degenerate as m_P is scaled while keeping T_c fixed. Therefore we expect stars with a surface temperature of 3500 K to remain partially degenerate as we decrease m_P keeping T_s fixed, so that indeed $M \propto m_\mathrm{P}^3$. These arguments justify the constraint Eq. (5.55).

It is illuminating to consider a different form of analysis. This applies to ionized stars where radiation pressure can be neglected but radiation dominates over convection in the transport of energy. Then the scaling of the stellar mass M with m_P for fixed T_c may be obtained for a class of stars (so-called homologous stars[4]) by applying homologous transformations to the equations of hydrodynamical equilibrium. To perform such an analysis, we must phenomenologically model the opacity of the star and the rate of energy generation per unit mass with the respective formulae [172],

$$\kappa \approx \kappa_0 \rho^n T^{-s}\,, \quad \epsilon \approx \epsilon_0 X^2 \rho T^\nu\,. \tag{5.56}$$

---

[4] This is a very restrictive class, since by definition the mass distribution for two homologous stars of mass M_i and radius R_i must satisfy $m_1(r)/M_1 = m_2(rR_2/R_1)/M_2$, where $m_i(r)$ is the mass contained within a sphere of radius r.

The terms κ_0 and ϵ_0 are constants while the exponents n, s, and ν depend on the physical properties of the star. The Kramers opacity of intermediate mass stars such as our sun is modeled using $n = 1$ and $s = 7/2$. Then it can be shown that, independent of ν, the stellar mass scales like $M \propto X^{1/3} m_{\rm P}^{10/3}$ for fixed central temperature [172]. For such stars T_s/T_c is not constant, so one cannot use Eq. (5.53). However, at fixed central temperature the homology transformations determine the scaling of the luminosity to be $L \propto X^2 m_{\rm P}^6 M^{-1} \propto X^{5/3} m_{\rm P}^{8/3}$. Inserting these scalings into Eq. (5.51) gives

$$\hat{\tau}'_\star \approx \hat{X}^{-1/3} m_{\rm P}^{2/3}\,, \tag{5.57}$$

where $\hat{\tau}'_\star$ is the main-sequence lifetime in units of an appropriate lifetime evaluated within our universe.

This gives a slightly weaker dependence on $m_{\rm P}$ than Eq. (5.55); however stars that are well-described by these approximations have shorter lifetimes than those described by an $i = 3/2$ polytrope. In addition, the physical characteristics that make these approximations applicable will not continue to describe stars as we decrease $m_{\rm P}$ with T_s fixed. (They do continue to describe stars as $m_{\rm P}$ is increased for fixed T_s, since this tends to lessen the importance of convection.) Nevertheless, this confirms the qualitative form of Eq. (5.55) and suggests that an analogous analysis would apply if surface temperatures closer to that of the sun were demanded.

5.2.6.2 Heavy Element Production

Supernovae are believed to be the exclusive source of heavy elements within our universe. However, the dynamics of supernovae are very complex and are still not fully understood (for reviews see for example Refs. [177, 178]). Therefore ensuring the existence of supernovae in universes with differing values of $m_{\rm P}$ is clearly speculative. In this section we simply provide some qualitative remarks in support of this possibility. There are many types of supernovae within our universe; for convenience we focus on what are called type Ia supernovae.

Type Ia supernovae are understood to erupt via the accretion of matter by a white dwarf star. Meanwhile, white dwarfs are created when a star has consumed all of its hydrogen and helium fuel but does not possess sufficient mass either to drive its central temperature high enough to ignite carbon fusion or to form a black hole. According to the scaling relationships discussed in the previous section, the first condition is always satisfied given that it is satisfied within our universe. On the other hand, the Schwarzschild radius and the physical radius of a star at fixed central temperature both scale as $R \propto m_{\rm P}^{-2} M$. Therefore white dwarfs will exist in all of the universes that we consider.

The supernova of an accreting white dwarf proceeds when its growing mass reaches the Chandrasekhar limit and the star becomes unstable through the nuclear ignition of its carbon. The relevant physical scales for this phenomena are set by the Chandrasekhar mass and the binding energy of a Chandrasekhar mass at about a Schwarzschild radius. Since these and the typical stellar mass all scale as $m_{\rm P}^3$, it seems plausible that type Ia supernovae would occur within universes with significantly differing values of $m_{\rm P}$. This ensures the production of heavy elements within these universes.

There is a possible caveat to this result. Within the context of another anthropic analysis, it has been remarked [120] that the relatively low production of oxygen by type Ia supernovae may significantly hinder the formation of life if oxygen is not generated elsewhere. However, it is not clear that this suppression, roughly 3% to 8% relative type II supernova [179, 180], is sufficient to render life overwhelmingly unlikely. The arguments of Ref. [120] were aimed against a scenario where type II supernova would definitely not occur. Since these supernova occur for a wide range of stellar masses within our universe, it is plausible that universes with $m_{\rm P}$ not too unlike ours will also contain type II supernovae. It is beyond the scope of this work to investigate more precisely for what values of $m_{\rm P}$ these supernovae will occur.

5.2.7 Stability of Stellar Systems

As is illustrated in Refs. [117, 121], an important anthropic constraint derives from requiring that stellar systems are stable against cosmic disruptions. Specifically, if a second star grazes too close to an existing stellar system, then a habitable planet may be thrown out of its anthropically fortuitous orbit. Here we seek a constraint to ensure that such encounters are typically too infrequent to interfere with the evolution of life. First, we define a destructive encounter rate,

$$\gamma \sim n_\star \sigma_\star v_p \,, \tag{5.58}$$

where $n_\star$ is the number density of stars, $\sigma_\star = \pi b^2$ is the cross section for an encounter with "fatal" impact parameter b, and v_p is the typical peculiar velocity of a star. Note that all of the stars within a given neighborhood have the same circular velocity; thus the circular velocity does not contribute to the encounter rate.

The typical peculiar velocity of the stars in a galaxy is approximately determined by the temperature of the constitutive baryons during the phase of star formation. Since the baryons in a galaxy quickly cool to about $T_{\rm H} \approx 10^4$ K and cool relatively slowly thereafter, we take this to be the relevant temperature. The corresponding peculiar velocity is then given by Eq. (5.42). For $\hat{m}_{\rm P} = 1$ this gives a typical peculiar velocity of about $v_p \simeq 20$ km/s which agrees well with observation. The number density of stars $n_\star$ is equal to the number density of baryons divided by the typical number of baryons within a star, $N_\star$. Recall from Section 5.2.5 that the number density of baryons within a galactic halo is equal to $f_b\rho_{\rm vir}/m_N\mu_b$. Therefore $n_\star$ can be written,

$$n_\star = \frac{f_\star f_b\,\rho_{\rm vir}}{m_N\mu_b N_\star}\,, \tag{5.59}$$

where $f_\star$ is a fudge factor inserted to account for the increased density of the galactic disk, for the fraction of baryons that do not end up within stars, and for any clustering that may be involved in the star formation process. A typical star within our universe contains 10^{57} baryons; therefore the results of Section 5.2.6 suggest $N_\star \sim 10^{57}\hat{m}_{\rm P}^3/\hat{\mu}_\star \sim 10^{57}\hat{m}_{\rm P}^3/\hat{\mu}_b$. To estimate $f_\star$ is somewhat more challenging.

The factor $f_\star$ accounts for several effects. For example, stars may form in clusters such that most stars exist in a neighborhood of higher density than the average density of stars in a galaxy. On the other hand, a significant fraction of baryons may compose a relatively diffuse interstellar gas and therefore not contribute to the stellar encounter rate. As it is beyond the scope of this work to compute the $m_{\rm P}$ dependence of these effects, we simply treat them as being independent of $m_{\rm P}$. Meanwhile, we also expect $f_\star$ to be proportional to the relative density of the galactic disk to that of the baryons in the halo. Rather than concern ourselves with the specific geometry of the galactic disk, we study a simple model to obtain the $m_{\rm P}$ dependence. We expect the factor $f_\star$ to roughly scale like

$$f_\star \propto \frac{R^3}{H_{\rm disk}R_{\rm disk}^2}\,, \tag{5.60}$$

where $H_{\rm disk}$ is the typical disk thickness. This can be written [166],

$$H_{\rm disk} \propto m_{\rm P}^2 M_{\rm disk}^{-1} R_{\rm disk}^2 v_p^2\,. \tag{5.61}$$

Meanwhile, to solve for $R_{\rm disk}$ we note that the circular velocity for stars is given by both by Eq. (5.44) and by,

$$v_c \propto m_{\rm P}^{-1} M_{\rm disk}^{1/2} R_{\rm disk}^{-1/2}\,. \tag{5.62}$$

Equating these expressions gives $R_{\rm disk} \propto \lambda^2 f_b^{-1} R$. Finally, putting all of this together gives,

$$f_\star \propto \frac{f_b^5}{\lambda^8}\frac{v_{\rm vir}^2}{v_p^2}\,. \tag{5.63}$$

Note the strong dependence of $f_\star$ on the spin parameter λ. The spin parameter is a stochastic variable with statistical properties related to those of $\zeta_{\rm eq}$. For example, the Milky Way appears to be characterized by $\lambda \approx 0.06$ [181] while typical galaxies may have λ a factor of two larger or smaller than this [167, 168, 169, 170]. This and other factors suggest that the factor $f_\star$ may vary widely among galactic environments within any particular universe. In addition, as explained above we have ignored several effects that might enter into $f_\star$. For concreteness we normalize $f_\star$ to the value that describes our solar environment in the Milky Way, $f_\star \sim 10^5$ [117]. This gives

$$f_\star \sim 10^5 \hat{m}_{\mathrm{P}}^{-2} \hat{\mu}_b \hat{f}_b^5 \mu^{2/3} \hat{\rho}_{\mathrm{eq}}^{1/3} \hat{\zeta}_{\mathrm{eq}} s(\mu)\,, \tag{5.64}$$

where the dependence of $f_\star$ on m_{P} stems entirely from its dependence on f_b and T_{vir}.

It is left to calculate the impact parameter for fatal encounters. We are specifically interested in the persistence of stellar systems that contain a planet in orbit about a star such as those considered in Section 5.2.6. In addition, we focus on planetary orbits that receive electromagnetic radiation with an intensity that is comparable to that from the sun at the orbit of the earth. We then assume that an encounter will not be devastating to such a stellar system if the gravitational field from the grazing star is less than a tenth that of the primary star in the vicinity of the orbiting planet. Therefore we approximate b to be roughly $\sqrt{10}$ times the radius of orbit for a planet receiving about the same stellar intensity as the earth but in orbit about the stars studied in Section 5.2.6. Note that the luminosity of a star is $L \propto R^2 T_s^4$ while the intensity at a distance r is $I \propto L/r^2$. Therefore

$$b = \sqrt{10}\, r_{\mathrm{au}} \frac{R_\otimes T_\otimes^2}{R_\odot T_\odot^2} \frac{R_\ominus}{R_\otimes}\,, \tag{5.65}$$

where the subscript $\odot$ designates a quantity for the sun, $\otimes$ designates a star with surface temperature $T_\otimes = 3500$ K in our universe, $R_\ominus$ is the radius of a star with this surface temperature within a universe with a different value for m_{P}, and r_{au} is one astronomical unit. Eq. (5.65) has been written so that every quantity can be evaluated within our universe except for $R_\ominus/R_\otimes \approx \hat{m}_{\mathrm{P}}$, which is deduced using the results of Section 5.2.6.

We put all these results together to obtain γ. The constraint that stellar systems typically survive a dangerous close encounter for long enough that life may evolve is $\tau_{\mathrm{evol}} \lesssim \gamma^{-1}$, where again $\tau_{\mathrm{evol}} = 5 \times 10^9$ yrs. In terms of cosmological parameters this is,

$$\hat{m}_{\mathrm{P}}^3 \hat{\mu}_b^{-1/2} \hat{f}_b^{-6} \hat{\rho}_{\mathrm{eq}}^{-4/3} \hat{\zeta}_{\mathrm{eq}}^{-4} \mu^{-2/3} s^{-4} \gtrsim 10^{-5}\,, \tag{5.66}$$

where we have used the models of Ref. [176] to substitute $R_\otimes T_\otimes^2 / R_\odot T_\odot^2 \approx \sqrt{L_\otimes / L_\odot} \approx 0.14$. Decreasing m_{P} reduces the cross section for dangerous impacts, since the "anthropically-favorable" radius decreases, yet increases the number density of stars. The net effect is a constraint against decreasing m_{P}. The explicit constraint implied for m_{P} is given by

$$\left(\frac{1}{6}\hat{m}_{\mathrm{P}}^{-\beta} + \frac{5}{6}\hat{m}_{\mathrm{P}}^{-1}\right)^{2/3} \hat{m}_{\mathrm{P}}^{3+4\alpha+6\beta} \mu^{-2/3} s^{-4} \gtrsim 10^{-5}\,, \tag{5.67}$$

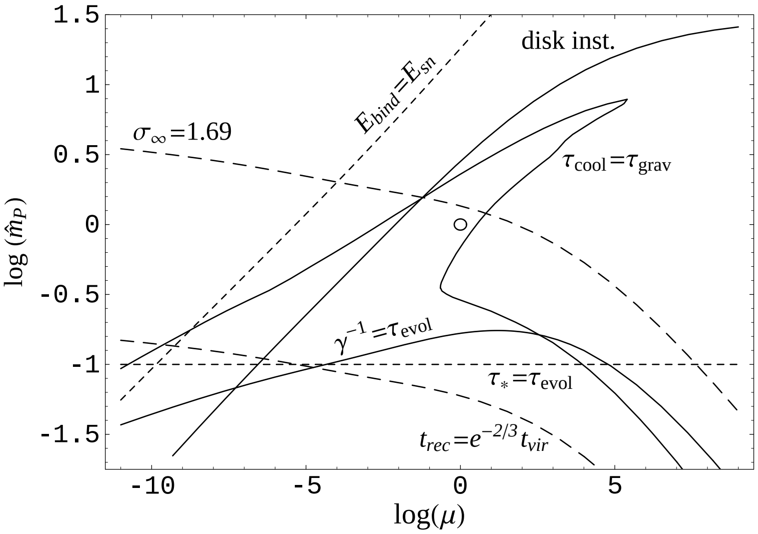


Figure 5.1: Anthropic constraints on $\hat{m}_{\mathrm{P}}$, plot as a function of halo mass scale μ, for $\alpha = 1$ and $\beta = 0$. The logarithms are base ten and the empty circle corresponds to the mass scale of our galaxy with m_{P} as observed within our universe. The region within the $\mu - \hat{m}_{\mathrm{P}}$ plane that is excluded by any constraint is the region that does not include this circle. The parameters α and β as well as the labels on the curves are defined in the text (see for example the summary of Section 5.2.8). Note that in this and subsequent figures the galactic cooling constraint does not include the effects of molecular cooling.

where $\hat{\mu}_b$ depends relatively weakly on m_{P} and has been ignored. The curves that saturate this inequality are displayed in Figs. 5.1–5.4 using the label "$\gamma^{-1} = \tau_{\mathrm{evol}}$."

5.2.8 Summary

Let us now summarize the results of the previous sections. Many anthropic constraints depend on the primordial curvature perturbation ζ_{eq} and on the baryon to photon ratio η. Lacking any standard model for the generation of either of these, we write them generically as $\zeta_{\mathrm{eq}} \approx \hat{m}_{\mathrm{P}}^{-\alpha}$ and $\eta \approx \hat{m}_{\mathrm{P}}^{-\beta}$, where $\hat{m}_{\mathrm{P}}$ is the ratio between the apparent Planck mass and the value obtained within our universe. For the most popular models of inflation, α ranges between one ($m_\varphi^2\varphi^2$ chaotic inflation) and three (hybrid and natural inflation). Meanwhile, popular models of baryogenesis give β between zero (efficient leptogenesis and electroweak baryogenesis) and 3/2 (specific models of SUSY Affleck-Dine baryogenesis).

Most of the anthropic constraints under consideration are displayed in Figs. 5.1–5.4, for representative values of α and β. We assume a WIMP candidate to dominate the dark matter density. Many constraints depend on the total mass within the galactic halo for which they are evaluated. This mass is denoted μ and is measured in units of the Milky Way mass, or 10^{12} solar masses. Note the empty circle in each panel of Figs. 5.1–5.4. This corresponds to a mass scale equal to the mass of the Milky Way with a Planck mass equal to the value obtained within our universe. The region

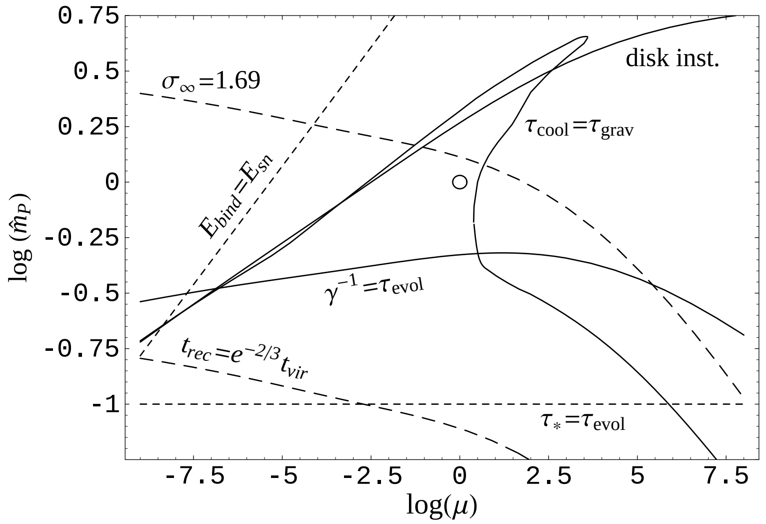


Figure 5.2: Anthropic constraints on $\hat{m}_{\rm P}$, plot as a function of halo mass scale μ, for $\alpha = 1$ and $\beta = 3/2$. For more details see the caption to Fig. 5.1.

within the μ–$\hat{m}_{\rm P}$ plane that is excluded by any constraint is the region that does not include this circle. For clarity we do not display the constraints that primordial black holes form a sub-dominant contribution to the energy density of the universe and that the dark matter dominates over baryonic matter. These are weaker constraints than those displayed in Figs. 5.1–5.4 and they are easy to calculate from Eq. (5.20) and Eqs. (5.32). Finally, we note that many of the constraints in Figs. 5.1–5.4 are deduced by assuming that other constraints are satisfied. For example, the "disk inst." curve is changed when the constraint represented by the curve "$t_{\rm rec} = e^{-2/3}t_{\rm vir}$" is not satisfied. The continuous curves in Figs. 5.1–5.4 are intended to guide the eye.

We annotate Figs. 5.1–5.4 as follows. A number of constraints come from the various levels of structure formation. The curve labeled "$\tau_{\rm cool} = \tau_{\rm grav}$" (this curve has a distinctive 'dorsal fin' shape) marks the separation between the mass scales of halos that contain baryons which cool faster than they (would) collapse and those that do not. As explained in Section 5.2.5.2, this is one among a set of sufficient, but perhaps not necessary, conditions that allow for galaxy formation. Another one of these conditions is that galactic disks be Jeans-unstable, which occurs below the curve labeled "disk inst." in Figs. 5.1–5.4. Meanwhile, structure formation requires that over-densities separate from the cosmic expansion before the domination of the cosmological constant halts their growth. This requirement is filled below the curve labeled "$\sigma_\infty = 1.69$." Finally, our analysis of structure formation assumes that galactic halos virialize after recombination, which occurs for $m_{\rm P}$ values located above the curve labeled "$t_{\rm rec} = e^{-2/3}t_{\rm vir}$." Alternative paths to structure formation are discussed in Section 5.5.

We also consider a few anthropic criteria that are not directly related to structure formation.

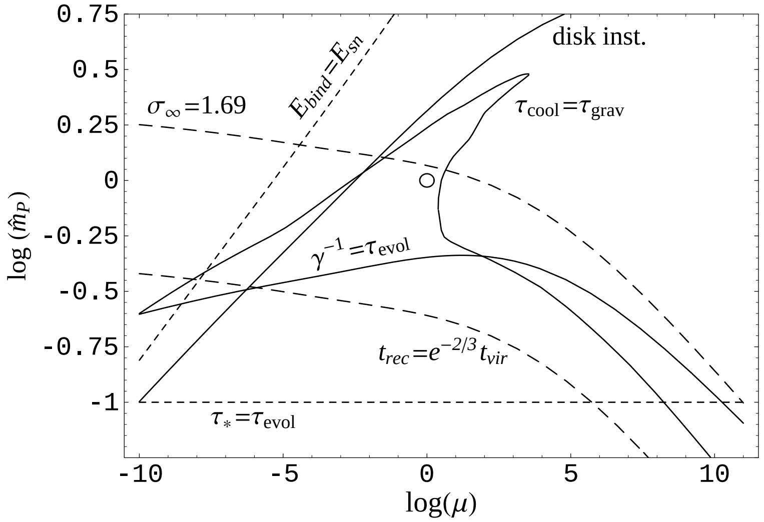


Figure 5.3: Anthropic constraints on $\hat{m}_{\rm P}$, plot as a function of halo mass scale μ, for $\alpha = 3$ and $\beta = 0$. For more details see the caption to Fig. 5.1.

For example, another constraint that we consider is that galaxies not be so small that they are blown apart by internal supernovae. This will not happen if the binding energy of a galactic halo well exceeds the energy released via supernovae. This condition is satisfied for μ and $\hat{m}_{\rm P}$ to the right of the curve labeled "$E_{\rm bind} = E_{\rm sn}$." In addition, one might require that collisions between stellar systems be such that impact parameters so small as to dislodge a habitable planet occur on a timescale that is much larger than the evolutionary timescale, here taken to be $\tau_{\rm evol} \approx 5 \times 10^9$ yrs. This constraint is satisfied above the curve labeled "$\gamma^{-1} = \tau_{\rm evol}$." Finally, one might wish to restrict attention to universes that contain stars that have surface temperatures in excess of about 3500 K and that have main-sequence lifetimes in excess of about four billion years. These correspond to positions above the line labeled "$\tau_\star = \tau_{\rm evol}$" in Figs. 5.1–5.4.

Except for the stellar lifetime constraint, every constraint displayed in Figs. 5.1–5.4 depends on the size of the initial over-density that eventually grows into a galaxy. The curves in Figs. 5.1–5.4 correspond to choosing this initial fluctuation to be the rms of the density perturbations at a scale μ evaluated at matter-radiation equality. However, the initial over-density describing any galaxy is a stochastic variable that may be larger or smaller than this. Therefore all of the curves in Figs. 5.1–5.4 will be shifted when one considers galaxies that are away from the norm. In addition, the disk instability and close encounters curves (labeled "disk inst." and "$\gamma^{-1} = \tau_{\rm evol}$" respectively) depends very strongly on other stochastic quantities, such as the galactic spin parameter (see Section 5.2.7). Therefore the range of $m_{\rm P}$ that is consistent with the above constraints is larger than the windows in Figs. 5.1–5.4 would suggest if one allows observers to arise in atypical environments within any given universe.

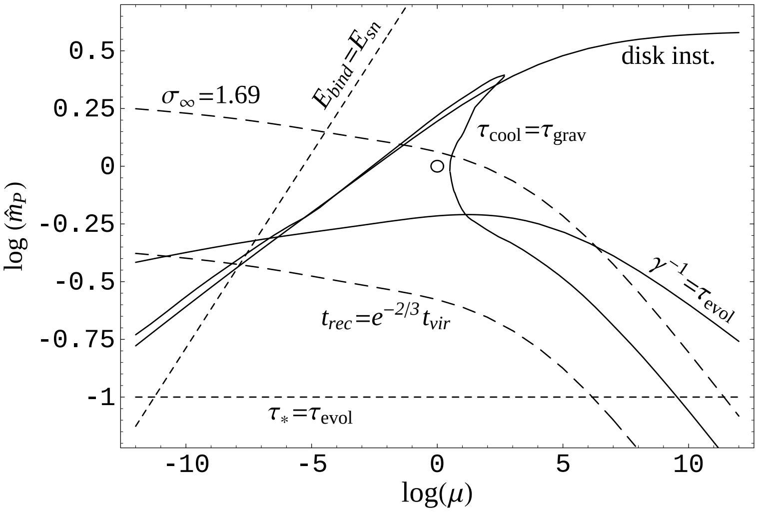


Figure 5.4: Anthropic constraints on $\hat{m}_{\rm P}$, plot as a function of halo mass scale μ, for $\alpha = 3$ and $\beta = 3/2$. For more details see the caption to Fig. 5.1.

5.3 The Probability Distribution for the Scale of Gravity

If the Planck mass $m_{\rm P}$ scans across a landscape of universes, then the value within any particular universe may not be uniquely determined. However, with an understanding of the landscape and a calculus to regulate over a conceivably infinite number of infinitely expansive universes, we may in principle calculate the distribution of $m_{\rm P}$. Since we cannot access any of the other universes within the landscape, such a distribution cannot be directly tested. Nevertheless, we may still use this distribution to calculate the likelihood that we should observe the value of $m_{\rm P}$ that we do. As we are forced to test this distribution using only our universe, we must be careful to account for any selection effects that would attenuate the distribution of $m_{\rm P}$.

These selection effects generate a factor $\mathcal{S}$ that multiplies the "prior" distribution $\mathcal{I}$. Thus we write the probability to measure the Planck mass to be $m_{\rm P}$,

$$P(m_{\rm P}) = \mathcal{S}(m_{\rm P})\mathcal{I}(m_{\rm P})\,. \tag{5.68}$$

The factor $\mathcal{I}(m_{\rm P})$ may be taken as the likelihood for universes with Planck mass $m_{\rm P}$ to arise within the multiverse, while $\mathcal{S}(m_{\rm P})$ may be understood as the likelihood for observers to arise within those universes. We restrict our prior $\mathcal{I}$ to account for only universes exactly like ours except for their value of $m_{\rm P}$. This is equivalent to restricting the selection criteria in $\mathcal{S}$. As mentioned in the introduction, there are many subtle issues that complicate the calculation of $\mathcal{S}$ and $\mathcal{I}$. Our purpose here is not to resolve any of these issues based on technical grounds. Instead we explore an empirical constraint that may complicate some proposals to address them.

One might expect the likelihood for a universe to support observers to be proportional to the total baryonic mass within galaxies in that universe. (Note that here and below we do not presume proportionality factors to be independent of m_{P}.) Meanwhile, the baryonic mass within galaxies is proportional to the total energy within a universe. This quantity diverges in proportion to the volume of the universe. Nevertheless, we may hope for a regularization scheme that allows for the volumes of universes to be compared. Since the energy density does not redshift during inflation, it is possible that when volumes are properly regulated, the ratio between the total energy densities of two universes will be proportional to the ratio of their inflationary expansion factors.

While this argument may be intuitively appealing, such a prescription for volume-based weighting presents well-known difficulties [131, 132, 133, 134, 135, 136]. For instance, its conclusion is crucially dependent on a specific global spacelike slicing, which is ambiguous outside the horizon of any one observer. For example, an observer can chose a spacelike slicing that is engineered to create a very large initial volume for the observer's own universe, while also suppressing the initial volume of the universes of casually disconnected observers. The suppression of an initial volume can be used to cancel the inflationary expansion factor, such that this slicing would give a dramatically different counting than the weighting described above.

Nevertheless, in at least one proposal this ambiguity has been overcome and the result includes a selection effect that weights universes according to their inflationary volume [135]. We assume that this result holds and write,

$$\mathcal{S}(m_{\mathrm{P}}) = \mathcal{A}(m_{\mathrm{P}})\mathcal{V}(m_{\mathrm{P}})\,. \tag{5.69}$$

Here $\mathcal{A}$ is proportional to the anthropic factor, which ultimately gives the likelihood per unit volume for some class of observer to arise within a universe. Depending on one's notion of an observer, $\mathcal{A}$ might include, for example, the baryonic mass fraction within galaxies, the fraction of stars with lifetimes in excess of few billion years, and/or other environmental specifications. The factor $\mathcal{V}$ is the inflationary expansion factor for the universe,

$$\mathcal{V}(m_{\mathrm{P}}) = e^{3N(m_{\mathrm{P}})}\,, \tag{5.70}$$

where $N(m_{\mathrm{P}})$ is the number of e-folds if inflation that typically occurs after a universe with Planck mass m_{P} has arisen within the multiverse.

If inflation is driven by a single canonical scalar field, then the number of e-folds of inflation is

$$N = \frac{1}{m_{\mathrm{P}}^2}\int_{\varphi_f}^{\varphi_i} \frac{V}{V_\varphi}\,d\varphi\,. \tag{5.71}$$

Here φ_f is the value of the inflaton when inflation ends, set by when the first slow-roll parameter equals unity, and φ_i is the value of the inflaton when inflation begins. Note that in general N depends explicitly on $m_{\rm P}$. As a specific example, consider chaotic inflation with an inflaton potential $V(\varphi) = \frac{1}{2}m_\varphi^2\varphi^2$. This gives

$$N = \frac{1}{4}\frac{\varphi_i^2}{m_{\rm P}^2} - \frac{1}{2}\,. \tag{5.72}$$

It seems evident that N will generically depend on $m_{\rm P}$. Yet without understanding the mechanism by which a universe is obtained within the multiverse, it is not clear what is the (typical) value of φ_i and what is its dependence on $m_{\rm P}$.

To illustrate that the explicit and implicit dependences of N on $m_{\rm P}$ are not expected to cancel, three models to determine φ_i are now considered. The first model sets φ_i to be the value where classical evolution of φ begins to dominate over the quantum fluctuations experienced whenever a mode exits the Hubble radius. This is set by the solution to

$$\frac{1}{\sqrt{12}\,\pi}\frac{V^{3/2}}{m_{\rm P}^3 V_\varphi} = 1\,. \tag{5.73}$$

If this is the case, the total number of e-folds is

$$N \simeq \sqrt{6}\,\pi\frac{m_{\rm P}}{m_\varphi}\,. \tag{5.74}$$

The second model assumes that φ_i is determined by where the inflaton energy density equals the Planck energy density. In this case,

$$N \simeq \frac{m_{\rm P}^2}{m_\varphi^2}\,. \tag{5.75}$$

On the other hand, if φ_i is determined by where the inflaton energy density equals M^4, then

$$N \simeq \frac{M^4}{m_\varphi^2 m_{\rm P}^2}\,. \tag{5.76}$$

Not only does N generally depend on $m_{\rm P}$, but the dependence is very strong for $m_{\rm P}$ near the value obtained within our universe. Consider for example the case of chaotic inflation with φ_i set by Eq. (5.73). Then $N \sim 10^5\,\hat{m}_{\rm P}$, where $\hat{m}_{\rm P}$ is the Planck mass in units of the value obtained within our universe. Meanwhile, if Eq. (5.75) sets the value of φ_i, then $N \sim 10^{10}\,\hat{m}_{\rm P}^2$. Clearly different choices for φ_i, and in particular different models of inflation, will in general give a different dependence of N on $m_{\rm P}$. However, the dependence is always strong. This is because our universe experienced a large number (at least about sixty) of e-folds of inflation.

The ambiguity over the $m_{\rm P}$ dependence of N is not of concern. The important result is that so long as the dependence on $m_{\rm P}$ of $\mathcal{A}$ and $\mathcal{I}$ is significantly weaker than the strong exponential dependence in $\mathcal{V}$, then we expect $m_{\rm P}$ to be most probably observed very near one of the boundaries of the anthropic range. The analysis of Section 5.2 reveals that this is not the case with at least the factor $\mathcal{A}$ and the value of $m_{\rm P}$ observed within our universe. We illustrate this with an explicit example in Section 5.6. There we show that even in the contrived case where $N \sim 60\,\hat{m}_{\rm P}$, the volume factor $\mathcal{V}(m_{\rm P})$ overwhelms what appears to be one of the tightest anthropic constraints. This pushes the expectation value for $m_{\rm P}$ well beyond what we estimated to be the anthropic boundary, while the value obtained within our universe sits far down the tail of the distribution. This is exactly analogous to the runaway "σ-problem" and the "Q catastrophe" introduced in Refs. [144, 145]. We refer to our example as the "$m_{\rm P}$-problem."

The σ-problem and Q catastrophe were motivated by the fact that in many models of inflation the total number of e-folds of inflation depends on the inflationary parameters that also set the level of density perturbations (the authors of Refs. [144, 145] use the notations $\sigma \sim Q \sim \zeta$). For example, in chaotic inflation with potential $V(\varphi) = \frac{1}{2}m_\varphi^2\varphi^2$ one finds $N \sim \zeta^{-1}$. Therefore if the inflationary parameters may scan over the landscape, by the same argument given above we expect ζ to be pushed to one of its anthropic boundaries, whereas in our universe it sits comfortably near the middle of the anthropic window [117]. It has been pointed out [182] that this argument is not completely satisfactory, since by hypothesis universes with an enormous number of e-folds are preferred. In such universes, ζ may plausibly depend on different parameters during a long stretch of early inflation than it does near the end of inflation, when scales important to the formation of structure are generated. Moreover, we note that the curvature perturbation is related to the first slow-roll parameter,

$$\zeta \sim \frac{1}{\sqrt{\epsilon_I}}\frac{V^{1/2}}{m_{\rm P}^2}\,. \tag{5.77}$$

Inflation of longer duration requires a smaller ϵ_I, yet for inflation to end at all requires that at some point ϵ_I evolve toward unity. Therefore $\zeta_{\rm eq}$ may be significantly decreased from its value during most of inflation by the necessary condition that ϵ_I interpolate between some very small value and unity by the end of inflation.

We emphasize that the $m_{\rm P}$-problem is not hampered by these issues. That is, unlike ζ, $m_{\rm P}$ is a constant within any given universe.[5] We also emphasize that if the model of inflation that describes our universe exhibits an $m_{\rm P}$-problem, then allowing more parameters to vary across the landscape cannot mitigate this problem. That is, although allowing more parameters to vary might

---

[5] Of course, the inflationary landscape hypothesis presumes that the fields ϕ described in the introduction, c.f. Eq. (5.1), will evolve within the multiverse. Indeed, this is how the landscape is populated. However, we assume that the vacua defined by the fields ϕ are selected prior to period of inflation in which we take interest, during which $m_{\rm P}$ is constant. Specifically, the factor $\mathcal{I}$ is assumed to account for any selection effects due to the field evolution prior to this period of inflation, and the terms $\mathcal{A}$ and $\mathcal{V}$ are defined to apply only after a particular metastable state, with a specific value of $m_{\rm P}$, has been selected.

dramatically shift the expectation value for $m_{\rm P}$ after the additional parameters have been marginalized, this can only happen if the overwhelming majority of universes near the new expectation value have values for the other parameters that are very different than ours. We would still be left with the challenge to explain why we find ourselves in a universe like ours, and not with these different parameter values.

There are significant caveats to this result. First of all, it is not clear that the selection effects in $\mathcal{S}$ should actually factorize as in Eq. (5.69). Since the diverging volumes of sub-universes is one of the circumstances that complicates making landscape predictions, we cannot be assured that the resolution of this problem will result in universes with greater inflationary expansion factors being more likely to harbor observers. Another caveat to this discussion is that little is known about the distribution $\mathcal{I}(m_{\rm P})$. As we have defined it, this term receives two separate contributions. One contribution comes from the distribution of $m_{\rm P}$ values over the landscape, that is the frequency of $m_{\rm P}$ values among the number of metastable states that are allowed by the underlying theory. A second contribution comes from the dynamics of the multiverse, which may prefer certain metastable states over others as the multiverse evolves in time. This is because the tunneling and diffusion rates of quantum fields will in general depend on $m_{\rm P}$, such that metastable states with certain values of $m_{\rm P}$ will appear more frequently within the multiverse than others. This $m_{\rm P}$ dependence within $\mathcal{I}(m_{\rm P})$ could be very strong; see for example the studies of quantum diffusion in Refs. [128, 129, 130].

Therefore $\mathcal{I}(m_{\rm P})$ may depend more sharply on $m_{\rm P}$ than does $\mathcal{V}(m_{\rm P})$, with a local peak within the anthropic range. This might at first seem incredibly fortuitous. However, the situation is very different from the case of the cosmological constant ρ_Λ. In that case we observe ρ_Λ to be very far from its 'natural' value and therefore we must presume a very diverse and densely packed landscape in order for the value that we observe to exist at all. However, since we do not know the natural value of $m_{\rm P}$, its landscape window could be much smaller.[6] In addition, it is possible that the landscape is not as densely populated as we have presumed, in particular once we restrict attention to metastable states in every way like ours except in the value of $m_{\rm P}$. For example, if the spacings between allowed values of $m_{\rm P}$ are significant next to the size of the anthropic window, then our value of $m_{\rm P}$ might be consistent with the shape of $\mathcal{I}(m_{\rm P})$.

Furthermore, the model of inflation that describes our universe may not actually exhibit an $m_{\rm P}$-problem. This would happen, for example, if the number of e-folds of inflation that describe this model were independent of $m_{\rm P}$ or had a maximum for some finite value of $m_{\rm P}$. An interesting

---

[6] It is tantalizing that within the context of weighting universes by inflationary expansion factors, the chaotic inflation model with N set by either Eq. (5.74) or Eq. (5.75) pushes $m_{\rm P}$ to larger values, while the largeness of $m_{\rm P}$ relative to other mass scales is well-noted in our universe. Let us assume a fundamental scale $M \sim M_{\rm GUT}$. Then in this case we expect $m_{\rm P} \gg M_{\rm GUT}$, and it is possible that $m_{\rm P} \sim 10^3 M_{\rm GUT}$ is simply the largest that the landscape allows. Furthermore, Eq. (5.74) pushes the inflaton mass m_φ to smaller values, and perhaps m_φ is the smallest that the landscape allows. Thus we obtain the apparent hierarchy

$$m_{\rm P} \gg M_{\rm GUT} \gg m_\varphi \,. \tag{5.78}$$

example of the latter case occurs when the effective Planck mass is not fixed within our metastable state, but evolves as in Brans-Dicke theory. This scenario has been studied in Refs. [128, 129, 130], where it is shown that for some non-minimally coupled models of inflation, the inflationary expansion factor is maximized when inflation ends at some finite value of $m_{\rm P}$. Note however that for this or any other model of inflation to avoid the $m_{\rm P}$-problem, it would have to generate more e-folds of inflation than all of the other anthropically viable possibilities within the landscape. Moreover, the value of $m_{\rm P}$ that maximizes N would have to lie within the anthropic window.

Finally, it is possible that the analysis of Section 5.2 missed or underestimated an important anthropic condition. This might appear as the most attractive possibility, but one must be careful to appreciate the strength of the exponential dependence within $\mathcal{V}(m_{\rm P})$. In order to cancel this exponential dependence and thus make the observed value of $m_{\rm P}$ reasonably likely, an anthropic constraint must appear to exponentially suppress the likelihood for observers to arise within our universe. The observed prevalence of galaxies, long-lived stars, supernovae, and planets, along with the observation that our solar system does not seem to occupy a particularly over- or under-dense region within the Milky Way, all seem to suggest that this is not the case. Since there do not yet exist experimentally confirmed theories for inflation, reheating, and baryogenesis, it is still possible that one of these processes presents an anthropic selection effect that provides this exponential suppression. This possibility is explored relative to the reheating temperature and baryogenesis in the context of the σ-problem in Ref. [183]. Since the reheating temperature in general also depends on $m_{\rm P}$, this analysis applies equally to our scenario.

5.4 Anthropic Constraints on Λ and the Scale of Gravity

It is straightforward to extend the analysis of Section 5.2 to the case where both $m_{\rm P}$ and the cosmological constant ρ_Λ may (independently) scan over a landscape. The only constraint that is affected by this generalization is the requirement that over-densities separate from the Hubble flow before their growth is halted by the domination of the cosmological constant. The maximum amplitude reached by a linear rms over-density in this scenario is,

$$\sigma_\infty \approx 1.44 \times \frac{3}{5}\frac{a_\Lambda}{a_{\rm eq}}\sigma_{\rm eq} \approx 3.20\, \hat{\rho}_{\rm eq}^{1/3}\hat{\rho}_\Lambda^{-1/3}\hat{\sigma}_{\rm eq}\,. \tag{5.79}$$

On the other hand, an over-density has separated from the Hubble flow when a linear analysis gives $\sigma \geq 1.69$ [37, 161]. Therefore an rms fluctuation will eventually form a halo if $\sigma_\infty \geq 1.69$, which gives the generalization of Eq. (5.28):

$$\hat{\rho}_{\rm eq}\hat{\sigma}_{\rm eq}^3\hat{\rho}_\Lambda^{-1} \gtrsim 0.1\,, \tag{5.80}$$

where we find it convenient to henceforth use $\hat{\sigma}_{\rm eq}$ instead of $\hat{\zeta}_{\rm eq} s$. This is the only result from Section 5.2 that changes when ρ_Λ may scan over the landscape.

Clearly, Eq. (5.80) is weakened as ρ_Λ is decreased from the value it obtains within our universe. In this case, Eq. (5.28) eventually ceases to be the strongest constraint and $m_{\rm P}$ is bounded from above by one of the other curves in Figs. 5.1–5.4. We may also interpret Eq. (5.80) as an upper bound on ρ_Λ for a specified value of $m_{\rm P}$. In universes with a larger value of $m_{\rm P}$, ρ_Λ is then more tightly bound than in our universe. However, in universes where $m_{\rm P}$ is smaller than in our universe, the bound on ρ_Λ may be significantly weakened. This effect can be dramatic. For example, if we take $\hat{m}_{\rm P} = 0.1$ and if $\alpha = 1$ and $\beta = 0$, then ρ_Λ may be increased by a roughly a factor of ten million and still satisfy Eq. (5.80). Of course, to determine the most likely range within which to observe ρ_Λ requires to determine the prior distribution $\mathcal{I}(\rho_\Lambda, m_{\rm P})$ and to incorporate all of the selection effects into a factor $\mathcal{S}(\rho_\Lambda, m_{\rm P})$, as described in Section 5.3. Both of these tasks are beyond the scope of this work.

Nevertheless, it is worthwhile to proceed but within a very simplified picture. While our level of analysis does not permit even an approximate landscape prediction, our results do imply restrictions on the dependence of $\mathcal{S}$ and $\mathcal{I}$ on $m_{\rm P}$. Our first assumption is that the landscape is so densely packed that we can approximate the prior distribution $\mathcal{I}(m_{\rm P}, \rho_\Lambda)$ to be a continuous and smooth function of both $m_{\rm P}$ and ρ_Λ. Then we can write the probability distribution for ρ_Λ in the form

$$P(\rho_\Lambda) \propto \int \mathcal{S}(m_{\rm P}, \rho_\Lambda)\mathcal{I}(m_{\rm P}, \rho_\Lambda) dm_{\rm P} \,. \tag{5.81}$$

We discuss in Section 5.3 and in Section 5.6 how our universe appears extremely unlikely to be observed if $\mathcal{S}$ contains a factor proportional to the inflationary expansion factor. Since we wish to expose additional restrictions on $\mathcal{S}$ and $\mathcal{I}$, we now assume that $\mathcal{S}$ does not contain this factor.

It is helpful to first consider the distribution $P(\rho_\Lambda)$ when $m_{\rm P}$ is fixed to the value obtained within our universe. This corresponds to taking $\mathcal{S} \propto \delta(\hat{m}_{\rm P} - 1)$ and thus eliminating the integral in Eq. (5.81). Refs. [32, 33] argue that it is appropriate to restrict attention to only positive values of ρ_Λ and to take the distribution $\mathcal{I}$ to be roughly independent of ρ_Λ. Although the rms fluctuation $\sigma_{\rm eq}$ is constrained by Eq. (5.80), any particular over-density may be larger or smaller than $\sigma_{\rm eq}$. This implies that galaxies of a given mass will form in universes even when ρ_Λ is larger than what is allowed by Eq. (5.80). On the other hand, galaxies of a given mass become statistically rarer as ρ_Λ is increased. To account for this, it is customary to speculate that the likelihood for a particular universe to be observed is proportional to the fraction of its total mass that collapses into galaxies with masses above some minimum $\mu_{\rm min}$ [32, 33]. This minimum galaxy mass is presumably set by other anthropic considerations.

The spectrum of density perturbations is at least approximately described by Gaussian statistics.

Therefore a randomly selected co-moving volume may or may not collapse, depending on the size of the matter over-density contained within the volume. We parameterize volumes using the mass μ that they enclose, measured in units of the Milky Mass, $10^{12}M_\odot$. Then the likelihood that a mass μ will eventually separate from the cosmic expansion is given by the Press-Schechter function [37, 161],

$$F(\mu) \;=\; \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_\infty(\mu)}\int_{1.69}^{\infty}\exp\left[-\frac{1}{2}\frac{z^2}{\sigma_\infty^2(\mu)}\right]dz \;=\; \mathrm{erfc}\left[\frac{0.373\,\hat\rho_\Lambda^{1/3}}{\hat\rho_{\mathrm{eq}}^{1/3}\,\hat\sigma_{\mathrm{eq}}(\mu)}\right]. \qquad (5.82)$$

The percentage of over-densities that eventually virialize is a function of the enclosed mass μ because the rms amplitude of the initial density perturbations σ_{eq} depends on μ (see Section 5.2.5). The fraction of galaxies that have mass between μ and $\mu + d\mu$ is $(dF/d\mu)d\mu$. Since $F(\mu \to \infty) = 0$, this means that the fraction of mass contained within galaxies with mass above the mass scale μ is simply the Press-Schechter function F evaluated at μ. When only the cosmological constant scans over the landscape, $\hat\rho_{\mathrm{eq}} = \hat\sigma_{\mathrm{eq}} = 1$. This gives $P(\rho_\Lambda) \propto \mathcal{S}(\rho_\Lambda) \propto F(\mu_{\min}, \rho_\Lambda)$ [33, 117].

In order to study the scenario where both ρ_Λ and m_{P} scan over the landscape, we adopt a very simplified picture. First, we assume that $\mathcal{I}$ is independent of both ρ_Λ and m_{P} over the anthropically allowed window. We emphasize that, unlike the case with ρ_Λ, we are unaware of any physical justification for this assumption regarding m_{P}. Second, we restrict our attention to galaxies with masses near the mass of the Milky Way. We perform this restriction simply so that we may ignore the scale dependence of anthropic constraints. It turns out that values of m_{P} somewhat larger than our own do not contribute significantly toward $P(\rho_\Lambda)$. To highlight this result we simply neglect all constraints on increasing m_{P}. On the other hand, the selection effects that bound m_{P} from below are very important when determining $P(\rho_\Lambda)$. For simplicity we consider selection effects from only one additional constraint; which is that stellar encounters are rare enough on average to allow for life to evolve in the intervening time. According to Figs. 5.1–5.4, this is usually the strongest constraint on decreasing m_{P}. The exception appears to be the case of low α and low β, where the galactic cooling constraint can interfere and the stellar lifetime constraint is not far below the close encounters constraint. We simply ignore the cooling constraint and note that we could just as well evaluate $P(\rho_\Lambda)$ for galaxy masses somewhat below the mass of the Milky Way to obtain a similar result. To account for stellar lifetimes, we impose a hard cut-off below $\hat m_{\mathrm{P}} = 0.1$.

As mentioned above, the mass fraction within galaxies with masses between μ and $\mu + d\mu$ is $\delta F \equiv (dF/d\mu)d\mu$. This quantity in general depends on the time at which one looks at the universe. We count galaxies in the infinite future, which is practically equivalent to counting galaxies at any time after the domination of ρ_Λ. Then

$$\frac{dF}{d\mu} \propto \frac{1}{\sigma_\infty^2}\left|\frac{d\sigma_\infty}{d\mu}\right| e^{-1.43/\sigma_\infty^2}\,. \qquad (5.83)$$

The only μ dependence within F stems from the dependence on $\sigma_\infty \propto \sigma_{\rm eq} \propto s(\mu)$, where $s(\mu)$ is given by Eq. (5.23). Within any given universe, to consider only galaxies with a particular mass μ in the far future is equivalent to selecting only over-densities with a particular amplitude at equality. This is because within that universe over-densities with smaller amplitudes will form galaxies with smaller mass while over-densities with larger amplitudes will form galaxies with larger mass (recall that we look at the universe after ρ_Λ domination when the growth in over-densities has halted). The amplitude of the initial over-density that is selected by looking at a particular galaxy mass μ is the one that gives $\sigma(\mu) = 1.69$ in the infinite future.

We must now account for the close encounter constraint mentioned above. This constraint is converted into a selection effect by noting that if the rate of disastrous encounters between stellar systems is γ, then the probability that a stellar system will survive for a time τ is $e^{-\gamma\tau}$. The rate γ is discussed in Section 5.2.7. Note that it depends on the amplitude of the initial over-density that seeded the galaxy. We restrict our attention to galaxies with masses near to the mass of the Milky Way. As described above, these galaxies only come from over-densities that satisfy $\sigma(\mu \approx 1) = 1.69$ in the infinite future. At equality, these over-densities have an amplitude

$$\sigma \approx 5 \times 10^{-4}\, \hat{\rho}_\Lambda^{1/3} \hat{\rho}_{\rm eq}^{-1/3} \,. \tag{5.84}$$

Now we can write the likelihood $P_{\rm ss}$ that a stellar system will survive for at least a time τ within this set of galaxies. We take $\tau = \tau_{\rm evol} \approx 5 \times 10^9$ yrs, which gives

$$P_{\rm ss} \approx \exp\left(-7 \times 10^{-7} \hat{m}_{\rm P}^{-3} \hat{\mu}_b^{1/2} \hat{f}_b^{6}\, \hat{\rho}_\Lambda^{4/3} \mu^{2/3}\right), \tag{5.85}$$

where the dependence on a general mass scale μ has been restored for future reference.

So far our assumptions correspond to weighting universes by the fraction of stellar systems that survive close encounters for longer than $\tau_{\rm evol}$ and that exist in galaxies with mass near to the Milky Way mass. We also require $m_{\rm P} \geq 0.1$ in order to ensure that sufficiently long-lived stars exist in these universes. Finally, we should account for the fact that the abundance of baryons relative dark matter will depend on the value of $m_{\rm P}$ within each universe. Putting all of this together gives the probability density,

$$P(\hat{\rho}_\Lambda) \propto \int_{\mu_{\rm min}}^{\mu_{\rm max}} d\mu \int_{0.1}^{\infty} dm_{\rm P} f_b(m_{\rm P}) P_{\rm ss}(m_{\rm P}, \hat{\rho}_\Lambda, \mu) \frac{d}{d\mu} F(m_{\rm P}, \hat{\rho}_\Lambda, \mu)\,. \tag{5.86}$$

The full $m_{\rm P}$ dependence of f_b, $P_{\rm ss}$, and $dF/d\mu$ is found by substitution of the results from Section 5.2. In general Eq. (5.86) is integrated over a window $\mu_{\rm min} \leq \mu \leq \mu_{\rm max}$, but as motivated above, in our main analysis we restrict to a narrow window about $\mu = 1$. Finally, as explained previously, Eq. (5.86) makes the simplifying but unrealistic assumption that $\mathcal{I}(m_{\rm P}, \rho_\Lambda) \approx$ constant.

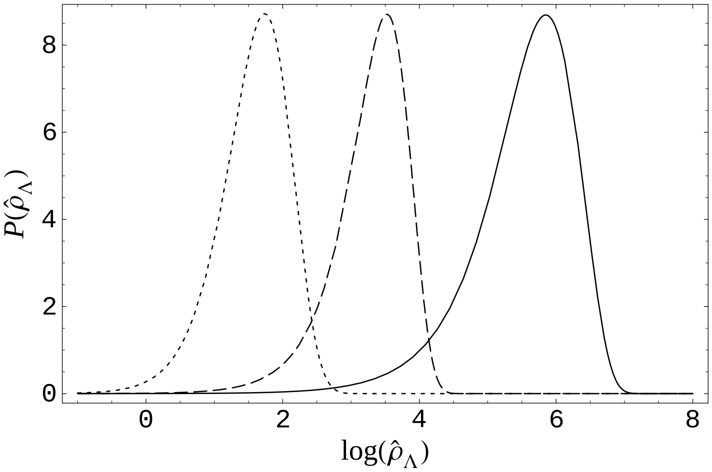


Figure 5.5: The distribution $P(\hat{\rho}_\Lambda)$ displayed against $\log(\hat{\rho}_\Lambda)$. The solid curve is $P(\hat{\rho}_\Lambda)$ marginalized over universes in which $m_{\rm P}$ may vary and with $\alpha = 1$ and $\beta = 0$, the longer-dashed curve is the same quantity but for $\alpha = 3$ and $\beta = 3/2$, while the shorter-dashed curve is $P(\hat{\rho}_\Lambda)$ evaluated when $m_{\rm P}$ is fixed to the value within our universe. All three distributions are for fixed galactic masses $\mu = 1$. See text for results obtained for a range of galactic masses. The normalizations are chosen for clarity.

The results of a numerical computation of $P(\rho_\Lambda)$ are displayed in Fig. 5.5. For reference, we also display the result when $m_{\rm P}$ is fixed to the value obtained within our universe (note that this corresponds to $dF/d\mu|_{\mu=1}$ and not $F(\mu = 1)$). Our value of ρ_Λ corresponds to the origin on this graph. Evidently the assumptions of this section render the observation of ρ_Λ at or below our value very unlikely. In fact, the fraction of $P(\rho_\Lambda)$ that sits below $\hat{\rho}_\Lambda = 1$ is about 7×10^{-5} for $\alpha = 1$, $\beta = 0$ and about 4×10^{-4} for $\alpha = 3$, $\beta = 3/2$. Since relatively large values of ρ_Λ receive significant weight only when $m_{\rm P}$ is relatively small, we also see that most of the weight of these distributions comes from values of $m_{\rm P}$ that are below the value obtained within our universe. This is evidence of a sort of 'statistical pressure' that gives greater weight to those values of marginalized parameters that permit a larger value of ρ_Λ. This is why it was unimportant to account for selection effects that constrain $m_{\rm P}$ from above.

Here we note the importance of the close encounters constraint and of recombination timing constraint (i.e. the "$t_{\rm rec} = e^{-2/3} t_{\rm vir}$" constraint) in bounding the anthropically allowed variation of $\hat{\rho}_\Lambda$. Inspecting Figs. 5.1–5.4 indicates that of the constraints limiting $m_{\rm P}$ from below, the close encounters bound is the strongest. For $\mu = 1$, it bounds $\hat{m}_{\rm P} \gtrsim 10^{-0.2}$ for $\alpha = 3$ and $\beta = 3/2$, and approximately $\hat{m}_{\rm P} \gtrsim 10^{-0.8}$ for $\alpha = 1, \beta = 0$. Inserting these limits on $\hat{m}_{\rm P}$ into Eq. (5.80) gives maximum values for $\hat{\rho}_\Lambda$ in good agreement with the peaks in Fig. 5.5. Had only the star lifetime constraint $\hat{m}_{\rm P} \gtrsim 0.1$ been imposed, the maximum allowed value of $\hat{\rho}_\Lambda$ would have been much larger. For $\alpha = 1$ and $\beta = 0$, $\hat{\rho}_\Lambda$ could be as large as 4×10^7. In the case of $\alpha = 3$ and $\beta = 3/2$, $\hat{\rho}_\Lambda$ can be as large as 2×10^{14}. But the anthropic constraints in Figs. 5.1–5.4 for this latter model indicate that the "$t_{\rm rec} = e^{-2/3} t_{\rm vir}$" constraint is stronger than the star lifetime constraint. Imposing the recombination constraint requires instead that $\hat{\rho}_\Lambda$ be no larger than about 4×10^8.

The analysis that leads to the curves in Fig. 5.5 gives at best a crude approximation for the actual probability distribution for ρ_Λ. One improvement to the analysis would be to weight universes by the mass fraction that collapses into galaxies that have a range of anthropically favorable masses, instead of the fraction that collapses into only galaxies with the Milky Way mass. Including galaxies with greater masses will tend to push the weight of the distributions toward smaller ρ_Λ, while including galaxies with smaller masses pushes the weight of the distributions toward larger ρ_Λ. We have checked that under the assumptions outlined above, allowing for a range of galaxy masses $0.1 \leq \mu \leq 10$ tends to push the weight of the distributions $P(\rho_\Lambda)$ to slightly larger values of ρ_Λ.

Previous calculations of the distribution $P(\rho_\Lambda)$ integrate over all galaxy masses equal to or larger than the Milky Way mass. Although our anthropic considerations offer no reason to ignore galaxies with mass below that of the Milky Way, and galactic cooling constraints limit the formation of galaxies with larger masses, we nevertheless consider the evaluation of Eq. (5.86) for a range of masses $1 \leq \mu \leq \infty$. Integrating over $0 \leq \hat{\rho}_\Lambda \leq 1$ gives the probability P_* that observers in such galaxies would observe a cosmological constant less than or equal to our own. Numerically, we find $P_* = 1 \times 10^{-3}$ for $\alpha = 3, \beta = 3/2$ and $P_* = 2 \times 10^{-4}$ for $\alpha = 1, \beta = 0$. For comparison, we find that in our universe $P_* = 0.06$. To reiterate, these calculations ignored any additional selection effects that might depend on μ, such as the effects of different galactic cooling rates.

Another improvement to the analysis would be to include more $m_{\rm P}$-dependent selection effects. The abundance of heavy elements and of long-lived stars with appropriate surface temperatures both seem important when determining the likelihood for observers to arise within a universe. However, the analysis of Section 5.2 does not shed light on how to calculate these selection effects. One thing that is clear is that the range of typical stellar masses scales as $m_{\rm P}^3$. This means that galaxies of a fixed mass will contain more stars as $m_{\rm P}$ is decreased. If all else were equal this would result in a greater number of observers per unit baryon mass in a galaxy, which would tend to push the weight of $P(\rho_\Lambda)$ toward larger values of ρ_Λ. In addition, the rate of destructive encounters γ is a function of stochastic variables, including for example the spin parameter λ. The statistical distribution of these variables could tend to strengthen or weaken the close encounter constraint as a function of $m_{\rm P}$. However, when everything else is equal the 'statistical pressure' alluded to below Eq. (5.86) tends to give greater weight to those values of stochastic variables that allow for a smaller $m_{\rm P}$ and larger ρ_Λ.[7]

Of course, a proper calculation of $P(\rho_\Lambda)$ requires an understanding of the prior distribution for

---

[7] We have confirmed this phenomena with the following simple example. The close encounter rate γ is proportional to a factor $f_\star$ that accounts for the increased density of the galactic disk relative the dark matter halo (see Section 5.2.5.3). This factor depends sensitively on λ, $f_\star \propto \lambda^{-8}$. In the preceding analysis, we normalized the factor $f_\star$ so as to give the correct stellar density within our neighborhood of the Milky Way. However, λ is a stochastic

$m_{\rm P}$, $\mathcal{I}(m_{\rm P})$. It must be emphasized that shape of $\mathcal{I}(m_{\rm P})$ could dramatically influence the shape of the distribution $P(\rho_\Lambda)$. Therefore the results of this section are best understood as an empirical restriction on the dependence of $\mathcal{I}(m_{\rm P})$ on $m_{\rm P}$. Since this is our main point in this section, let us be very explicit. The curves displayed in Fig. 5.5 suggest that within a very simplified landscape picture, it is very unlikely to observe a value of ρ_Λ that is at or below the value within our universe. This means that if a landscape picture is to describe our universe, it should contain important ingredients that were neglected in our analysis. In addition, these additional ingredients should provide a strong emphasis for larger values of $m_{\rm P}$. Thus we conclude that for a landscape picture to describe our universe as among those that are likely to be observed, it is necessary that $\mathcal{I}(m_{\rm P})$ or some other neglected selection effect must receive the vast majority of its weight for values of $m_{\rm P}$ that are very near to or larger than the value obtained within our universe. We emphasize that the analysis of this section did not assume that the inflationary expansion factor enters into landscape calculations.

5.5 Nonstandard Paths toward Structure Formation

One might wonder what are the constraints on structure formation if we do not assume that dark matter dominates over baryonic matter, or that virialization occurs after recombination. If dark matter does not dominate over baryonic matter, then the evolution of over-densities in the dark matter does not significantly affect the evolution of over-densities in baryons. Without appreciable dark matter potential wells, baryon over-densities do not grow (even logarithmically) until after recombination. This is because in the era before recombination, the Jeans length for the tightly coupled baryons,

$$R_J = \sqrt{\tfrac{8}{3}\pi} H^{-1} v_s \,, \tag{5.88}$$

where $v_s = 1/\sqrt{3}$ is the speed of sound prior to recombination, is always larger than the Hubble radius. Growth therefore does not occur in either the radiation or baryon-dominated era until after recombination. Between recombination and the domination of cosmological constant the evolution

---

variable. N-body simulations suggest that the distribution for λ can be approximated using [184]

$$P_\lambda(\lambda) d\lambda \propto \frac{d\lambda}{\lambda} \exp\left[-2\ln^2(28.6\lambda)\right] \,. \tag{5.87}$$

This distribution has a peak at about $\lambda \approx 0.03$, while the Milky Way appears to be described by $\lambda \approx 0.06$ [181]. This implies that typical values of λ more tightly constrain $m_{\rm P}$ than the value represented in Fig. 5.5. Therefore, one might expect that when we treat λ as a stochastic variable with distribution $P_\lambda(\lambda)$, that the weight of the distributions $P(\rho_\Lambda)$ will shift to smaller values of ρ_Λ. In fact, the opposite trend occurs, as the previously mentioned 'statistical pressure' is such that the weight of the distributions $P(\rho_\Lambda)$ actually shifts the location (in $\hat{\rho}_\Lambda$) of the peak by approximately an order of magnitude toward larger values of ρ_Λ.

of over-densities may be approximated by

$$\sigma \approx (a/a_{\rm rec})\sigma_{\rm rec}\,. \tag{5.89}$$

The spectrum of fluctuations at recombination $\sigma_{\rm rec}$ is scale dependent in the sense that it is constant for scales larger than the Hubble radius at recombination but rapidly decreases to zero as one looks at smaller distance scales. This is because of the tight coupling between baryon and radiation over-densities, and because the latter decay after they enter the Hubble radius.

As in the standard picture, after the domination of cosmological constant over-densities will grow by a factor of 1.44 and then stop. Thus the maximum amplitude achieved by a linear analysis of an rms fluctuation is

$$\sigma_\infty \approx 1.44 \times (a_\Lambda/a_{\rm rec})\sigma_{\rm rec} \approx 5\times 10^{-2}\,\hat\rho_{\rm eq}^{1/3}\hat\zeta_{\rm eq}\,. \tag{5.90}$$

Here we have used that $\sigma_{\rm rec} \approx 5\times 10^{-5}\hat\zeta_{\rm eq}$ on scales larger than the Hubble radius at recombination, and that recombination occurs at a temperature $T_{\rm rec} \approx 3000$ K, where we ignore the logarithmic dependence of $T_{\rm rec}$ on $m_{\rm P}$ and η. The formation of structure still requires that a linear analysis gives $\sigma_\infty \geq 1.69$ before the growth in over-densities is halted by the domination of cosmological constant. This gives the constraint,

$$\hat\rho_{\rm rec}\hat\zeta_{\rm eq}^{\,3} \gtrsim 5\times 10^5\,. \tag{5.91}$$

Eq. (5.91) constrains $m_{\rm P}$ according to

$$\left(\frac{1}{6}\hat m_{\rm P}^{-\beta} + \frac{5}{6}\hat m_{\rm P}^{-1}\right)\hat m_{\rm P}^{-3\alpha} \gtrsim 5\times 10^5\,. \tag{5.92}$$

This constraint is much stronger than the constraint it replaces, Eq. (5.29).

Allowing for baryons to dominate the matter density of the universe may affect the other constraints in Figs. 5.1–5.4 in two ways. First, the halo density and background density at virialization, $\rho_{\rm vir}$ and ρ_*, are now reduced by a factor of 3×10^{-6} due to the difference between $\sigma_{\rm rec}$ and $\sigma_{\rm eq}$. Second, structure formation only occurs on scales greater than the Hubble radius at recombination, since sub-horizon perturbations are suppressed. Ignoring the $m_{\rm P}$ dependence in $T_{\rm rec}$, this implies a minimum halo mass set by the horizon mass at recombination, corresponding to a scale of roughly $\mu_{\rm min} \sim 10^6$. It can be shown that no value of $m_{\rm P}$ satisfies all of the constraints displayed in Figs. 5.1–5.4 after these effects have been included. Dropping the constraint that virialization precede recombination does not change this result.

We now turn to the second assumption of Section 5.2.5, which is that recombination occurs at least an e-fold of expansion before virialization. To investigate what happens when virialization occurs before recombination, we adopt the following simplified picture. Dark matter over-densities grow when they enter the Hubble radius, and we assume that they become non-linear and virialize as they would within our universe. However, growth in the baryon over-densities is hampered by their interaction with the photon Hubble flow before recombination. Therefore we approximate that baryons do not participate at all in the over-densities of the dark matter and are rarefied relative the halo density as they follow the Hubble flow.

Within this simplified model, the final baryon fraction within a halo will be at most about $f_b/18\pi^2$, and will decrease by a factor of $e^{-3} \approx 0.05$ for each e-fold of expansion between virialization and recombination. However, it turns out that only two of the constraints that we consider depend significantly on the baryon fraction of the halo. These are the disk instability constraint of Section 5.2.5.3 and the close encounters constraint of Section 5.2.7. (The explicit f_b dependence in the galactic cooling constraint of Section 5.2.5.2 is canceled by an implicit dependence within Λ_c.) To explore whether this situation opens a new window for allowed values of m_{P}, it is helpful to adopt the following picture. Instead of simply eliminating the recombination timing constraint of Eq. (5.31), we continuously weaken it. For example, we may demand that recombination occur at most N_{rec} e-folds of expansion after virialization and then study the above constraints as N_{rec} is increased.

When we do this, we find that the curves in Figs. 5.1–5.4 corresponding to the disk instability constraint, the close encounters constraint, and the recombination timing constraint all slide downward as N_{rec} is increased. This shifts the allowed window for m_{P} such that larger values of m_{P}, including the value obtained within our universe, become excluded as lower values become allowed. It turns out that the disk instability curve slides downward at a faster rate than that of the recombination timing curve, so that as the window for allowed m_{P} moves to smaller m_{P} it also grows smaller. Ultimately, the window gets pushed against other constraints, such as the stellar lifetime constraint or the galactic cooling constraint, and disappears. This happens at about $N_{\mathrm{rec}} \approx$ a few.

5.6 Analysis of a Structure Formation Constraint

In Section 5.3 it is argued that if the probability to observe a particular value of m_{P} is weighted in part by the inflationary expansion factor of universes that contain that value of m_{P}, then it is overwhelmingly preferred that m_{P} should be measured at one of the boundaries of its anthropic window. It is clear from the discussion of Section 5.2 that our value of m_{P} is not at either of its anthropic boundaries. Nevertheless, it is worthwhile to investigate more quantitatively just how 'far' is our value of m_{P} from its anthropic boundaries. For simplicity we investigate the selection effect from only one anthropic constraint. Specifically, we look at the structure formation requirement

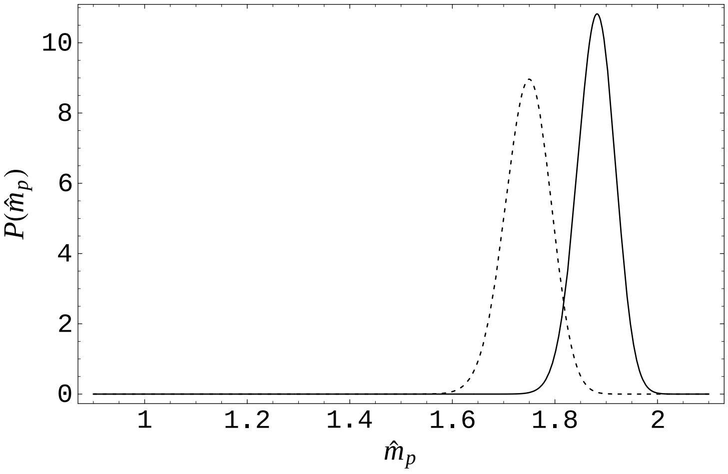


Figure 5.6: The normalized distribution $P(\hat{m}_{\rm P})$ for $N = 60\,\hat{m}_{\rm P}$ (solid) and $N = 60\ln(\hat{m}_{\rm P})$ (dashed) for when the landscape distribution for $m_{\rm P}$ depends on the inflationary expansion factor. Our universe corresponds to $\hat{m}_{\rm P} = 1$. In both cases F is defined using $\alpha = 3$ and $\beta = 3/2$. Although in both cases selection effects appear to prefer a specific range for $m_{\rm P}$, our value is far outside of this range.

that halos virialize before the domination of cosmological constant (Section 5.2.5.1). Note that this provides the tightest constraint on $m_{\rm P}$ according to the curves in Figs. 5.1–5.4.

The arguments of Section 5.3 are appropriate primarily when the landscape is so densely packed that we can approximate the prior distribution $\mathcal{I}(m_{\rm P})$ to be a continuous function of $m_{\rm P}$ within the anthropic window. The probability to observe $m_{\rm P}$ to lie within the range $dm_{\rm P}$ can then be written,

$$P(m_{\rm P})dm_{\rm P} \propto \mathcal{A}(m_{\rm P})\mathcal{V}(m_{\rm P})\mathcal{I}(m_{\rm P})dm_{\rm P}\,, \tag{5.93}$$

where the factors on the right-hand side are defined in Section 5.3. Of course, we are assuming that universes are weighted in part by their inflationary expansion factor $\mathcal{V}$. Since we have no knowledge about the shape of $\mathcal{I}(m_{\rm P})$, we take $\mathcal{I}(m_{\rm P}) \approx$ constant. As suggested above, we take the anthropic factor $\mathcal{A}$ to be conditioned by only the constraint that halos virialize before the domination of cosmological constant prevents this.

To proceed, we assume that the likelihood to observe a given value of $m_{\rm P}$ is proportional to the baryon fraction within galaxies of mass greater than or equal to the mass of the Milky Way, $10^{12}M_{\odot}$. Allowing for smaller galaxies or allowing for observers that do not require a galactic environment can only expand the window of allowed $m_{\rm P}$. The Press-Schechter function [37] gives the fraction of matter that collapses into a galaxy of mass greater than or equal to a given scale. It is derived in

Section 5.4 and given by Eq. (5.82). We reproduce it here for convenience,

$$F = \operatorname{erfc}\left[0.373\left(\frac{\hat{\rho}_\Lambda}{\hat{\rho}_{\rm eq}\hat{\sigma}_{\rm eq}^3}\right)^{1/3}\right] . \tag{5.94}$$

The prefactor comes in part from evaluating F in the infinite future and at the Milky Way mass scale. The $m_{\rm P}$ dependence is given by

$$\left(\frac{\hat{\rho}_\Lambda}{\hat{\rho}_{\rm eq}\hat{\sigma}_{\rm eq}^3}\right)^{1/3} = \left(\frac{1}{6}\hat{m}_{\rm P}^{-\beta} + \frac{5}{6}\hat{m}_{\rm P}^{-1}\right)^{-4/3} \hat{m}_{\rm P}^{\alpha} . \tag{5.95}$$

Note that when α and β are positive, F is a decreasing function of increasing $m_{\rm P}$. We are interested in the baryonic matter within galaxies. Therefore $\mathcal{A}$ should contain a factor of the baryon fraction f_b, given by Eq. (5.18), along with F.

Finally, we take $\mathcal{V}(m_{\rm P}) \propto e^{3N}$ for N e-folds of inflation. We require that $N(m_{\rm P})$ be an increasing function of $m_{\rm P}$ so that $m_{\rm P}$ is pushed to larger values, saturating the constraint in Eq. (5.29). In addition, we want $\alpha = 3$ and $\beta = 3/2$ so that this constraint on increasing $m_{\rm P}$ is as strong as possible. Rather than propose a specific model of inflation, we assume that one can contrive a model with the relatively weak dependence $N \approx 60\,\hat{m}_{\rm P}$. Then putting all of our assumptions together gives the probability distribution,

$$P(m_{\rm P}) = \mathcal{N} f_b(m_{\rm P}) F(m_{\rm P}) e^{3N(m_{\rm P})} , \tag{5.96}$$

where $\mathcal{N}$ is a normalization factor. The normalized distribution $P(m_{\rm P})$ is displayed in Fig. 5.6 for $\alpha = 3$ and $\beta = 3/2$. Although it is intriguing that in this scenario the erfc function overcomes the exponential volume factor, this happens for a value of $m_{\rm P}$ somewhat larger than the value that we observe. This $\mathcal{O}(1)$ change in $m_{\rm P}$ is significant due to the exponential sensitivity of P on $m_{\rm P}$. Indeed, the fraction of the distribution function $P(m_{\rm P})$ that sits below $\hat{m}_{\rm P} = 1$ is completely negligible compared to that which sits above (explicitly, this fraction is roughly 10^{-53}). This distribution is so sharply peaked because N is relatively large. For example, expanding about the local maximum gives $P \sim \exp[-cN\Delta^2]$ where $c \sim \mathcal{O}(1)$ and Δ is the difference between $\hat{m}_{\rm P}$ and its value at the maximum. Even if the number of e-folds depends very weakly on $m_{\rm P}$, for example $N \approx 60\ln(\hat{m}_{\rm P})$, we still find the preference for larger $m_{\rm P}$ to be overwhelming.

We now provide a final point of clarification. A careful reader may notice that according to Fig. 5.6, the values of $m_{\rm P}$ that are most likely to be observed lie well outside the anthropically allowed windows of Figs. 5.1–5.4. This is because the relevant curves in Figs. 5.1–5.4 are calculated by assuming that all over-densities have initial amplitudes equal to the rms amplitude. Meanwhile, Fig. 5.6 takes into account that the initial amplitude of an over-density is at least approximately a

Gaussian random variable. The discrepancy between the results in Figs. 5.1–5.4 and 5.6 reflect that under the assumptions of this section, the overwhelming majority of galaxies stem from over-densities that begin with amplitudes many standard deviations away from the norm. Although these galaxies result from relatively unlikely initial over-densities, the fact that they arise within enormously larger universes more than compensates for this. This result stems from the sharp dependence on $m_{\rm P}$ in the inflationary expansion factor. If the proper landscape measure does not contain this factor, then the distribution for $m_{\rm P}$ would be very different than Fig. 5.6 indicates.

5.7 Conclusions

If the magnitude of the apparent Planck mass $m_{\rm P}$ may scan across a landscape of possibilities, then there may exist universes with physical parameters and interactions in every way like those within our universe except for their value of $m_{\rm P}$. We have calculated the range over which $m_{\rm P}$ may scan over such universes while still satisfying a number of anthropic constraints. Perhaps not surprisingly, if we combine all of the anthropic constraints we find a rather narrow window for allowed $m_{\rm P}$. The results for WIMP dark matter and representative models of inflation and baryogenesis are displayed in Figs. 5.1–5.4. Of course, the window for allowed $m_{\rm P}$ is expanded if one loosens the anthropic criteria.

More interestingly, this window will expand if an important cosmological quantity is determined by a stochastic process. For example, many scenarios to generate a primordial curvature perturbation depend on the local vev of a light scalar field, as does the density of dark matter when it is determined by the axion. If these models apply, then the curvature perturbation and/or dark matter density are not correlated with changes in $m_{\rm P}$, and a much larger window for $m_{\rm P}$ may be able to satisfy anthropic constraints. Our purpose has been to calculate a minimal window for allowed $m_{\rm P}$, so we have not considered these possibilities in detail.

Even a very small window for allowed values of $m_{\rm P}$ has important implications for the landscape paradigm. In particular, the probability to observe a particular value of $m_{\rm P}$ may be weighted by the inflationary expansion factor of universes that contain that value. This effect inputs a strong exponential dependence on $m_{\rm P}$ into the probability distribution, which must be offset by another strong selection effect near the peak of the distribution. This other selection effect could be a very sharp peak or boundary to the underlying landscape distribution; otherwise the effect must come from an exponentially strong anthropic dependence on $m_{\rm P}$. Such a strong anthropic dependence on $m_{\rm P}$ would be in conflict with the observation that $m_{\rm P}$ has even a narrow anthropic window in our universe. This is another example of the runaway inflation problem discussed in the recent literature.

We also consider the anthropic window for the cosmological constant Λ when both Λ and $m_{\rm P}$ are allowed to independently scan over the landscape. Even when the allowed range for $m_{\rm P}$ is relatively narrow, it still allows for a significant broadening of the allowed range for Λ. This is because Λ is only constrained by the necessity that cosmic structures separate from the Hubble flow before Λ domination. Meanwhile, the time at which structures separate from the cosmic expansion is proportional to a high power of m_{P}. The result is that even for values of m_{P} within the small allowed windows of Figs. 5.1–5.4, Λ may be over ten million times larger in other universes than it is within ours. Just because Λ may be larger does not automatically imply that our value of Λ is less likely to be observed, since selection effects may ultimately weight smaller values of Λ more than larger values. We perform a very basic calculation which suggests that anthropic selection effects tend to make larger values of Λ more likely to be observed. This suggests that the observation of a cosmological constant at or below the level obtained within our universe is very unlikely unless unknown anthropic selection effects or the underlying landscape distribution of m_{P} is dominated by values very near to or larger than the value obtained within our universe.

Acknowledgments

We would like to thank Stearl Phinney, Jonathan Pritchard, Scott Thomas, and Andrei Linde for helpful discussions. This work was supported by the U.S. Department of Energy under contract Nos. DE-FG03-92ER40689 and DE-FG03-92ER40701.

Chapter 6

Quark Masses and Mixings from the Landscape

We propose that the flavor structure of the Standard Model may arise due to random selection from a landscape of solutions to a more fundamental theory. Specifically, we study a set of “Gaussian landscapes” where the Standard Model fields have localized (Gaussian) zero mode wavefunctions over some geometry of extra dimensions, and the Standard Model Yukawa couplings derive from overlap integrals involving these wavefunctions. Focusing on the quark sector, we find that the observed generation structure and pairing structure of the Standard Model are typical of landscapes where the central locations of the quark wavefunctions are allowed randomly scan over the extra dimensions.

Based on work in preparation, L. J. Hall, M. P. Salem, and T. Watari.

6.1 Introduction

A minimal extension of the Standard Model describes all laboratory data with 26 free parameters. Of these, 20 arise from the Yukawa matrices $\lambda^{u,d,e}$ and the coupling matrix C that appear in the flavor interactions

$$\mathcal{L}_{\text{flavor}} = \lambda^u u\, q\, h + \lambda^d d\, q\, h^* + \lambda^e e\, l\, h^* + \frac{C}{M}\, l\, l\, h\, h\,, \tag{6.1}$$

where q, l (u, d, e) are the left (right) handed quark and lepton fields. Of the flavor interactions, the most well known are those in the quark sector, with all ten quark flavor parameters having been measured. The present experimental limits for these parameters are given in Table 6.1. Given the continued progress on improving the accuracy of these experimental limits, along with the limits on other flavor parameters, the most striking fact is that there is nothing approaching a standard theory of the origin of these parameters. We lack not only a convincing explanation of any of the numerical values for these parameters, but we remain ignorant about the overall picture of flavor.

A complete theory of flavor would address both the quark and lepton sectors and would provide answers to three very different questions: (1) What is the origin of the fermion quantum numbers and

λ_u	$(3.0 \pm 1.0) \times 10^{-6}$	$\log_{10} \lambda_u$	$-5.53^{+0.13}_{-0.17}$
λ_c	$(1.4 \pm 0.1) \times 10^{-3}$	$\log_{10} \lambda_c$	-2.87 ± 0.03
λ_t	$(4.9 \pm 0.3) \times 10^{-1}$	$\log_{10} \lambda_t$	-0.31 ± 0.02
λ_d	$(6.7 \pm 2.7) \times 10^{-6}$	$\log_{10} \lambda_d$	$-5.18^{+0.15}_{-0.22}$
λ_s	$(1.3 \pm 0.4) \times 10^{-4}$	$\log_{10} \lambda_s$	$-3.90^{+0.10}_{-0.13}$
λ_b	$(5.7 \pm 0.1) \times 10^{-3}$	$\log_{10} \lambda_b$	-2.24 ± 0.01
$\frac{1}{\pi}\,\theta_{12}^{\text{CKM}}$	$(7.31 \pm 0.03) \times 10^{-2}$	$\log_{10}\left(\frac{1}{\pi}\,\theta_{12}^{\text{CKM}}\right)$	-1.136 ± 0.002
$\frac{1}{\pi}\,\theta_{23}^{\text{CKM}}$	$(1.344^{+0.003}_{-0.025}) \times 10^{-2}$	$\log_{10}\left(\frac{1}{\pi}\,\theta_{23}^{\text{CKM}}\right)$	$-1.872^{+0.001}_{-0.008}$
$\sin\theta_{13}^{\text{CKM}}$	$(4.01 \pm 0.09) \times 10^{-3}$	$\log_{10}\left(\sin\theta_{13}^{\text{CKM}}\right)$	-2.397 ± 0.010
δ^{CKM}	$1.00^{+0.06}_{-0.10}$	$\log_{10}\left(\delta^{\text{CKM}}\right)$	$0.00^{+0.03}_{-0.05}$

Table 6.1: The ten quark sector flavor parameters. For comparison to distributions provided throughout this chapter, the measured values of these parameters have been run up to the Planck scale assuming no physics beyond the Standard Model enters up to this scale. All data comes from the pdgLive feature from the Particle Data Group [185].

why are there three generations? (2) What determines the qualitative pattern of the quark and lepton mass matrices? For example, why do the charged fermion masses and mixings have a hierarchical pattern, while in the neutrino sector there are large mixing angles? (3) What determines the precise values of the 20 flavor parameters? Restricting attention to the quark sector, the couplings $\lambda^{u,d}$ are symmetry breaking parameters of the flavor symmetry group $U(3)^3$, where one $U(3)$ factor acts on each of q, u, d. The dominant approach to constructing theories of flavor is to use symmetries to reduce the number of free parameters, n [38, 39]. For example, an underlying flavor group $G_f \subset U(3)^3$ and a more unified gauge symmetry both limit n, leading to precise predictions if $n < n_{obs}$, the number of observables. A hierarchy of symmetry breaking scales can lead to small dimensionless parameters that explain qualitative features of the mass matrices [38], and there are many realizations with G_f both Abelian and non-Abelian. Still, it is striking that the progress along these lines is limited, even as more precise data have become available. Perhaps this is a sign that a completely new approach is needed.

The cosmological dark energy apparently has little to do with flavor. However, it may be the first evidence for a huge landscape of vacua, with the observed value of the cosmological constant resulting from environmental selection for large scale structure [32, 33]. If we take string theory to be a theory of a landscape of meta-stable states rather than a theory of a single vacuum, then what are the implications for flavor? The enormous number of vacua, required for a sufficiently fine scan of the cosmological constant, results from the large number of ways background gauge fluxes can link non-trivial compact manifolds of extra spatial dimensions [88]. The low energy four dimensional theory is likely to vary enormously from one vacuum to another. This is seen very directly in the light fermion sector, since the number of light fermions is precisely determined by these background gauge fluxes. Furthermore, in many vacua the background gauge fluxes lead to a localization of the wavefunctions of the light fermions in these extra dimensions, $\phi_i^a(y)$ for fermion of type a and generation i. The value of some entry of the usual Yukawa coupling matrices is then given by an integral over the extra dimensions y, for example for the up sector

$$\lambda_{ij}^u \propto \int \phi_i^{u_R}(y)\phi_j^{q_L}(y)\phi^h(y)\,dy\,, \tag{6.2}$$

where $\phi^h(y)$ is the profile of the Higgs. If the various wavefunctions in Eq. (6.2) are peaked at different locations, then it is very easy to see that the overlap integral could lead to a small Yukawa matrix element. Hence localization in extra dimensions is an alternative to symmetries for fermion mass hierarchies [40]. However, the localization depends on the background gauge fluxes, so matrices such as $\lambda^{u,d}$ would all scan from one universe to another. Thus the problem of flavor is drastically altered: symmetry breaking patterns are replaced with landscape probability distributions, together with possible environmental selection effects.

While there may be strong environmental effects acting on the electron, up and down masses [186], and possibly also on the top mass [100] and on the lepton sector from leptogenesis, most of the flavor parameters do not seem to be strongly selected. Therefore it could be that most flavor parameters have values that simply reflect the underlying probability distribution over an enormous number of solutions to the fundamental theory. The precise value of such observables is accidental, not fundamental, since any nearby value would be just as probable. Although it may be unappealing that there is nothing fundamental or beautiful relating the flavor parameters in our universe, so far this possibility cannot be dismissed.

In any particular meta-stable state, the Yukawa couplings have a relevant dependence on $n = n_S + n_F$ parameters, where n_S of the parameters scan and n_F of them are fixed. If $n < n_{obs}$ then data can determine the subset of vacua in which we happen to live and $n_{obs} - n$ predictions can be made. A certain simple model with a single extra dimension has $n_{obs} - n = 1$, giving a single prediction [187]. We will assume that $n > n_{obs}$, so that no such precise predictions are possible. In this case one must consider the probability distributions for the n_S parameters that scan. In a particular landscape the distributions for the flavor observables may be quite constrained, so that it is possible to determine whether the observed pattern of fermion masses is typical of this landscape. In this chapter we study the emergence of qualitative features of quark and masses and mixings that arise from scanning over a simple toy landscape that is based on the overlap of wavefunctions in extra dimensions, as in Eq. (6.2). An investigation of the lepton sector is in progress.

The large mixing angles observed in atmospheric and solar neutrino oscillations inspired the idea that the relevant Yukawa couplings of the neutrino sector are governed by randomness rather than by flavor symmetries [188]. By introducing a simple probability distribution for the elements in the Dirac and Majorana neutrino mass matrices, probability distributions for the neutrino observables were generated and found to agree well with data. In particular, realistic modest hierarchies in the ratio of Δm^2 for atmospheric and solar oscillations and in θ_{13} were found to be quite probable in this picture of Neutrino Anarchy. In other work a statistical analysis was performed on both the charged and neutral fermion sectors [189]. This is much bolder, since the hierarchies are much stronger in the charged sectors. Yet this hierarchy was found to result from choosing a simple three-parameter probability distribution for the Yukawa couplings. In Section 6.2 we analyze this scheme in some detail, studying both its accomplishments and limitations, to gain insight into some features of landscape models.

In Sections 6.3 through 6.6 we introduce simple "Gaussian landscapes" and study the resulting distributions for quark masses and mixings. These landscapes share features expected from certain string landscapes, for which they can be viewed as simplified or toy models. The key feature of a Gaussian landscape is that all quark fields and the Higgs field have zero mode wavefunctions with Gaussian profiles in the extra dimensions, and that the centers of these profiles all scan independently with flat probability distributions over the volume of the extra dimensions. For simplicity, the number of free parameters used to describe the geometry of the extra dimensions and the widths of the Gaussian profiles is kept to a minimum. The main question we address is: are the observed quark masses and mixings typical of these simple Gaussian landscapes?

In Section 6.3 we introduce a Gaussian landscape for quarks on a circular extra dimension S^1, with all Gaussian profiles having the same width. Numerical probability distributions for the nine CP conserving flavor observables are provided, and a qualitative semi-analytic description of these distributions is derived. The results are compared with those that result from introducing approximate flavor symmetries, with some similarities and some differences emerging. Finally, the effects of possible environmental selection on the top mass is studied. The large number of flavor parameters in the Standard Model allows for a reasonably significant evaluation of goodness-of-fit between a Gaussian landscape and the observed flavor structure. This is described in Section 6.4, with the S^1 Gaussian landscape used for illustration. The effects on the quark sector from adding more dimensions to the Gaussian landscape are examined in Section 6.5, together with a preliminary study of the effects of geometry. Section 6.6 contains some analytic results for reference. Finally, we provide discussion and concluding remarks in Section 6.7.

6.2 Prelude: Hierarchy without Flavor Symmetry

In previous work it has been suggested that the components of the Yukawa matrices $\lambda^{u,d,e}$ and the coupling matrix C are selected randomly and independently of each other [188, 189]. For example, in Neutrino Anarchy [188] one finds that the large mixing angles underlying neutrino oscillation are typical of the lepton interactions that arise when each element of the matrices λ^e and C is independently selected from simple probability distributions such as

$$\frac{dP(\lambda^e)}{d\ln\lambda^e} = \text{const.} \quad \text{for } \lambda^e_{\min} < \lambda^e < \lambda^e_{\max}\,, \qquad \frac{dP(C)}{d\ln C} = \text{const.} \quad \text{for } C_{\min} < C < C_{\max}\,, \tag{6.3}$$

for $\lambda_{\min,\max}$ and $C_{\min,\max}$ order unity, or from distributions such as

$$\frac{dP(\lambda^e)}{d\lambda^e} = \text{const.}\,, \qquad \frac{dP(C)}{dC} = \text{const.} \tag{6.4}$$

(for more details see [188, 190]). Such an absolute anarchy of lepton couplings tends to result in degenerate mass eigenvalues. On the other hand, Ref. [189] introduced a power-law probability distribution for the Yukawa matrix elements, $dP(\lambda)/d\lambda \propto \lambda^{\delta-1}$ for $\lambda_{\min} < \lambda < \lambda_{\max}$ and $dP(\lambda)/d\lambda = 0$ otherwise. By assuming $\lambda_{\min} \ll \lambda_{\max}$ and choosing δ appropriately, quark Yukawa matrices with each matrix element following such a distribution can roughly accommodate the hierarchical pattern of quark mass eigenvalues [189]. According to Ref. [189], $\delta = -0.16$ provides the best fit to the quark sector.

In this work we propose a significant modification to these ideas. However as a prelude to discussing our proposal it is worthwhile to first study the model of Ref. [189] in greater detail. To this we devote the remainder of this section. To simplify the analysis we specialize to the particular case $\delta = 0$, such that

$$\frac{dP(\lambda)}{d\log_{10}\lambda} \propto \begin{cases} 0 & \text{for } \lambda > \lambda_{\max}\,, \\ 1/\log_{10}(\lambda_{\max}/\lambda_{\min}) & \text{for } \lambda_{\min} < \lambda < \lambda_{\max}\,, \\ 0 & \text{for } \lambda < \lambda_{\min}\,. \end{cases} \tag{6.5}$$

Henceforth we refer to this distribution as a scale-invariant distribution. In this section we also restrict attention to the quark sector. Note that CP-violating phases are not introduced in this landscape because all of the matrix elements are real-valued. A pair of 3×3 Yukawa matrices is generated by choosing each of the 18 matrix elements randomly according to the distribution Eq. (6.5). Just like in the Standard Model, the quark masses and mixings are calculated from the eigenvalues of the Yukawa matrices and the unitary transformations that diagonalize them. This process is then repeated to generate an ensemble of these nine observables.

6.2.1 The Distribution of Mass Eigenvalues

Let us first study the probability distributions of the quark masses, which are just the eigenvalues of the Yukawa matrices. Since the up-type and down-type Yukawa matrices are generated indepen-

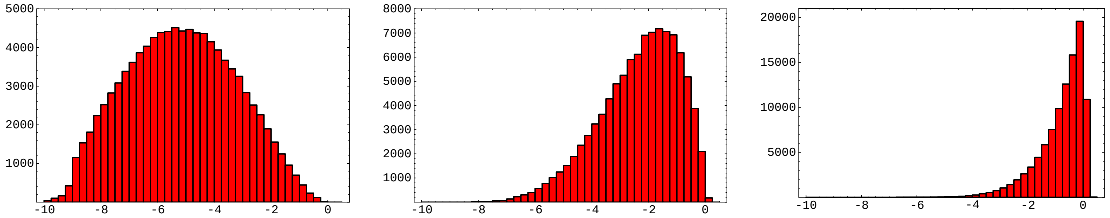


Figure 6.1: Distributions of the three eigenvalues of Yukawa matrices whose each element follows the distribution Eq. (6.5). From left to right the three panels correspond to the smallest, middle and largest eigenvalues. The sum of all three distributions reproduces Figure 9a of [189]. We used $\log_{10}\lambda_{\rm min} = -9$ and $\log_{10}\lambda_{\rm max} = 0.1$ for this simulation.

dently of each other and in exactly the same way, the distributions of eigenvalues for the up sector and those of the down sector are exactly the same. Thus we only need to study one of the two sectors. Results of a numerical study are shown in Figure 6.1, where we have chosen $\lambda_{\rm min}/\lambda_{\rm max} = 10^{-9.1}$ and have generated an ensemble of 10^5 3×3 Yukawa matrices.

With some approximations we can understand the shapes of the distributions in Figure 6.1. Let λ_3' denote the largest element of the 3×3 matrix λ. Meanwhile, the largest element of the 2×2 sub-matrix of λ that excludes λ_3' is denoted λ_2'. For example, in the matrix

$$\lambda = \left(\begin{array}{c|cc} \lambda_{23} & \lambda_2' & \lambda_{21} \\ \lambda_{13} & \lambda_{12} & \lambda_1' \\ \hline \lambda_3' & \lambda_{32} & \lambda_{31} \end{array}\right), \tag{6.6}$$

λ_3' is the largest among the nine entries and λ_2' the largest in the upper right 2×2 sub-matrix. Given this characterization, the probability distribution for the variables $\lambda'_{1,2,3}$ and λ_{ij} is

$$\begin{aligned} dP(x'_{1,2,3}, x_{ij}) \;&=\; 36\,\Theta(x_3'-x_2')\Theta(x_3'-x_{32})\Theta(x_3'-x_{23})\Theta(x_3'-x_{31})\Theta(x_3'-x_{13})\Theta(x_2'-x_1') \\ &\quad \times\Theta(x_2'-x_{21})\Theta(x_2'-x_{12})\,dx_1'dx_2'dx_3'dx_{12}dx_{21}dx_{13}dx_{31}dx_{23}dx_{32}\,, \end{aligned} \tag{6.7}$$

where Θ is the step function. The factor of 36 comes from the nine possible locations for λ_3' times the four possible locations for λ_2'. In addition we have introduced the notation

$$x_i' \equiv \frac{\ln(\lambda_i'/\lambda_{\rm min})}{\ln(\lambda_{\rm max}/\lambda_{\rm min})}\,, \qquad x_{ij} \equiv \frac{\ln(\lambda_{ij}/\lambda_{\rm min})}{\ln(\lambda_{\rm max}/\lambda_{\rm min})}\,. \tag{6.8}$$

The largest eigenvalue of Eq. (6.6) is approximately λ_3'; this approximation is poor if one of λ_{32}, λ_{31}, λ_{23} and λ_{13} is almost as large λ_3', but this is unlikely if $\ln(\lambda_{\rm max}/\lambda_{\rm min})$ is large. We call this largest eigenvalue λ_3 and define x_3 analogously to x_3'. The probability distribution of x_3 is therefore

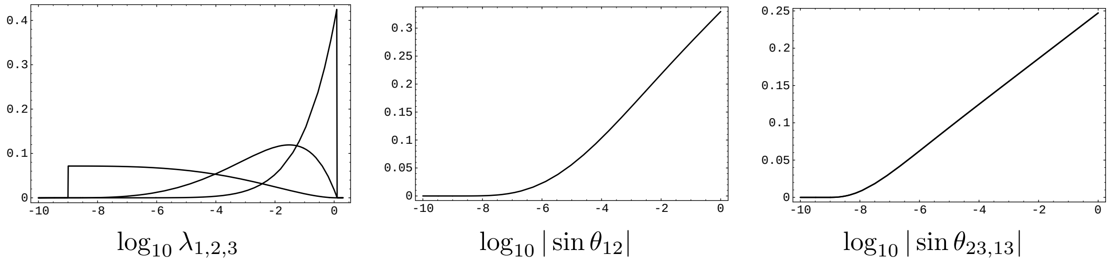


Figure 6.2: The approximate distribution of the three eigenvalues given in Eqs. (6.9–6.11), and those of mixing angles given in Eqs. (6.16, 6.17). We have used the same $\lambda_{\rm max}$ and $\lambda_{\rm min}$ as are used in Figure 6.1.

approximated by integrating out from Eq. (6.7) all the variables except x'_3:

$$dP(x_3) \simeq 9x_3^8\,dx_3\,. \tag{6.9}$$

Meanwhile, we approximate the middle eigenvalue λ_2 by λ'_2 and the smallest eigenvalue λ_1 by λ'_1. This approximation is poor when the see-saw contributions $(\lambda_{i3}\lambda_{3j})/\lambda'_3$ and $(\lambda_{12}\lambda_{21})/\lambda'_2$ are larger than λ'_2 and λ'_1, respectively. Thus we do not expect this approximation to be reliable for small values of x_2 and x_1. Nevertheless, integrating out all of other variables we find

$$dP(x_2) = \frac{36}{5}x_2^3(1-x_2^5)\,dx_2, \tag{6.10}$$

$$dP(x_1) = \frac{36}{5}\left(\frac{1-x_1^3}{3}-\frac{1-x_1^8}{8}\right)dx_1\,. \tag{6.11}$$

The probability distributions Eqs. (6.9–6.11) are shown in Figure 6.2. Remarkably, they capture the gross features of the numerical results in Figure 6.1. Therefore we use these distributions to examine the qualitative aspects of the mass distributions that follow from the landscape Eq. (6.5).

The average $\langle x_i\rangle$ and the standard deviation σ_i of the three Yukawa eigenvalues (both on a logarithmic scale) can be calculated from the distributions Eqs. (6.9–6.11):

$$\langle x_3\rangle = 0.90, \quad \sigma_3 = 0.09, \quad x_3 \sim [0.81-0.99]\,, \tag{6.12}$$

$$\langle x_2\rangle = 0.72, \quad \sigma_2 = 0.16, \quad x_2 \sim [0.56-0.88]\,, \tag{6.13}$$

$$\langle x_1\rangle = 0.36, \quad \sigma_1 = 0.22, \quad x_1 \sim [0.14-0.58]\,. \tag{6.14}$$

The three eigenvalues are on average well-separated and they overlap with neighboring eigenvalues only slightly at one standard deviation. Even this slight overlap between the distributions is misleading. Recall that by definition for any particular set of Yukawa matrices we have $x_3 > x_2 > x_1$. Thus the combined distribution for these eigenvalues is not a naive product of Eqs. (6.9–6.11) but

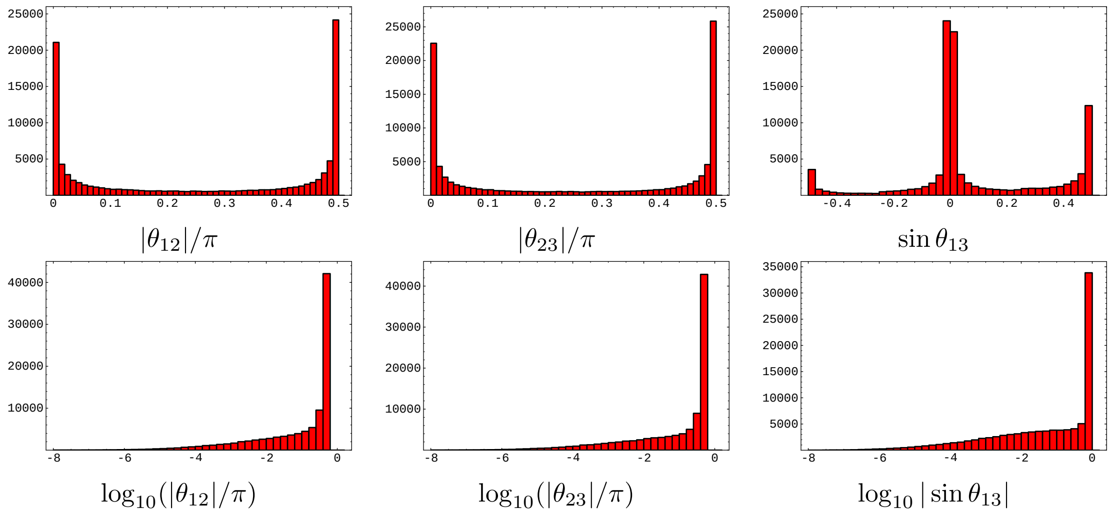


Figure 6.3: Distribution of the three mixing angles of the CKM matrix. A similar figure is found in [189], where the distributions are shown over the range $0 \leq \theta_{ij} \leq \pi/4$, $\delta = -0.16$, and presumably a different $\lambda_{\rm max}/\lambda_{\rm mix}$ is chosen.

is given by integrating the other six variables out of Eq. (6.7). This gives

$$dP(x_1, x_2, x_3) = 36\, x_3^4 x_2^2\, \Theta(x_3 - x_2)\Theta(x_2 - x_1)\, dx_1 dx_2 dx_3\,. \tag{6.15}$$

Thus it happens that in the the subset of cases where x_3 is small, the distribution of x_2 is pushed to even smaller values. Hence the three eigenvalues tend to be well-separated even in the logarithmic scale, and only rarely are adjacent eigenvalues comparable. Note that none of this depends on the choice of $\lambda_{\rm min}$ and $\lambda_{\rm max}$. In short, hierarchical structure (Yukawa eigenvalues well-separated in logarithmic scale) is generated statistically in a landscape where each matrix element independently follows the scale-invariant distribution Eq. (6.5). Whether the hierarchy is large or small is determined by whether $\log_{10}(\lambda_{\rm max}/\lambda_{\rm min})$ is large or small.

6.2.2 Pairing Structure in Electroweak Interactions

Let us now study the mixing angles. Figure 6.3 shows the distributions of mixing angles in the quark sector that result from a numerical simulation where each element of both the up-type and down-type Yukawa matrices is assumed to follow the distribution Eq. (6.5) independently.[1] The probability distribution function of the mixing angles are shown against the axes of $d\theta_{12}$, $d\theta_{23}$ and

---

[1] Using Eqs. (6.9–6.11) we find these distribution functions to be given by the approximate analytic form:

$$dP(t) \sim \frac{3}{50}(10 - 15t + 6t^4 - t^9)(5 - 8t^3 + 3t^8)\, dt\,, \qquad \text{where } t \equiv \frac{\ln|\sin\theta_{12}|}{\ln(\lambda_{\rm min}/\lambda_{\rm max})}\,, \tag{6.16}$$

$$dP(t) \sim \frac{9}{32}(8 - 9t + t^9)(1 - t^8)\, dt\,, \qquad \text{where } t \equiv \frac{\ln|\sin\theta_{23,13}|}{\ln(\lambda_{\rm min}/\lambda_{\rm max})}\,. \tag{6.17}$$

$d(\sin\theta_{13})$ because the invariant measure[2] of SO(3) mixing matrices is $d\theta_{12}\wedge d\theta_{23}\wedge d(\sin\theta_{13})$. Since in this model the Yukawa couplings are all real- and positive-valued, the CKM quark mixing matrices are SO(3) matrices.

The prominent feature of these distributions is the twin peaks at $\theta_{ij} = 0$ and $|\theta_{ij}| = \pi/2$ for all three mixing angles. This feature is straightforward to understand. Suppose that the randomly generated Yukawa matrices are of the form

$$\lambda^u_{ij} \sim \begin{pmatrix} * & \lambda'^u_2 & \bullet \\ * & \bullet & \lambda'^u_1 \\ \lambda'^u_3 & * & * \end{pmatrix}, \qquad \lambda^d_{ij} \sim \begin{pmatrix} \bullet & * & \lambda'^d_2 \\ * & \lambda'^d_3 & * \\ \lambda'^d_1 & * & \bullet \end{pmatrix}, \tag{6.18}$$

where the $*$'s are assumed to be less than λ'^u_3 or λ'^d_3, and the $\bullet$'s less than λ'^u_2 or λ'^d_2. Ignoring the see-saw contributions to eigenvalues, i.e. when $(**/\lambda'_3) \ll \lambda'_2$ and $(\bullet\bullet/\lambda'_2) \ll \lambda'_1$, we find that the three left-handed quark doublets $q_j = (u_L, d_L)_j$ $(j = 1, 2, 3)$ are approximately

$$q_1 = (t_L, d_L), \qquad q_2 = (c_L, b_L), \qquad q_3 = (u_L, s_L), \tag{6.19}$$

where t_L, c_L and u_L stand for left-handed components of the heaviest, middle and lightest mass eigenstates of up-type quarks. The down-type mass eigenstates b_L, s_L and d_L are defined similarly. The CKM matrix for the up-type and down-type Yukawa matrices Eq. (6.18) is roughly

$$V_{CKM} \sim \begin{pmatrix} & 1 & \\ & & 1 \\ 1 & & \end{pmatrix}, \tag{6.20}$$

corresponding to $\theta_{12} \sim \theta_{23} \sim \pi/2$ and $\theta_{13} \sim 0$. This explains one of the 2^3 combinations of peaks in Figure 6.3. When the $\lambda'^{u/d}_{3,2,1}$ are found in different entries of the 3×3 Yukawa matrices, other peak combinations are obtained. Continuous distributions connecting $\theta_{ij} \sim 0$ to $|\theta_{ij}| \sim \pi/2$ originate from the see-saw contributions that we have ignored. Thus the distributions become more and more localized around the peaks as $\lambda_{\max}/\lambda_{\min}$ is increased and see-saw contributions become less important. Therefore the small mixing angles of the observed CKM matrix are not atypical of distributions with large values of $\lambda_{\max}/\lambda_{\min}$.

How then do we interpret the peaks at $\pi/2$? The flavor structure of the Standard Model is characterized by two general features. On the one hand, the W-boson current connects three approximately distinct pairs of quarks that have hierarchical masses—we refer to this as "generation structure." Moreover, the W-boson current connects pairs such that the lightest up-type quark is

---

[2] The importance of the invariant measure is emphasized in Ref. [190].

approximately paired with the lightest down-type quark, the middle up-type quark is approximately paired with the middle down-type, etc. We refer to this as the “pairing structure” of the Standard Model. Mixing angles near $\pi/2$ maximally violate this pairing structure; for example the set of angles $\theta_{12} \sim \theta_{23} \sim \pi/2$ and $\theta_{13} \sim 0$ of Eq. (6.20) corresponds to the quark pairings in Eq. (6.19). There are more combinations of mixing angles (2^3) than the 3! combinations of quark pairs because some choices of mixing angles correspond to the same combination of quark pairs.

6.2.3 Problems

Is the existence of mixing angle peaks about $|\theta_{ij}| = \pi/2$ really a problem? The landscape that we have discussed so far may reproduce the pairing structure of mass eigenstates in the W-boson current (when $\theta_{12} \sim \theta_{23} \sim \theta_{13} \sim 0$), but more often it does not. Although it might be argued that this is just a $1/2^3$ coincidence problem; it is still difficult to accept that the pairing structure of flavor is not revealing something important about the underlying theory. On the other hand, even if one accepts this coincidence, in this landscape it is still difficult to understand why $V_{ub} \sim \theta_{13} \sim 4 \times 10^{-3}$ is so small compared to the Cabibbo angle θ_{12}. It does not appear that this is a shortcoming that can be solved by a more ideal choice of $\lambda_{\text{max}}/\lambda_{\text{min}}$.

An even bigger problem is to understand how the probability distribution Eq. (6.5) arises or what is the correct distribution to replace it. That is, although the phenomenology of the landscape that we have considered may be deemed acceptable, we do not have a solid theoretical ground on which to base it.[3] In the remaining sections of this chapter we analyze some landscape models that successfully reproduce the phenomenology of generation and pairing structure, while making progress on the problems described above.

6.3 A Toy Landscape: Quarks in One Extra Dimension

The landscape discussed in section 6.2 assumes that all 18 elements of λ^u and λ^d are scanned independently. Yet without any correlation between these two Yukawa matrices, pairing structure will never be obtained. For example, in order to ensure that the heaviest up-type quark t_L is contained in the same $\mathrm{SU}(2)_L$ doublet as the heaviest down-type quark b_L, we require the following. When an (i, j) element of the up-type Yukawa coupling $\lambda^u_{ij}\, u^i_R\, q^j_L\, h^*$ is large, at least one of the three down-type Yukawa couplings $\lambda^d_{kj}\, d^k_R\, q^j_L\, h$ $(k = 1, 2, 3)$ involving the same quark doublet q^j_L should be large. A landscape of vacua must realize such a correlation between the up-type and down-type

---

[3] Such an attempt is made in [191]. However, we consider that correlation among Yukawa couplings and the number of dimension of extra dimensions are crucial ingredients in understanding flavor physics, and these are missing in [191]. The intersecting $D6$–$D6$ system mentioned in [191] is dual to a T^3-fibered compactification of Heterotic string theory and is simulated by the $D = 3$ toy landscape model of this chapter. On the other hand the scale invariant distribution Eq. (6.5) is derived from a $D = 1$ toy landscape model in section 6.3 of this chapter, but not from $D = 3$ toy landscape models.

Yukawa matrices in order to explain the pairing structure.

Perhaps one of the simplest ideas to introduce such a correlation is to introduce an extra dimension. If a large Yukawa coupling of $\lambda^u_{ij}\, u^i_R\, q^j_L\, h^*$ is due to a substantial overlap of the wavefunctions of q^j_L and h^*, then the down-type Yukawa couplings involving the same q^j_L tend to be larger because of the overlap of q^j_L and h. At the same time, for highly localized wavefunctions the overlap of some triplets of u_R, q_L and h^* can be very small, and so there is hope to explain the hierarchically small Yukawa couplings necessary to account for light quarks.

In this section we present a simple toy landscape based on Gaussian wavefunctions spanning a circular extra dimension. Although a single extra dimension is introduced for simplicity, this model captures the essence of what one expects more generally from such "Gaussian landscapes" based on multiple extra-dimensional field theories. Later, in Section 6.5, we will study what are the effects of more complicated geometries of extra dimensions. Through numerical simulation and an approximate analytical analysis, we find that both the generation structure and the pairing structure of quarks are obtained statistically in this landscape. No flavor symmetry is needed.

6.3.1 Emergence of Scale-Invariant Distributions

We introduce a single extra dimension with the simplest geometry: S^1. The wavefunctions for all of the quarks and the Higgs boson are assumed to be Gaussian with a common width d. Each quark and the Higgs may have its wavefunction centered at an arbitrary point on S^1, for any one of these fields the wavefunction is

$$\varphi(y; y_0) \simeq \frac{1}{\pi^{1/4}(M_5 d)^{1/2}} e^{-\frac{(y-y_0)^2}{2d^2}} . \tag{6.21}$$

Here y is the coordinate of S^1 and M_5 is the cut-off scale of a field theory in $3+1$ dimensions. This wavefunction is normalized so that

$$M_5 \int_0^L dy\, \varphi^2(y) = 1, \tag{6.22}$$

where L is the circumference of S^1. The wavefunction Eq. (6.21) should be made periodic on S^1, while maintaining the normalization in Eq. (6.22). Yet as long as the width of the Gaussian profile d is parametrically smaller than the circumference L, the wavefunction is almost Gaussian. One should examine whether Gaussian wavefunctions arise as solutions of equations of motion of field theories in extra dimensions, but we defer this theoretical study to future work. Presently we study whether the assumption of Gaussian wavefunctions on extra dimensions leads to a successful explanation of the physics of quark masses and mixing angles.

We calculate the up-type and down-type Yukawa matrices with the overlap integrals

$$\begin{aligned}\lambda^u_{ij} &= gM_5 \int_{S^1} dy\, \varphi^{u_R}_i(y;a_i)\, \varphi^{q_L}_j(y;b_j)\, \varphi^h(y;y_0)\,, \\ \lambda^d_{kj} &= gM_5 \int_{S^1} dy\, \varphi^{d_R}_k(y;c_k)\, \varphi^{q_L}_j(y;b_j)\, \varphi^h(y;y_0)\,, \end{aligned} \tag{6.23}$$

where g is an overall constant. $\varphi^{q_L}_j(y)$, $\varphi^{u_R}_i(y)$, $\varphi^{d_R}_k(y)$ and $\varphi^h(y)$ are wavefunctions of left-handed quark doublets q_j ($j = 1, 2, 3$), right-handed up-quarks $\bar{u}_i$ ($i = 1, 2, 3$), right-handed down-quarks $\bar{d}_k$ ($k = 1, 2, 3$) and of the Higgs boson, respectively, all of the Gaussian form Eq. (6.21). The center coordinates of these Gaussian wavefunctions are b_j, a_i, c_k and y_0, respectively. The matrices $\lambda^{u,d}$ are real, so that CP is conserved in this toy landscape. We defer to future work the study of CP-violating complex Gaussian wavefunctions.

In this section, we analyze a toy landscape where the center coordinates b_j, a_i, c_k and y_0 are scanned freely and independently from one another on S^1. Because of the translational symmetry of S^1, only the relative difference between these center coordinates affects observables. Thus the effective number of scanning parameters is $n_S = 9$. On the other hand, there are nine observables determined from the Yukawa matrices in the quark sector: 3×2 mass eigenvalues and three mixing angles. Thus the scanning parameters cover the space of observables and no precise prediction among the observables is available. However, since this Gaussian landscape covers the space of observables, our vacuum is unlikely to be missed in this ensemble.

The other model parameters, namely the width d, circumference L, cut-off scale M_5 and overall coupling g, are treated as fixed. This treatment must appear quite arbitrary, and it really is. Among the myriad of other possibilities are to scan some or all of these parameters, to allow the up-sector and the down-sector to have different values of g or to allow their couplings to be scanned independently, to choose a different width d for different wavefunctions, etc. An extreme version of the landscape would allow everything to scan, leaving no fixed parameters to be input by hand. We adopt the assumption above—namely four fixed parameters—because of a practical attitude. Nothing is known about the distribution of these parameters within string theory, let alone what are the appropriate weighting factors coming from cosmological evolution and environmental selection effects. If the distribution of these parameters is sharply peaked around a certain point in the four-dimensional parameter space, then these parameters can be treated as effectively fixed. Such a possibility is not implausible because some toy landscapes predict Gaussian distributions for some parameters, and moreover, cosmological evolution may yield exponentially steep weight factors.

Note that out of these four fixed parameters there are only two independent combinations that affect the Yukawa matrices. First, the coordinate of the extra dimension can be made dimensionless by defining $\tilde{y} \equiv M_5 y$. Now the circumference is $M_5 L$ and the width of the wavefunction is $M_5 d$. Second, the new coordinate $\tilde{y}$ can be rescaled: $\tilde{y}' \equiv c\tilde{y}$. Under this redefinition of the coordinate, both M_5L and M_5d scale, but the ratio L/d does not. Thus two fixed parameters, g and L/d, control the Yukawa couplings of this toy landscape. (Note that L and M_5 do affect the low-energy values of gauge couplings and Newton's constant.)

The Yukawa couplings, given by the overlap integrals Eq. (6.23), can be expressed more explicitly in terms of the underlying parameters in a restricted region of the parameter space. Suppose that $L \gg d$. Then the compactness of S^1 is not important in the calculation of the Yukawa couplings, as long as the center coordinates of quarks, a_i (or c_k) and b_j, are close to that of the Higgs boson y_0 (which, by translational invariance, we set as the origin of the coordinate y). For such a vacuum, the Yukawa couplings are given by

$$\lambda_{ij} \simeq \left(\frac{4}{9\pi}\right)^{\frac{1}{4}} \frac{g}{\sqrt{M_5 d}}\, e^{-\frac{1}{3d^2}(a_i^2+b_j^2-a_ib_j)}. \tag{6.24}$$

It is useful to compare this result to the form for the Yukawa couplings that results from approximate Abelian Flavor Symmetries (AFS) [38, 39]. In the most general AFS scheme there is a small symmetry breaking factor associated with each quark field, $\epsilon_i^{q_L,u_R,d_R}$, which leads to Yukawa matrix elements

$$\lambda_{ij} = g_{ij}\, \epsilon_i^q \epsilon_j^{\bar{q}}\,, \tag{6.25}$$

where the g_{ij} are all of order unity. A mass hierarchy among the generations is realized by imposing $\epsilon_3 \gg \epsilon_2 \gg \epsilon_1$ in the left, right, or both sectors. Models with fewer parameters can be constructed and then the symmetry breaking parameters are not all independent; consider for example a single Abelian symmetry with a symmetry breaking parameter ϵ that appears in different entries with different powers due to generation dependent charges. Generation charges (0,2,3) then give $\epsilon_3 : \epsilon_2 : \epsilon_1 = 1 : \epsilon^2 : \epsilon^3$. No matter how the model is arranged, the mass hierarchy arises because the first generation feels much less flavor symmetry breaking than the third. Note that AFS theories are very flexible—any hierarchical pattern of fermion masses can be described by an appropriate AFS.

We can understand the result Eq. (6.24) within this context. First notice that a_i and b_j can be both positive and negative and therefore the factor of a_ib_j in Eq. (6.24) is statistically neutral. Anticipating this statistical averaging, we cast Eq. (6.24) into the form of Eq. (6.25) with the identification

$$\epsilon_i^{\bar{q}} = e^{-\frac{a_i^2}{3d^2}}\,, \qquad \epsilon_j^{q} = e^{-\frac{b_i^2}{3d^2}}\,. \tag{6.26}$$

An important feature is that the AFS factor ϵ_j^q is shared by all elements of both the up-type and the down-type Yukawa couplings that involve the left-handed quark doublet q_L^j. This introduces a correlation between the up-type and down-type Yukawa matrices.

We first study the probability distribution for a single entry in the Yukawa matrix, ignoring correlations with other entries. This allows us to determine the the analogue of Eq. (6.5) for this toy

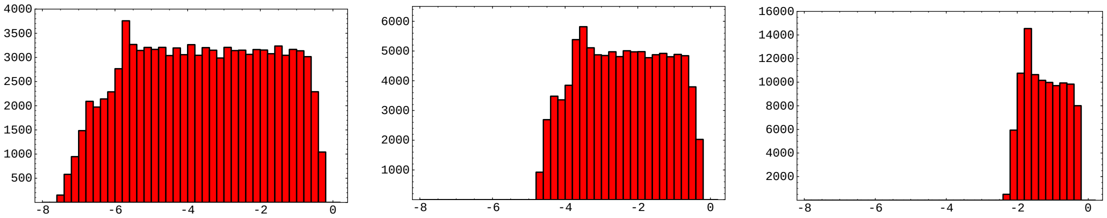


Figure 6.4: Distribution of 10^5 randomly generated Yukawa matrix elements. From left to right the panels correspond to $(d/L, g) = (0.08, 1)$, $(0.1, 1)$, and $(0.14, 1)$.

landscape. Here we do not have a distinction between the up-type and down-type Yukawa matrices because we assume that the center coordinates of both the u_R and d_R wavefunctions are distributed randomly over S^1. The Yukawa coupling Eq. (6.24) is a function on the two-dimensional parameter space $\{(a, b) \in [0, L] \times [0, L]\}$. The probability that the Yukawa coupling is larger than some value λ_0 is given by the area of an ellipse,

$$\frac{(a+b)^2}{4} + \frac{3(a-b)^2}{4} = 3d^2 \ln\left[\frac{g}{\sqrt{M_5 d}}\frac{1}{\lambda_0}\left(\frac{4}{9\pi}\right)^{\frac{1}{4}}\right]. \tag{6.27}$$

When the signs of a and b are opposite, the overlap of the three wavefunctions is small and the Yukawa coupling becomes small. This is why the region of $\lambda > \lambda_0$ is short in the $(a - b)$ axis and long in the $(a + b)$ axis. The probability that $\lambda > \lambda_0$ is given by

$$P(\lambda > \lambda_0) \simeq 2\sqrt{3}\pi\left(\frac{d}{L}\right)^2 \ln\left(\frac{g}{\sqrt{M_5 d}}\frac{1}{\lambda_0}\left(\frac{4}{9\pi}\right)^{\frac{1}{4}}\right), \tag{6.28}$$

and hence the distribution is flat:

$$\frac{dP(\lambda)}{d\ln\lambda} = 2\sqrt{3}\pi\left(\frac{d}{L}\right)^2 \simeq 11\left(\frac{d}{L}\right)^2. \tag{6.29}$$

The distribution may cease to be flat as a or b approaches $\pm L/2$, where the parameter space ends, because we ignored the periodic boundary condition in the calculation that led to Eq. (6.24). Setting this point aside we see that the probability distribution of Yukawa couplings in this toy landscape is flat on the $\ln\lambda$ axis and has an approximate span of

$$\Delta\ln\lambda = \ln\left(\frac{\lambda_{\rm max}}{\lambda_{\rm min}}\right) \simeq \frac{1}{11}\left(\frac{L}{d}\right)^2, \tag{6.30}$$

arising from the inverse of the height of the distribution function Eq. (6.29). The overall hierarchy among Yukawa couplings $\Delta\ln\lambda = \ln(\lambda_{\rm max}/\lambda_{\rm min})$ is proportional to $(L/d)^2$; the narrower the wavefunctions become, the further the wavefunctions can be separated, and the smaller Yukawa couplings

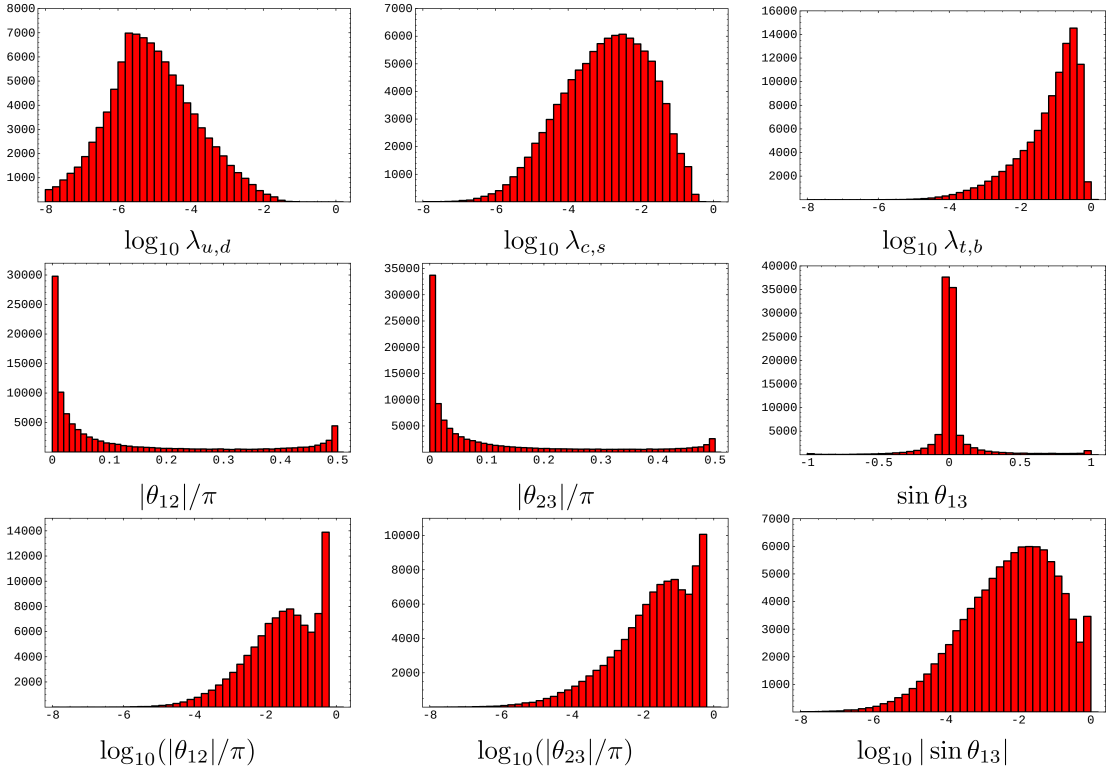


Figure 6.5: Distribution of three quark Yukawa eigenvalues and mixing angles, based on a numerical analysis with $(d/L, g) = (0.08, 1)$.

can be. As seen from Eq. (6.24), the upper end of this scale-invariant distribution $\lambda_{\max}$ is roughly $g/\sqrt{M_5 d}$. Note that the scale-invariant distribution Eq. (6.5) was introduced almost by hand in [189] in order to account for the large hierarchy among Yukawa couplings. It is interesting that this distribution is a natural prediction of our simple toy landscape.

We performed a numerical calculation to confirm the semi-analytical analysis above, taking account of the compactness of S^1 by making the wavefunction Eq. (6.21) periodic. The center coordinates of the quark wavefunctions a, b, c were generated randomly 10^5 times, and the Yukawa coupling was calculated through Eq. (6.23). This process was repeated for three different sets of the $(d/L, g)$ parameters: $(0.08, 1)$, $(0.10, 1)$ and $(0.14, 1)$. The resulting distributions, shown in Figure 6.4, are all roughly scale-invariant (flat on a logarithmic scale), with heights proportional to $(d/L)^2$, just as we expected from the semi-analytical discussion.

6.3.2 Quark-Sector Phenomenology of the Gaussian Landscape

Let us now move on to study the probability distributions of the mass eigenvalues and mixing angles. Figure 6.5 shows the result of a numerical simulation with $(d/L, g) = (0.08, 1)$. The distributions of Yukawa eigenvalues in Figure 6.5 are similar to those in Figure 6.1, but with narrower distributions

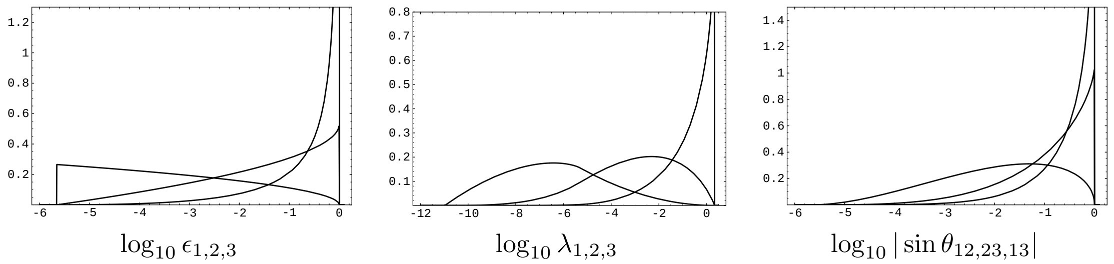


Figure 6.6: Distribution of the three AFS suppression factors, the three Yukawa eigenvalues, and the CKM mixing angles; the latter calculated naively from Eq. (6.38) and Eqs. (6.40–6.42). The CKM mixing angle θ_{23} is the one most sharply peaked at zero, while θ_{13} is most spread out. These figures correspond to $d/L = 0.08$ and thus $\Delta\log_{10}\epsilon = 5.65$. Recall that these distribution functions are not reliable for small eigenvalues.

of $\lambda_{u,d}$ and with the distributions of $\lambda_{c,s}$ shifted downward. The prominent difference between the mixing-angle distributions in Figure 6.5 and those in Figure 6.3 is the absence of the unwanted peaks at $|\theta_{ij}| \simeq \pi/2$. Thus we find the pairing structure of the quark sector follows from this toy landscape; introducing correlation between the up-type and down-type Yukawa matrix elements works perfectly. Moreover, in contrast to Figure 6.3 the distribution of θ_{13} in Figure 6.5 has a clear peak at $\theta_{13} \ll \mathcal{O}(1)$ when displayed on a logarithmic scale.

The distributions of Yukawa eigenvalues and mixing angles in Figure 6.5 can be understood analytically if we allow ourselves to make an approximation. We have seen in section 6.3.1 that both the up-type and down-type Yukawa matrices have an AFS structure. Thus, we begin by determining the probability distribution of the AFS suppression factor $\epsilon^{q,\bar{q}}$ in Eq. (6.26). The value of ϵ^{q} ($\epsilon^{\bar{q}}$) is determined by the distance $|b|$ ($|a|$, $|c|$) of the left-handed (right-handed) quark wavefunction from the Higgs boson wavefunction. Since the center coordinates are scanned randomly over the extra dimension S^1, the probability measure is

$$dP(b) = \frac{2}{L} d|b|\,, \qquad \text{for } 0 \le |b| \le \frac{L}{2}. \tag{6.31}$$

The measure for the right-handed quarks is the same, and we only deal with the left-handed quarks hereafter. Converting the variable from $|b|$ to $\ln\epsilon^{q}$ using Eq. (6.26), we find

$$dP(y) = \frac{dy}{2\sqrt{y}}\,, \qquad \text{for } 0 \le y \le 1\,, \tag{6.32}$$

where $y \equiv \ln\epsilon^{q}/\Delta\ln\epsilon$ and we have defined $\Delta\ln\epsilon \equiv -(L/d)^2/12$.

The center coordinates of the three left-handed quark wavefunctions are chosen randomly, thus three AFS suppression factors follow Eq. (6.32) independently. The smallest of these corresponds to the suppression factor ϵ_1 for the lightest quark, while the middle factor ϵ_2 and largest factor ϵ_3

correspond to the middle and the heaviest quarks. The distribution of $\epsilon_{1,2,3}$ is given by

$$dP(y_1, y_2, y_3) = \frac{3!}{2^3} \frac{dy_1\, dy_2\, dy_3}{\sqrt{y_1 y_2 y_3}}\, \Theta(y_1 - y_2)\, \Theta(y_2 - y_3)\,, \qquad \text{for } 0 \leq y_i \leq 1\,, \tag{6.33}$$

where $y_i \equiv \ln \epsilon_i / \Delta \ln \epsilon$. Note that in this notation $y_1 > y_2 > y_3$. The distributions of the individual AFS suppression factors ϵ_3, ϵ_2, and ϵ_1 are obtained by integrating Eq. (6.33) with respect to the other two variables:

$$\begin{aligned} dP(y_3) &= \frac{3}{2} \frac{(1 - \sqrt{y_3})^2}{\sqrt{y_3}}\, dy_3, & (6.34)\\ dP(y_2) &= 3\,(1 - \sqrt{y_2})\, dy_2, & (6.35)\\ dP(y_1) &= \frac{3}{2} \sqrt{y_1}\, dy_1. & (6.36) \end{aligned}$$

These distribution functions are shown in Figure 6.6. The average of each of these AFS suppression factors is given by

$$\frac{\langle \ln \epsilon_3 \rangle}{\Delta \ln \epsilon} = \langle y_3 \rangle = 0.1\,, \quad \frac{\langle \ln \epsilon_2 \rangle}{\Delta \ln \epsilon} = \langle y_2 \rangle = 0.3\,, \quad \frac{\langle \ln \epsilon_1 \rangle}{\Delta \ln \epsilon} = \langle y_1 \rangle = 0.6\,, \quad \frac{\langle \ln(\epsilon_1/\epsilon_2) \rangle}{\langle \ln(\epsilon_2/\epsilon_3) \rangle} = 1.5\,. \tag{6.37}$$

In a sense, this toy landscape predicts the ratio of the AFS charges for the three generations: 6:3:1. However the distribution functions Eqs. (6.34–6.36) contain more information.

The distributions of Yukawa eigenvalues follow from Eq. (6.33) with the approximation

$$\ln(\lambda_i / \lambda_{\max}) \sim \ln \epsilon_i^q \epsilon_i^{\bar{q}}. \tag{6.38}$$

Explicit expressions of the distribution functions derived in this way are given in Section 6.6, Eqs. (6.64–6.66) and are plotted in Figure 6.6. From Eqs. (6.64–6.66) we see that for small values of $z_i \equiv \ln(\lambda_i/\lambda_{\max})/\Delta \ln \epsilon$ (these correspond to large eigenvalues), the distribution functions behave as $\neq 0$, $\propto z_2$, and $\propto z_1^2$, which is confirmed in the numerical simulation in Figure 6.5.

On the other hand, over the range of small eigenvalues the distribution in Figure 6.6 disagrees with the numerical results in Figure 6.5. For example, the distributions based on the AFS approximation extend all the way down to $\ln(\lambda/\lambda_{\max}) \sim 2\Delta \ln \epsilon$, while the numerical results cover a logarithmic range close to Eq. (6.30). Since $\Delta \ln \epsilon \simeq -\Delta \ln \lambda$, the former ranges over twice the logarithmic scale of the latter. This discrepancy arises from the compactness of the extra dimension. That is, what really matters in the exponent of Eq. (6.24) is

$$\min \left[(a + nL)^2 + (b + mL)^2 - (a + nL)(b + mL) \right]\,, \qquad \text{for } n\,, m \in \mathbb{Z}\,. \tag{6.39}$$

As the center coordinates a and b approach $\pm L/2$, non-zero choices of n and m may become just as important as $n = m = 0$ in Eq. (6.24). Indeed, when $|a| \sim |b| \sim L/2$ integers n and m can be chosen so that the last term is negative. Thus when the compactness of the extra dimension is taken into account this term is not statistically neutral and the full expression Eq. (6.39) cannot be larger than $(L/2)^2$. This is why the distributions of Yukawa couplings and eigenvalues in Figure 6.4 and 6.5 span over $\ln(\lambda/\lambda_{\max}) \sim \Delta \ln \epsilon$.

Distribution functions of the mixing angles can also be obtained by pursuing the AFS approximation, along with the additional (crude) approximations:

$$\ln V_{us} \sim \ln\left(\max\left\{(\epsilon^q_1/\epsilon^q_2)_{\text{u-sector}},\ (\epsilon^q_1/\epsilon^q_2)_{\text{d-sector}}\right\}\right), \tag{6.40}$$

$$\ln V_{cb} \sim \ln\left(\max\left\{(\epsilon^q_2/\epsilon^q_3)_{\text{u-sector}},\ (\epsilon^q_2/\epsilon^q_3)_{\text{d-sector}}\right\}\right), \tag{6.41}$$

$$\ln V_{ub} \sim \ln\left(\max\left\{(\epsilon^q_1/\epsilon^q_3)_{\text{u-sector}},\ (\epsilon^q_1/\epsilon^q_3)_{\text{d-sector}}\right\}\right). \tag{6.42}$$

Explicit expressions are given in Section 6.6, Eqs. (6.68–6.70), and are plotted in Figure 6.6. These approximate analytic distribution functions capture qualitative features of the numerical results. Note that the distribution function of $\sin\theta_{13}$ becomes zero at $\log_{10}(\sin\theta_{13}) \sim 0$ because $\sin\theta_{13}$ can be of order unity only when all three eigenvalues are almost degenerate in either the up- or down-sector, and the probability for this to occur is small. The averages of the mixing angles in logarithm axes are ordered

$$\langle\theta_{13}\rangle < \langle\theta_{12}\rangle < \langle\theta_{23}\rangle\ , \tag{6.43}$$

both in the numerical simulation and in the analytic distributions; see Eq. (6.71). This is regarded as a consequence of the AFS charges in Eq. (6.37); indeed in Eq. (6.37) we have

$$\langle\ln(\epsilon_1/\epsilon_3)\rangle < \langle\ln(\epsilon_1/\epsilon_2)\rangle < \langle\ln(\epsilon_2/\epsilon_3)\rangle\ . \tag{6.44}$$

Whether the assignments in Eq. (6.37) are observationally acceptable or not is debatable, and we will return to this issue in section 6.4. Related discussion is also found in section 6.5.

To summarize, we see that the qualitative expectations from an AFS-type mass matrix hold true in this landscape. In particular, the similarities between the landscape generated Yukawa couplings Eq. (6.24) and those of Eq. (6.25) allow us to understand Gaussian landscapes, at least to some degree, using intuition from models of AFS. Of course, between these approaches the origin of small parameters is completely different: in the landscape they arise from small overlaps of wavefunctions well separated in the extra dimension, whereas in AFS they arise from small symmetry breaking parameters. Let us now emphasize this distinction.

A crucial general feature of all AFS models is that the mass hierarchy between generations, $m_3 \gg m_2 \gg m_1$, arises because there is a hierarchy in the amount of symmetry breaking coupled to these generations. This is true in the general Abelian case by the choice $\epsilon_3 \gg \epsilon_2 \gg \epsilon_1$. In more restricted versions having $\epsilon_i \approx \epsilon^{Q_i}$, the hierarchy is imposed by a choice of charges $Q_1 > Q_2 > Q_3$. If the flavor symmetry is non-Abelian, then there is a hierarchy of symmetry breaking, for example the rank may be broken from i to $i-1$ with strength ϵ_i. Thus AFS can in principle describe any flavor pattern, for example one heavy generation with Yukawas of order unity and two very light generations with Yukawas of order 10^{-10}. The situation with this Gaussian landscape is very different. Each Yukawa coupling is a statistical quantity, with a probability distribution that is approximately scale invariant over a range determined by a single small parameter, d/L. Relative to this range, the hierarchy of mass eigenvalues, including the typical intergenerational mass ratios and mixing angles, arises purely from statistics. Unlike with the AFS parameters ϵ_i, there is no sense in which the fundamental theory distinguishes between generations. Therefore unlike with AFS, this Gaussian landscape cannot accommodate one heavy generation and two very light generations of comparable mass. Within statistical uncertainties, the landscape determines the AFS charges.

6.3.3 Environmental Selection Effects

It is a formidable task to understand or even simply to enumerate all of the environmental effects that would be associated with a landscape scanning over the flavor parameters of the Standard Model. Furthermore, without a specific theory for the landscape it is unclear whether certain qualitative features of the flavor sector arise due to environmental selection, due to systematic features of the landscape distributions, or due to accident. Consider these examples. In the Gaussian landscape of this section, λ_1/λ_2 tends to be smaller than λ_2/λ_3. Therefore if L/d is chosen so as to explain the hierarchy λ_2/λ_3, then the relative lightness of the up and down quarks are explained. The value of L/d required to reproduce the quark masses we measure may be selected dynamically within the fundamental theory or it may be selected due to environmental effects associated with having very light up and down quarks. Likewise, the unexpected hierarchy m_t/m_b may be due to the dynamical or accidental selection of different coupling constant g's for the up and down sectors,[4] or environmental selection effects could favor a very large top quark mass as described in [100]. Selection of a heavy top increases the likelihood for an "accidental" hierarchy of the charm quark mass over the strange quark mass.

We now describe qualitative effects associated with one possibility for environmental selection, which is the selection of a large top mass to ensure the stability of our present Higgs phase [100]. Since the distributions in Figure 6.5 are rather broad, the precise form of this environmental condition is not very important and consider a simple cut on the top Yukawa:

$$\log_{10} \lambda_t \geq -0.3\,. \tag{6.45}$$

---

[4] This hierarchy may also be due to weak-scale supersymmetry with a large $\tan\beta$; however in this case one has to assume that the wavefunctions for the up-type and down-type Higgs are located at the same position in the extra dimension.

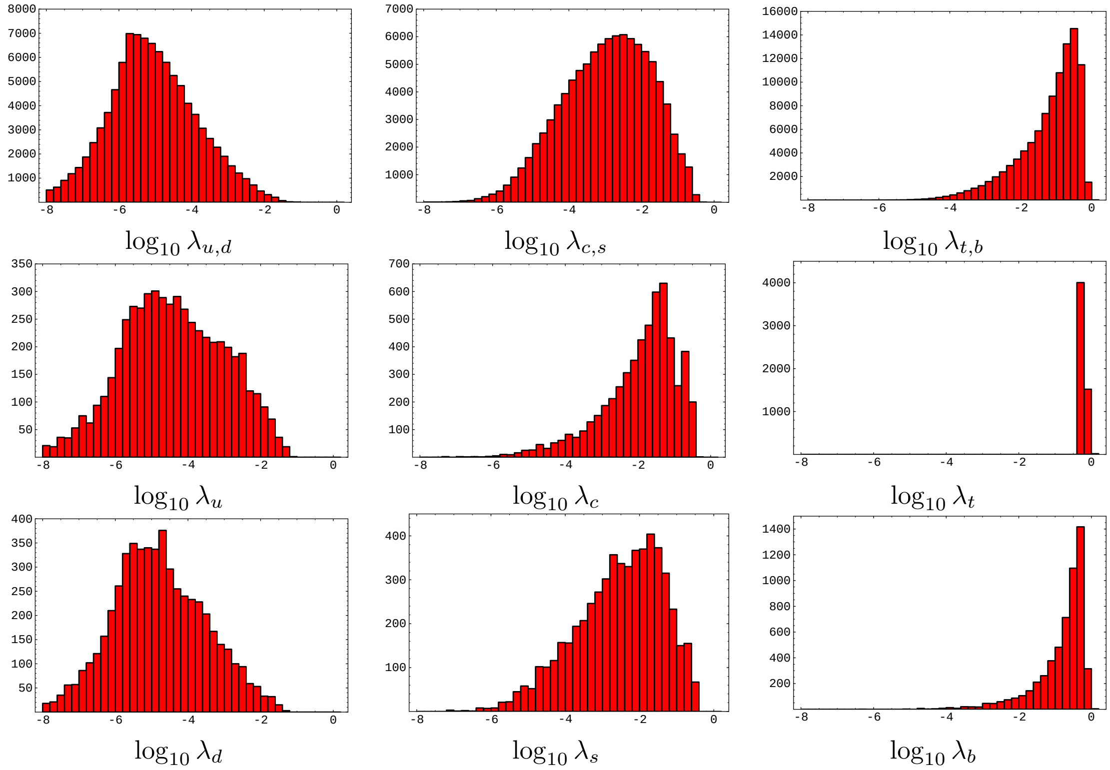


Figure 6.7: Distribution of quark Yukawa eigenvalues. The first row shows the distributions from Figure 6.5 before the t-cut is imposed. The last two rows show the distributions of the roughly 6% of Yukawa matrices that pass the t-cut.

We emphasize that we impose this t-cut to study qualitative effects; it is not intended to be precise. We first study the distributions of the quark Yukawa eigenvalues, which follow the distributions shown in Figure 6.7. The three distributions in the first row correspond to Yukawa eigenvalues before the environmental cut is imposed. Since the our toy model has not introduced any difference between the up- and down-type sectors, the three distributions are the same in both sectors. Imposing the t-cut reduces the sample size by 6%. After the t-cut is imposed, the eigenvalue distributions are modified into those displayed in the second and third rows of Figure 6.7. A notable effect of the t-cut is that the distributions of the other Yukawa eigenvalues are dragged upward. This effect is more evident in the up-sector than in the down-sector, which fits well with the observation that the charm quark is heavier than the strange quark. The effects of environmental selection are also seen in the distributions of the mixing angles; see Figure 6.8. The distributions of all three mixing angles are significantly reduced near $|\sin\theta_{ij}| \sim 1$.

6.3.4 Summary

This simple toy model addresses some of the major problems with the scheme of [189]. For one, it eliminates the peaks at $\pi/2$ in the distributions of the mixing angles and thus provides a high

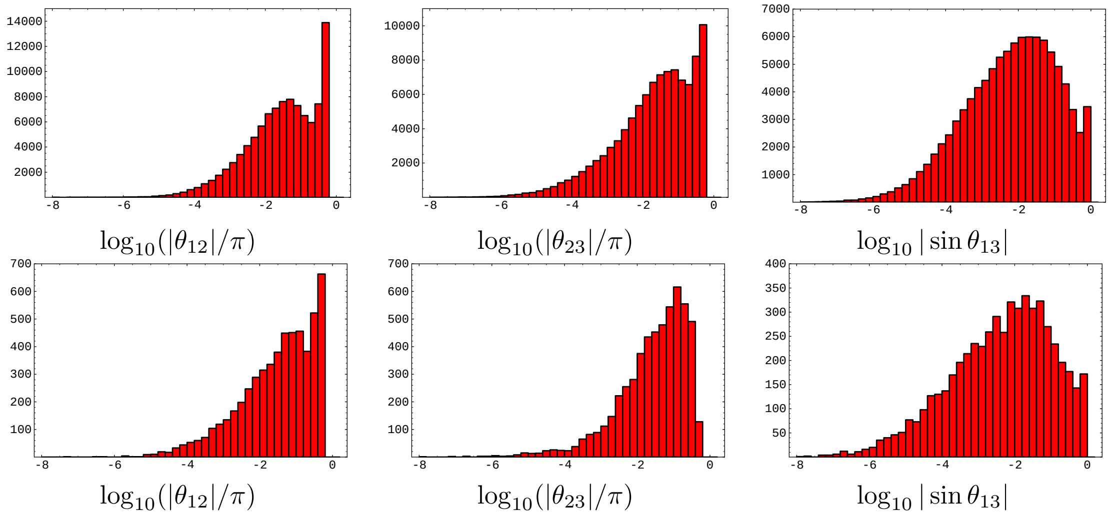


Figure 6.8: Distribution of mixing angles. The first row shows the distributions from Figure 6.5 before the t-cut is imposed. The second row shows the distributions corresponding to the roughly 6% of Yukawa matrices that pass the t-cut.

probability for the correct pairing between up-type and down-type quarks. In addition, it provides an explanation (though still at the toy-model level) for the underlying distribution of Yukawa matrix elements. Finally, it is not hard to achieve distributions that seem to fit the observed masses and mixings. Indeed, there is improvement in the fit to very small mixing angles and the observations $m_c > m_s$ and $m_t > m_b$ may be understood in the context of a selection effect. This comes from a model with only two free parameters, g and L/d.

Not every aspect of flavor structure is reproduced in this Gaussian landscape. In particular, it appears that a good fit to the top Yukawa coupling tends to predict a bottom Yukawa coupling that is too large. When the same underlying model is applied to the charged lepton sector, the tau lepton will also be predicted too heavy. We consider this problem not merely due to an un-optimized choice of parameters L/d and g. Note for example that the distribution of Yukawa eigenvalues in Figure 6.4 has a sharp cutoff at the largest allowed eigenvalue, "λ_{max}." This holds regardless of the choice of L/d and g. In addition, the distributions of the largest eigenvalues $\lambda_{b,t,\tau}$ are sharply peaked at this largest eigenvalue. This implies that a very large λ_t cannot be understood as a rare statistical upward fluctuation, even if environmental selection prefers such a fluctuation. Meanwhile, it is very unlikely that both the bottom and the tau lepton correspond to rare statistical downward fluctuations. These problem comes from the scale-invariant distribution Figure 6.4 and cannot be addressed by choosing parameters differently.

There may be a number of ways to modify the toy model of this section and thereby alleviate this problem. Some of these are presented in section 6.5, where it is discussed how the geometry of extra dimensions can affect the underlying distribution of eigenvalues in a Gaussian landscape.

Before proceeding to that discussion, we take a moment to discuss how such landscape models can be more quantitatively tested.

6.4 Testing Landscape Model Predictions

We here take an aside to discuss some possibilities for testing landscape models of family structure. The key observation of this section is that these models predict correlated distributions for many low energy parameters. In addition, for at least several of these parameters we expect dynamical and environmental selection effects to be relatively weak. This situation allows for landscape predictions to be compared to experimental measurements at higher significance than could be done for previous landscape models. Although we view this possibility as very promising, we emphasize that in this work we provide only toy models of flavor in the landscape. Thus the Gaussian landscape of section 6.3 is referred to below only for illustrative purposes. Henceforth we refer to this as the S^1 model.

The distributions for the nine quark sector flavor parameters that are predicted by the S^1 model are displayed in Figure 6.5. The simplest evaluation of how well these distributions fit observation is to simply compare the measured value of each parameter to its predicted distribution. On the other hand, a key feature of these distributions is the existence of correlations between the various flavor parameters. Thus it is possible for each flavor parameter to be reasonably typical of its distribution, yet for the complete set of flavor parameters to be atypical among the predicted distribution of complete sets. A simple way to explore this feature is to study specific relationships among the flavor parameters and compare to the predicted distributions for these relationships. For example, we may select the ratios λ_c/λ_b and λ_u/λ_s to characterize the size of the hierarchies between the generations, and select the ratios λ_b/λ_t, λ_s/λ_c, and λ_d/λ_u to characterize the discrepancies between masses within each generation. Rather than present the profile of each of these distributions, in Figure 6.9 we display the median of each distribution with error bars representing the 16th and 84th percentile values, alongside the values determined by running experimentally measured quantities (including their uncertainties) up to the Planck scale.

Figure 6.9 reveals that although neither λ_t nor λ_b are very atypical of their distributions, the ratio λ_b/λ_t is over two standard deviations below the median of its distribution. This discrepancy is reduced but not eliminated if we speculate that environmental selection strongly prefers a relatively large top quark mass, as suggested in Ref. [100] and implemented by the t-cut of section 6.3. The results for this scenario are displayed in the right panel of Figure 6.9. It is interesting that the measured value of λ_s/λ_c is also low relative the median of its distribution. This suggests a sub-pattern to flavor hierarchy that might be addressed for example by allowing different couplings g apply to the up and down sectors.

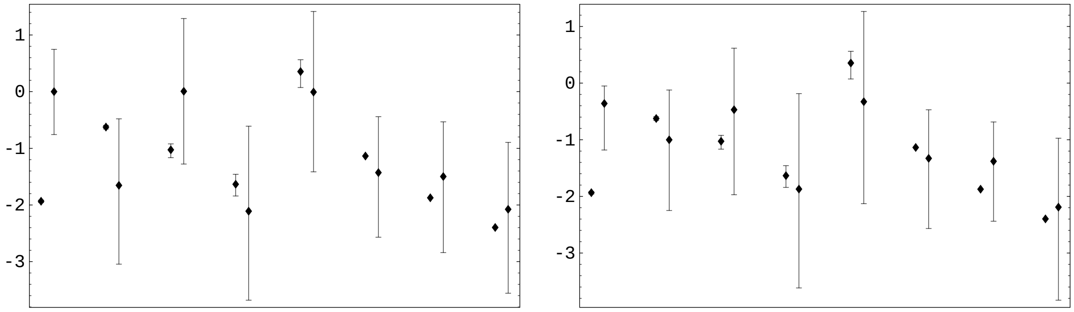


Figure 6.9: Typical values of several flavor parameters. From left to right the pairs of error bars correspond to λ_b/λ_t, λ_c/λ_b, λ_s/λ_c, λ_u/λ_s, λ_d/λ_u, θ_{12}/π, θ_{23}/π, and $\sin\theta_{13}$; the vertical scale is logarithmic. The figure to the left corresponds to no selection cuts whereas the figure to the right corresponds to flavor parameters that survive the t-cut of section 6.3. For each pair of error bars the first corresponds to the measured value and the second to the middle 68% of the Gaussian landscape of section 6.3.

6.4.1 The Chi-Square Statistic

We now pursue more quantitative analyses. One example is to use the so-called chi-square statistic as a measure of goodness-of-fit between the observed flavor parameters and their hypothetical landscape distributions. For concreteness consider again the S^1 model of section 6.3. It is convenient to denote the relevant set of quark flavor parameters by

$$\{x_i\} \equiv \log_{10}\{\lambda_u,\, \lambda_c,\, \lambda_t,\, \lambda_d,\, \lambda_s,\, \lambda_d,\, |\theta_{12}|/\pi,\, |\theta_{23}|/\pi,\, |\sin\theta_{13}|\}\,, \tag{6.46}$$

with the logarithm acting on every element of the list. The chi-square for this list is then

$$\chi^2 = \sum_{i,\,j} \left(\hat{x}_i - \langle x_j\rangle\right) \left(\sigma^{-1}\right)_{ij} \left(\hat{x}_j - \langle x_j\rangle\right), \tag{6.47}$$

where hats denote the measured value of a parameter, brackets denote the landscape average, and the co-variance matrix σ is given by

$$\sigma_{ij} = \langle x_i x_j\rangle - \langle x_i\rangle \langle x_j\rangle\ . \tag{6.48}$$

If the distributions describing each element of the list $\{x_i\}$ were Gaussian, then the quantity χ^2 would follow the chi-square distribution. This distribution is a function of the number of independent random variables that enter the quantity χ^2: if χ^2 is determined by N random variables then the chi-square distribution has a mean of N and a variance of $2N$. Our S^1 model relies on nine random variables (the a_i, b_i, and c_i) but contains two free parameters (g and L/d) that can essentially be tuned to remove two of these variables.[5] Therefore if the distributions describing the $\{x_i\}$ of

---

[5] In fact the free parameters L/d and g have not been fully exploited, since we have not performed a maximum

Eq. (6.46) were Gaussian, we would expect the χ^2 represented by Eq. (6.47) to be approximately $\chi^2 \approx 7 \pm 4$. Values of χ^2 well above this range would imply that, after accounting for correlations in the underlying distributions, too many of the measured flavor parameters are too far from the means of their distributions for it to be likely that these parameters were derived from such distributions. Meanwhile, values of χ^2 well below this range would indicate that the measured flavor parameters do not exhibit the variance about their means that would be expected from the underlying distributions.

In fact the S^1 model does not predict Gaussian distributions for any of the flavor parameters. However, the statistical interpretation of the chi-square statistic is so useful that it may be worthwhile to use Eq. (6.47) as a rough indicator of goodness-of-fit. The closer to a Gaussian profile are the underlying distributions, the more precise becomes this interpretation. This is our motivation to choose as the $\{x_i\}$ basis the logarithms of flavor parameters. With this in mind, we find $\chi^2 \approx 5$ for the S^1 Gaussian landscape, with about half of the contributions to χ^2 coming from the anomalously large measured value of the top quark Yukawa. It is tempting to use the quantity χ^2 to compare our S^1 model to the model of Ref. [189], or to models presented later in this chapter. However we must keep in mind the appropriate interpretation of the chi-square statistic, which is a measure of how well a data set is described by a set of distributions for that data. Thus it is not appropriate to use χ^2 to compare one theory to another; but instead to compare each model to the data, independent of any other models. In the above example, $\chi^2 \approx 5$ means that overall the observed set of flavor parameters is typical of those predicted by the S^1; that half of χ^2 comes from the top indicates that most of the observed flavor parameters do not exhibit quite the variance about the means of their distributions that would be expected from the S^1 model, with the unusually high mass of the top "making up" for this.

6.4.2 The P-Value Statistic

Although the chi-square statistic has a familiar interpretation, this interpretation is only precise when the underlying distributions are Gaussian. For non-Gaussian distributions we run into the problem that the quantity χ^2 of Eq. (6.47) depends on how we represent the measured quantities $\{x_i\}$. For example, it depends on whether we choose the first six $\{x_i\}$ to be the quark Yukawas or their logarithms, or whether we replace some of these parameters with ratios of quark Yukawas. This ambiguity stems from the distributions of Figure 6.5 being non-Gaussian.

This particular ambiguity is not present in the p-value statistic, which can be applied to a correlated set of distributions with any profile. The p-value of a single flavor parameter x_i is simply the fraction of the distribution that is more atypical than the measured value $\hat{x}_i$. For example we may consider the most typical value of a distribution with a single peak to be the mode (the most

---

likelihood analysis to determine the values that minimize χ^2 in Eq. (6.47). However, the parameter values we use are chosen to qualitatively fit our expectations for the measured values of $\{x_i\}$.

frequent value) in the peak, which we denote $\overline{x}$. Then the p-value is the fraction of the distribution that is farther from $\overline{x}_i$ than the observed value $\hat{x}_i$. More precisely, if we approximate the distribution as continuous with profile $P(x_i)$, then the p-value p_i is

$$p_i = \int_0^{\hat{x}_i} 2P(x_i)\,dx_i \quad \text{for } \hat{x}_i < \overline{x}_i\,, \qquad p_i = \int_{\hat{x}_i}^{\infty} 2P(x_i)\,dx_i \quad \text{for } \hat{x}_i > \overline{x}_i\,. \tag{6.49}$$

Note that like χ^2 the p-value is a random variable. Whereas χ^2 is distributed according to the chi-square distribution, the p-value is distributed uniformly between zero and one.

The p-value itself suffers from ambiguity, but this is simply the ambiguity in how one defines what is typical. For example, one may choose the most typical value to be the mode, the median, or define the p-value in a more complicated way to handle the case of more than one peak in a distribution. The purpose of the p-value is to generate a uniform distribution between zero and one, such that values very close to zero or very close to one are interpreted as unlikely due to their unusual distance or proximity to what is viewed the most typical value(s). The definition in Eq. (6.49) is appropriate for the quark Yukawa distributions of Figure 6.5. For the mixing angles of Figure 6.5 it is reasonable to take zero as the most typical value and define the p-value to be the fraction of mixing angles that is greater than any measured mixing angle.

Above we have described the p-value for a single quantity x_i relative its hypothetical distribution. We seek the p-value for a set of quantities $\{x_i\}$ relative their distributions. Consider first the product of p-values each calculated as described above, $k \equiv \prod_i p_i$. In the case where the quantities x_i are independent of each other, k represents the fraction of the distribution of $\{x_i\}$ for which each element is more atypical than the measured value. The quantity k is indeed a measure of how typical is the measured set $\{\hat{x}_i\}$; however it is not uniformly distributed between zero and one. A quantity uniformly distributed between zero and one is the fraction of possible sets $\{x_i\}$ for which $k \leq \hat{k}$, where $\hat{k}$ denotes the value of k coming from the measured set $\{\hat{x}_i\}$. We take this fraction to be the p-value, which gives

$$p = \int^{k \leq \hat{k}} \prod_i dp_i = \hat{k} \prod_{a=0}^{N-1} \frac{1}{a!} \left(-\ln \hat{k} \right)^a . \tag{6.50}$$

Here a is simply an index used to simplify the last expression, and N represents the number of random variables in $\{x_i\}$. For the S^1 model of section 6.3 $N = 7$ and we find $p = 0.53$.

The result Eq. (6.50) does not include possible correlations between the parameters in $\{x_i\}$. These can be included by changing the way in which k is defined. As already mentioned, the quantity k represents the fraction of $\{x_i\}$ that is more atypical than the measured set $\{\hat{x}_i\}$. Whereas this fraction is equal to $\prod_i p_i$ when the x_i are uncorrelated, for correlated x_i one simply finds the fraction $\{x_i\}$ for which each parameter is more atypical than the measured value in $\{\hat{x}_i\}$. The primary drawback to this method is that for data sets with large N it requires precise knowledge of the underlying distributions. For example, we required 300,000 random sets of flavor parameters to find 29 that were more atypical than the observed set. This corresponds to the reasonably typical p-value $p = 0.19$.

6.5 Geometry Dependence

In section 6.3 we introduced and analyzed a toy landscape based on a single extra dimension, and we found that it could provide both the generation structure and the pairing structure of observed quark sector. However, the specific geometry that we studied is unlikely to derive from superstring theory, where gauge fields and fermions of the low-energy effective theory may have as many as six extra dimensions in which to propagate. In addition, the compactification geometry that results in the Standard Model gauge groups may not be unique. Therefore, we initiate a study into how the compactification geometry of extra dimensions affects the probability distributions of observables, while also identifying some aspects of these distributions that do not depend on the choice of geometry. We find that the qualitative results of section 6.3 can be achieved by Gaussian landscapes in other geometries of extra dimension(s).

6.5.1 A Gaussian Landscape on T^2

To see how robust is the success of the Gaussian landscape on S^1, we study the toy model on another geometry. We first look at the simplest extension to more than one dimension, $T^2 = S^1 \times S^1$, and focus on a "square torus" where the two periods of the torus are both L. We assume that each of the quarks and the Higgs have rotation-symmetric Gaussian wavefunctions of the form

$$\varphi(\vec{y};\vec{y}_0) \propto e^{-\frac{|\vec{y}-\vec{y}_0|^2}{2d^2}}, \tag{6.51}$$

where the center coordinates $\vec{y}_0$ of each particle are randomly scanned over the internal space T^2. The up-type and down-type Yukawa matrices are calculated by the overlap integration Eq. (6.23), which is naturally generalized to integration on T^2. We defer to future work an investigation into how reasonable are these assumptions in a dynamical field theory on an extra-dimensional space-time, and for the moment focus on the phenomenology of this toy landscape.

Figure 6.10 shows the distribution of Yukawa matrix elements for a numerical simulation of this model. As in the S^1 model, the only parameters relevant to these distributions are g and L/d. In addition, larger hierarchy is generated when the wavefunctions are more localized, i.e. when L/d is larger. The key difference from S^1 is that the distribution of Yukawa matrix elements is no longer scale invariant. Instead, on T^2 the probability density for the largest and the smallest matrix

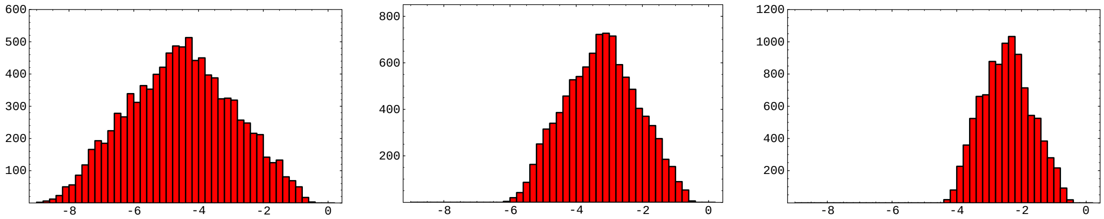


Figure 6.10: Distribution of 10^4 randomly generated Yukawa matrix elements on T^2. From left to right the panels correspond to $(d/L, g) = (0.1, 1)$, $(0.12, 1)$, and $(0.14, 1)$.

elements is depleted. With regard to the Yukawa eigenvalues and mixing angles, we find that a hierarchical pattern of Yukawa eigenvalues is generated (Figure 6.11) and that the distributions of mixing angles are peaked at $\theta_{ij} = 0$ (Figure 6.12). Thus the generation and pairing structures of the quark sector follow from the Gaussian landscape on T^2. Despite the apparent difference between the distribution of the Yukawa matrix elements of the two toy landscapes (e.g. Figure 6.4 and Figure 6.10), we see that the distribution of masses and mixing angles are roughly the same when we compare Figures 6.11 and 6.12 to Figures 6.7 and 6.8. In particular, these distributions all come with a width of about an order of magnitude, and the differences between the distributions from the two toy landscapes is not statistically significant compared to this width. This demonstrates that generation structure and pairing structure in the quark sector are robust features of Gaussian landscapes.

The biggest difference between the distribution of observables in the two toy landscapes is the right-hand-side tail of the distributions of $\log_{10} \lambda_{b,t}$ and $\log_{10} \lambda_{s,c}$. These tails are more significant in the Gaussian landscape on T^2 than in the toy landscape on S^1. This is a consequence of the difference in the distribution of individual Yukawa matrix elements, Figure 6.10 vs. Figure 6.4. The depleted probability density of the largest Yukawa matrix elements in the Gaussian landscape on T^2 results in a softer cut-off of the distributions for the largest values of the largest Yukawa eigenvalues. This applies to the heaviest and the middle quarks.

The effects of the possible environmental selection for a large top Yukawa coupling are studied in Figure 6.11 and in Figure 6.12. The distributions of the mixing angles θ_{23} and θ_{13} are shifted toward smaller angles, just as on S^1. Since a cut in favor of a large top Yukawa coupling is in favor of a larger hierarchy between the lighter quarks and the heaviest quarks, smaller θ_{23} and θ_{13} are natural consequences. The inequality $\langle \lambda_c \rangle > \langle \lambda_s \rangle$ also follows from the cut, just like in section 6.3. The most important difference between the two landscapes may be in the distribution of the bottom Yukawa coupling after the cut is imposed. It was rare that λ_b be less than 10^{-2} in the lower-right distribution of Figure 6.7, but a significant fraction this distribution is below 10^{-2} in Figure 6.11. Therefore the observed hierarchy λ_t / λ_b may be understood within the context of a landscape with a large-top-Yukawa environmental selection, if the background geometry is chosen properly.

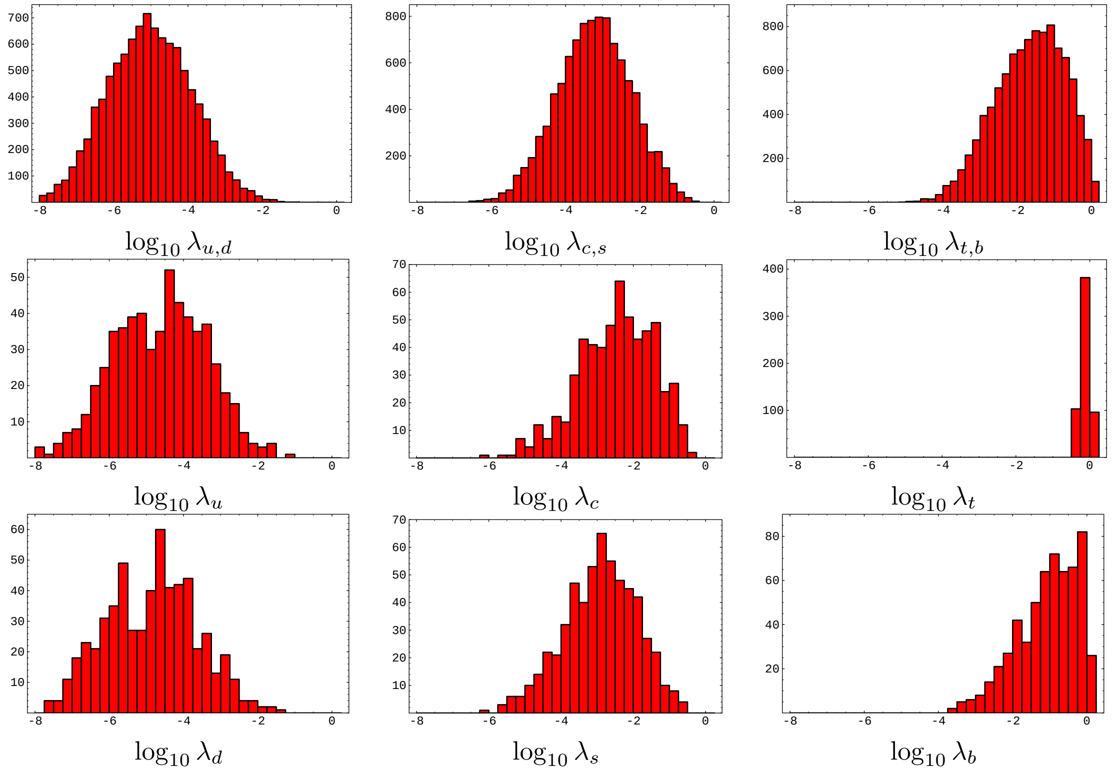


Figure 6.11: Distribution of Yukawa eigenvalues in the Gaussian landscape on T^2, based on a numerical simulation with $(d/L, g) = (0.1, 5)$. The first row shows the distribution of the three eigenvalues of the up (and down) sector. The second and third rows display the eigenvalues of the roughly 6% of matrices that survive the *t*-cut of section 6.3.3.

6.5.2 Changing the Number of Dimensions

The studies in sections 6.3 and 6.5.1 demonstrate that generation structure and pairing structure are obtained in the quark sector, once we assume a landscape based on localized wavefunctions on extra dimensions. For the examples studied, differences in the distributions of observables were not significant compared to the statistical widths of these distributions. However, looking more carefully at some details of the distributions we found some interesting geometry dependence. Therefore it will be that some geometries fit the observed flavor parameters better than others. The problem of flavor may be changed from the historical approach of seeking a specific theory to predict the precise values of flavor observables, to a new approach based on understanding the various compactification geometries and what flavor parameter distributions they imply.

Within this context we would like study Gaussian landscapes based on manifolds in any number of dimensions. However in practice the integration time grows very large as we move to manifolds with more dimensions. Operating with *Mathematica* on desktop computers, we find it impractical to study toy landscapes on manifolds with three or more dimensions.[6] On the other hand, in

---

[6] We are referring to the number of dimensions in which gauge fields are free to propagate and fermion wavefunctions

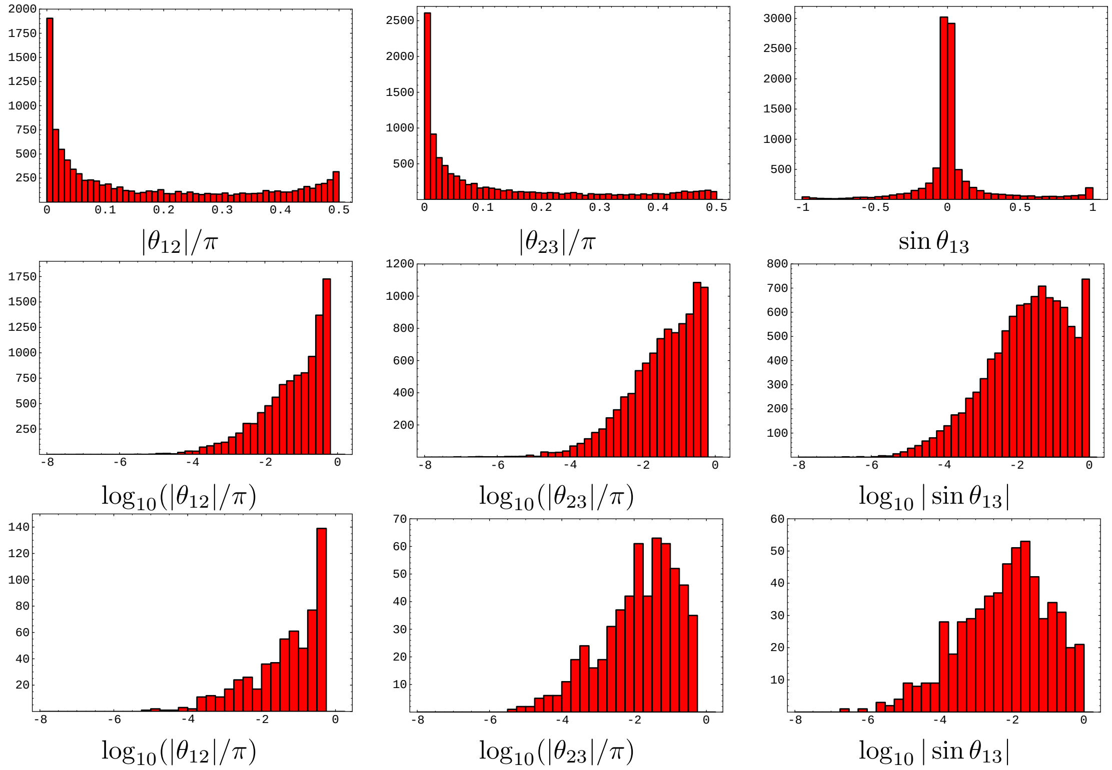


Figure 6.12: Distribution of mixing angles in the Gaussian landscape on T^2, based on a numerical simulation with $(d/L, g) = (0.1, 5)$. The bottom row displays the mixing angles of the roughly 6% of matrices that survive the t-cut of section 6.3.3.

sections 6.3.1 and 6.3.2 we made significant progress by studying flavor parameter distributions in the limit where the quark wavefunctions were more localized and near the Higgs wavefunction, such that the compactness of the extra dimension was not important. We now generalize that analysis to geometries in more than one extra dimension.

Let us consider a D-dimensional internal space with local coordinates $\vec{y}$. Then in the limit where we can ignore the size of the internal space the Yukawa coupling matrix is given by

$$\lambda_{ij} \propto e^{-\frac{1}{3d}(|\vec{a}_i|^2+|\vec{b}_j|^2-\vec{a}_i\cdot\vec{b}_j)}, \tag{6.52}$$

which replaces Eq. (6.24) for the Gaussian landscape on S^1. Here $\vec{a}_i$ and $\vec{b}_j$ are the central coordinates of the quark wavefunctions, relative to those of the Higgs. The Yukawa matrix again has the AFS form Eq. (6.25), with suppression factors

$$\epsilon^q_j = e^{-\frac{|\vec{b}_j|^2}{3d^2}}\,, \qquad \epsilon^{\bar{q}}_i = e^{-\frac{|\vec{a}_i|^2}{3d^2}}\,, \tag{6.53}$$

---

are localized. This number of dimensions can in principle be different than the number of dimensions in a compactified manifold or a stack of D-branes that realize the gauge groups of the Standard Model in string theory.

replacing Eq. (6.26). In the extra factor $g_{ij} = e^{\vec{a}_i \cdot \vec{b}_j / 3d^2}$ (which is not necessarily of order unity) the vectors $\vec{a}_i$ and $\vec{b}_j$ are sometimes unaligned, sometimes parallel, and sometimes anti-parallel. This generates a random coefficient to each Yukawa coupling that is statistically neutral in the AFS approximation. We note again that this analysis is valid only when the compactness of the internal space is unimportant and the local geometry can be approximated as a flat D-dimensional space. This is equivalent to focusing on only largest Yukawa matrix elements.

Ignoring the statistically neutral factor $e^{\vec{a}_i \cdot \vec{b}_j / 3d^2}$, the Yukawa matrix elements are roughly

$$\lambda \sim e^{-\frac{|\vec{\mathbf{r}}|^2}{3d^2}} , \qquad - \ln \lambda = \frac{|\vec{\mathbf{r}}|^2}{3d^2} , \tag{6.54}$$

with $\vec{\mathbf{r}} = (\vec{a}, \vec{b})$ scanning a $2D$-dimensional space. The natural probability measure is

$$dP \propto \frac{d^{2D} \vec{\mathbf{r}}}{L^{2D}} \propto \frac{|\vec{\mathbf{r}}|^{2D-2} \, d|\vec{\mathbf{r}}|^2}{L^{2D}} \propto \left(\frac{d}{L} \right)^{2D} (- \ln \lambda)^{D-1} d |\ln \lambda| , \tag{6.55}$$

where L is the typical size of the extra dimensions. Indeed, the distribution of the Gaussian landscape on S^1 $(D-1=0)$ is flat, see for example Figure 6.4, and that of the Gaussian landscape on T^2 $(D-1=1)$ is linear in $|\ln \lambda|$ at its upper end. Both of these results are independent of the value of L/d. The logarithmic range of the distribution of Yukawa couplings scales as $(L/d)^2$.

The distribution of the AFS suppression factors, ϵ_j^q, can be obtained as in section 6.3. In more than one dimension the measure Eq. (6.31) is generalized to

$$dP(b) \sim \frac{d^D b}{L^D} \sim L^{2-2D} \frac{dV(b)}{db^2} \frac{db^2}{L^2} , \tag{6.56}$$

where $b \equiv |\vec{b}|$ and $V(b)$ is a volume enclosed within a distance b from a given point. Analogous to the analysis in section 6.3, we represent this distribution

$$dP(y) = f(y) \, dy , \qquad f(y) \sim L^{2-2D} \frac{dV(b)}{db^2} , \tag{6.57}$$

where $y \equiv \ln \epsilon / \Delta \ln \epsilon$ with $\Delta \ln \epsilon \equiv -\frac{1}{3}(b_{\max}/d)^2$ now representing the logarithmic range of ϵ in the D-dimensional geometry. Note that what is $b_{\max}$ will depend on the geometry—for example $b_{\max} = L/\sqrt{2}$ for the "square torus" and $b_{\max} = \pi R = L/2$ for the S^2 geometry discussed below—however this only affects the distributions of small eigenvalues. In addition the density function $f(y)$ depends on the geometry of the extra dimensions, but its behavior for small y, i.e. $\epsilon \sim \mathcal{O}(1)$, depends only the number of extra dimensions: $f(y) \propto y^{D/2-1}$. The smallest, middle and largest AFS suppression factors then follow the probability distribution

$$dP(y_1, y_2, y_3) = 3! f(y_1) f(y_2) f(y_3) \Theta(y_1 - y_2) \Theta(y_2 - y_3) \, dy_1 dy_2 dy_3 , \tag{6.58}$$

where $y_i \equiv \ln\epsilon_i/\Delta\ln\epsilon$ and again $y_1 > y_2 > y_3$. Thus for small y_i,

$$dP(y_1) \propto y_1^{3D/2-1} dy_1\,, \qquad dP(y_2) \propto y_2^{D-1} dy_2\,, \qquad dP(y_3) \propto y_3^{D/2-1} dy_3\,. \tag{6.59}$$

Distribution functions of the masses and mixing angles can be obtained by using the approximations Eq. (6.38) and Eqs. (6.40–6.42). In the limit of large eigenvalues and large mixing angles these distribution functions behave as

$$\begin{aligned} &\frac{dP(z_1)}{dz_1} \propto z_1^{3D-1}\,, && \frac{dP(z_2)}{dz_2} \propto z_2^{2D-1}\,, && \frac{dP(z_3)}{dz_3} \propto z_3^{D-1}\,, \\ &\frac{dP(t_{12})}{dt_{12}} \neq 0\,, && \frac{dP(t_{23})}{dt_{23}} \neq 0\,, && \frac{dP(t_{13})}{dt_{13}} \propto t_{13}\,, \end{aligned} \tag{6.60}$$

where $z_i \equiv \ln(\lambda_i/\lambda_{\max})/\Delta\ln\epsilon$ and $t_{ij} \equiv \ln|\sin\theta_{ij}|/\Delta\ln\epsilon$. The behavior of the distribution functions on the mixing angles do not even depend on the number of dimensions.

For small Yukawa eigenvalues and mixing angles, the AFS approximation breaks down and the shape of geometry becomes important. But for the purpose of studying the zeroth order effects of the number of dimensions, we introduce a crude approximation:

$$f_D(y) = (D/2)\, y^{D/2-1} \qquad \text{for } 0 \le y \le 1\,, \qquad f_D(y) = 0 \qquad \text{otherwise.} \tag{6.61}$$

The distribution functions of the individual suppression factors ϵ_i are obtained by integrating out the other two variables from Eq. (6.58). Figure 6.13 shows these distributions for $D = 1$, $D = 2$, and $D = 3$. As the number of extra dimensions increases, distributions of Yukawa eigenvalues shift to smaller values, and the middle eigenvalue becomes closer to the smallest eigenvalue. Meanwhile, the mixing angle between the first and second generations, i.e. θ_{12}, becomes larger than the second and third generation mixing angle θ_{23}. These two phenomena are both sides of the same coin in Yukawa matrices with the AFS structure. Since we measure

$$\frac{\lambda_d}{\lambda_s} > \frac{\lambda_s}{\lambda_b}, \qquad \text{and} \qquad \theta_{12} > \theta_{23}\,, \tag{6.62}$$

this may be regarded as an indication that $D > 1$. Note however that the compactness of extra dimensions affects the distributions of the smallest eigenvalues and therefore the value of the ratio $\langle\ln(\epsilon_1/\epsilon_2)\rangle \,/\, \langle\ln(\epsilon_2/\epsilon_3)\rangle$ predicted from Eq. (6.58, 6.61) cannot reliably be used to infer the number of extra dimensions.

Although the distribution functions of the flavor parameters depend on the number of extra dimensions D, there is an important common feature that is independent of D. As we have seen, all the distribution functions (in their approximate forms) are given by power, polynomial, or logarith-

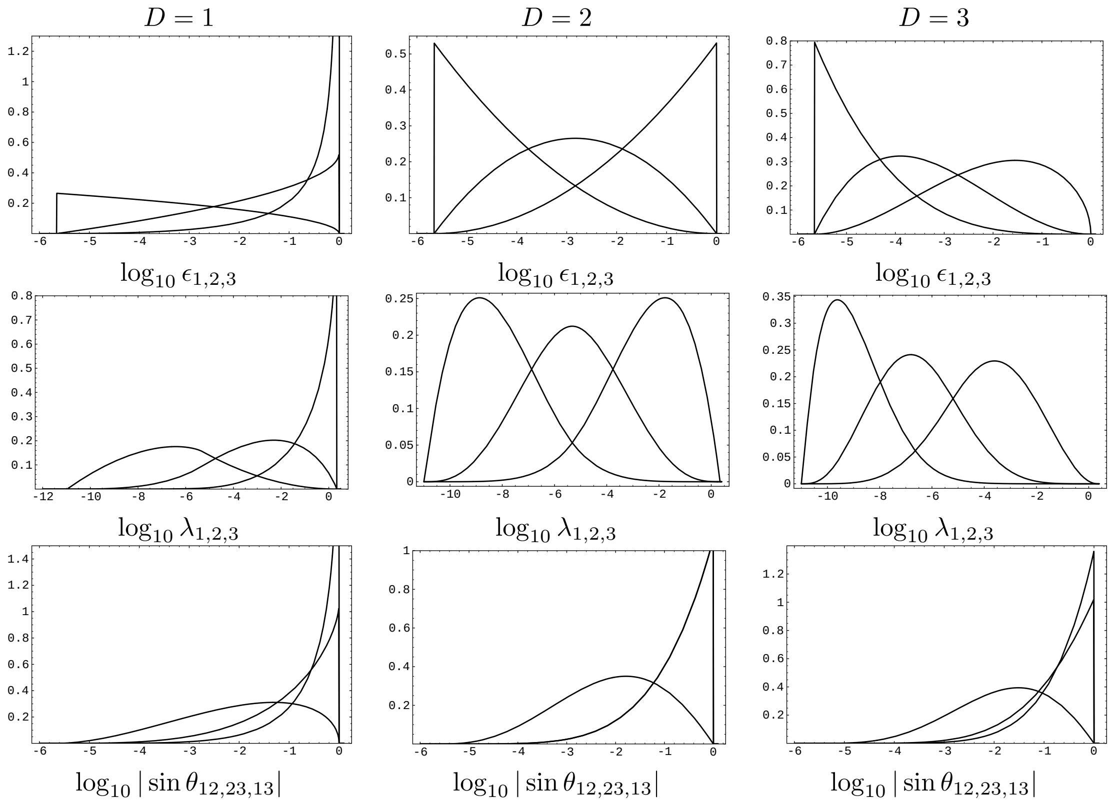
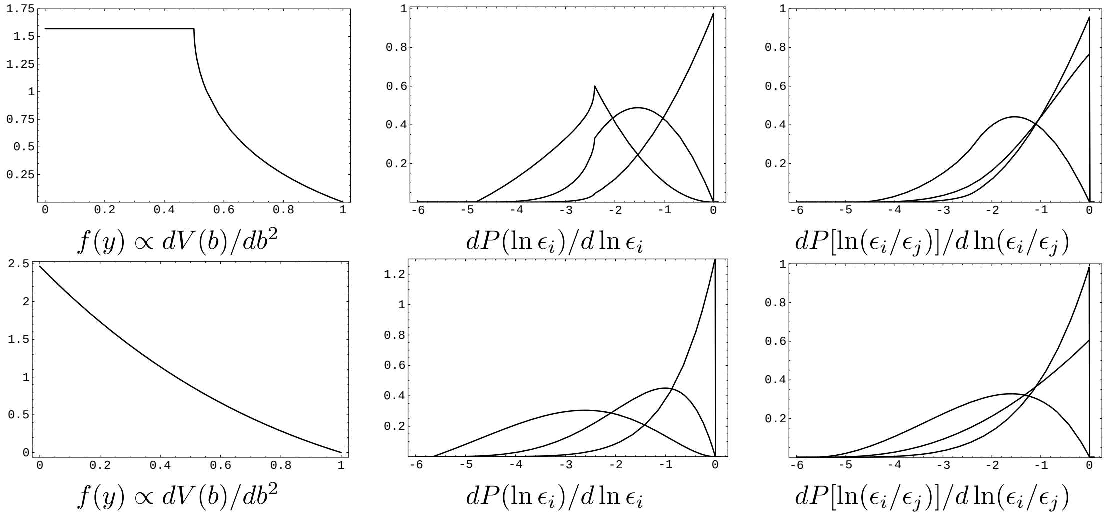


Figure 6.13: Comparing $D = 1, 2$, and 3. The first row shows distributions of $\log_{10} \epsilon_{1,2,3}$, the second row of $\log_{10} \lambda_{1,2,3}$, and the third of $\log_{10} \sin \theta_{ij}$. In all cases the broadest mixing angle distribution corresponds to θ_{13}, while the distribution most sharply peaked at zero corresponds to θ_{23} in the left panel and θ_{12} in the right panel (in the center panel the distributions of θ_{12} and θ_{23} coincide). We used approximations Eq. (6.38) and Eqs. (6.40–6.42), which are not reliable for small Yukawa eigenvalues. For clear comparison each distribution uses $d/L = 0.08$ and $b_{\max} = L/2$.

mic functions of the logarithmic variables $\log_{10} \lambda_i$ or $\log_{10} |\sin \theta_{ij}|$. Specifically, These distribution functions are not powers of λ_i or exponential functions of the logarithmic variables. Therefore, the Gaussian landscapes described in this work are different from those in [192], where all of the dimensionless coupling constants of the standard model are supposed to have narrow-width Gaussian distributions. This common feature of our toy landscapes is traced back to our assumption that the localized wavefunctions become exponentially small as one moves away from the center of localization.

6.5.3 Information Not Captured by the Number of Dimensions

Although for a given number of extra dimensions the distribution of Yukawa matrix elements has a universal form at its upper end, the full distribution depends on the the global background geometry of the extra dimensions. The density function $f(y)$ introduced in Eq. (6.56) looks like those in the first column of Figure 6.14 for the two-dimensional manifolds T^2 and S^2. Indeed, they appear quite



$f(y) \propto dV(b)/db^2$ $dP(\ln \epsilon_i)/d \ln \epsilon_i$ $dP[\ln(\epsilon_i/\epsilon_j)]/d \ln(\epsilon_i/\epsilon_j)$

$f(y) \propto dV(b)/db^2$ $dP(\ln \epsilon_i)/d \ln \epsilon_i$ $dP[\ln(\epsilon_i/\epsilon_j)]/d \ln(\epsilon_i/\epsilon_j)$

Figure 6.14: Distribution of density functions $f(y)$ (left column), AFS suppression factors ϵ_i (middle column), and the angles of quark diagonalization matrices ϵ_i/ϵ_j (right column). The top row corresponds to T^2 geometry while the bottom row corresponds to S^2. The sizes of T^2 and S^2 are set so that the average of $\ln \epsilon_2 = 0.275$. The distribution function of $\ln(\epsilon_2/\epsilon_3)$ is larger than that of $\ln(\epsilon_1/\epsilon_2)$ at $\ln(\epsilon_i/\epsilon_j) \sim 0$ for both T^2 and S^2.

different. Therefore we would like to build some intuition for how choosing different geometries with the same number of extra dimensions affects the distributions of the flavor parameters. To do this we now compare the Gaussian landscape defined on T^2 to a Gaussian landscape defined on S^2.

It is important to note that to obtain distribution functions for the observable flavor parameters requires a number of consecutive integrations. In the AFS approximation, it is the density function $f(y) \propto dV(b)/db^2$ that ultimately carries the information about the geometry of the extra dimensions. In order to obtain approximate distribution functions of the flavor suppression factors, we need to integrate two of the variables $y_{1,2,3}$ out of the distribution Eq. (6.58). Since for any one of the variables y_i the density $f(y_i)$ is not integrated out, singularities in the original density function $f(y_i)$ remain in the distribution function of $\ln \epsilon_i$. See for example the second column of Figure 6.14. However, because of the other integrations, the detailed information carried by $f(y)$ is smeared out, and the two plots of Figure 6.14 are quite similar.

The Yukawa matrix eigenvalues λ_i are approximated by $\epsilon_i^{\text{u-sector}} \epsilon_i^{\text{d-sector}}$, and hence their distribution functions are given by a convolution of the distribution functions of ϵ_i. Therefore the original density function $f(y)$ is integrated at least once. The same is true for the distribution functions of the three angles of the matrices that diagonalize the up-type and down-type Yukawa matrices. These angles are approximately ϵ_i/ϵ_j $(i < j)$, such that their distribution functions are obtained by integrating two variables out of Eq. (6.58) while keeping the distance $\Delta_{ij} \equiv y_i - y_j$ fixed. The third column of Figure 6.14 shows the distribution functions of Δ_{ij}. The singularity in $f(y)$ for T^2 is already smeared out, and we see that the distributions of Δ_{ij} are very similar between T^2 and

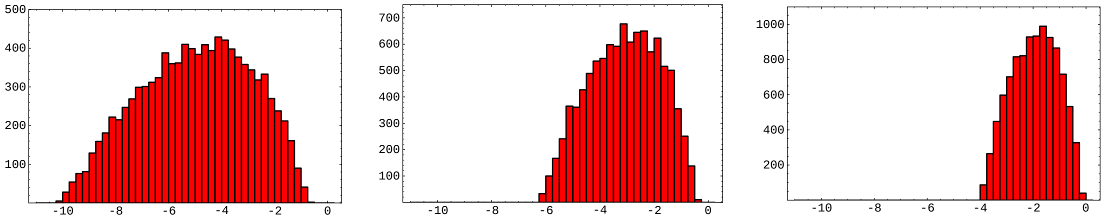


Figure 6.15: Distribution of 10^4 randomly generated Yukawa matrix elements on S^2. From left to right the panels correspond to $(d/R, g) = (0.4, 1)$, $(0.5, 1)$, and $(0.6, 1)$.

S^2, despite the differences in their density functions $f(y)$.

The theoretical argument above relies on the AFS approximation. To support its conclusions, we ran a numerical simulation of the Gaussian landscape on S^2. As before, the quarks and Higgs are represented by localized wavefunctions with width d. For example, on S^2 a wavefunction centered at $\theta = 0$ is given by[7]

$$\varphi(\theta, \phi) \propto e^{-\frac{\theta^2}{2d^2}} . \tag{6.63}$$

Of course, to generate an ensemble of Yukawa matrices the central coordinates of each wavefunction are scanned uniformly over the geometry S^2. Figure 6.15 shows the distribution of Yukawa matrix elements for this model. Note that the overall shape of the distributions in Figure 6.15 are remarkably similar to those in Figure 6.10, which involves the toy landscape on T^2. The distribution functions are approximately proportional to $|\ln \lambda|$ for large Yukawa matrix elements, similar to Figure 6.10 and in agreement with the theoretical discussion in section 6.5.2. That the distributions on S^2 are not as precisely linear for large matrix elements as those on T^2 can be understood in terms of geometry: S^2 has positive curvature while T^2 is flat. Figure 6.16 displays the distributions of the three Yukawa eigenvalues for the Gaussian landscapes on S^2 and T^2. These distribution are quite similar between S^2 and T^2, confirming the conclusion based on the AFS approximation.

Let us briefly summarize the conclusions of this section in the context of the above results. The number of extra dimensions can be visible in the distribution functions of observable flavor parameters because it determines the behavior of the density function $f(y)$ at a boundary $y \sim 0$. That is, boundary conditions survive (for distributions we have seen so far) regardless of how many times the distribution functions are integrated, convoluted, etc. This is why we obtain D-dependence in the distribution functions of observable flavor parameters, see Eq. (6.60). Meanwhile, although details about the geometry of the extra dimensions are still left in the distribution functions of the AFS factors ϵ_i, they are smeared out through integration processes, and do not survive to the level of observable flavor parameters. This means that we cannot learn very much about the geometry of extra dimensions from the observed masses and mixing angles. On the other hand, it appears that

---

[7] Although this wavefunction is not smooth at $\theta = \pi$, this is not of present concern. That is, the purpose of this numerical simulation is not to determine the distribution precisely but to study its qualitative aspects.

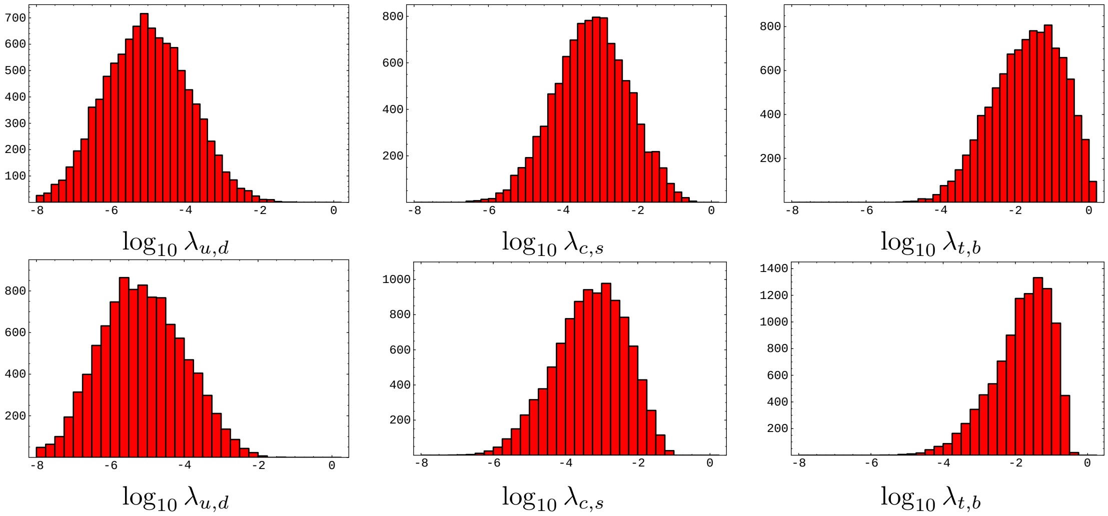


Figure 6.16: Comparison between the distributions of Yukawa eigenvalues from the Gaussian landscape on T^2 with $(d/L, g) = (0.1, 5)$ (top row) and the Gaussian landscape on S^2 with $(d/R, g) = (0.5, 1)$ (bottom row).

we can understand the qualitative pattern of masses and mixing angles without knowing the details of the underlying geometry.

6.6 Approximate Probability Distribution Functions

In this section we collect several probability distribution functions calculated using the AFS approximation, i.e. using Eq. (6.38) and Eqs. (6.40–6.42), as a reference for discussion within the main text of this chapter.

6.6.1 Gaussian Landscape with One Extra Dimension

We first collect the results from the Gaussian landscape on one extra dimension: S^1. Below we list the distribution functions of the Yukawa eigenvalues that follow from Eq. (6.33) and the approximation Eq. (6.38):

$$\frac{dP(z_3)}{dz_3} = \begin{cases} \frac{9}{4}\left(\pi - 8\sqrt{z_3} + (4+\pi)z_3 - \frac{8}{3}z_3^{3/2} + \frac{\pi}{8}z_3^2\right) & \text{for } 0 \le z_3 \le 1\,, \\ -\frac{3}{16}\Big[32 - 76\sqrt{z_3 - 1} + 48z_3 - 26z_3\sqrt{z_3-1} & \\ \quad - 3\left(8 + 8z_3 + z_3^2\right)\operatorname{arccot}\left(\sqrt{z_3-1}\right) & \\ \quad - 3\left(8 + 8z_3 + z_3^2\right)\arctan\left(\sqrt{z_3-1}\right)\Big] & \text{for } 1 < z_3 \le 2\,. \end{cases} \tag{6.64}$$

$$\frac{dP(z_2)}{dz_2} = \begin{cases} \frac{3}{8} z_2 \left(24 - 32\sqrt{z_2} + 3\pi z_2\right) & \text{for } 0 \le z_2 \le 1\,, \\ \frac{3}{2}\left(4 - 2\sqrt{z_2 - 1} - 6z_2 + 5z_2\sqrt{z_2 - 1}\right) & \\ -\frac{9}{4} z_2^2 \Big[\arctan\left(\sqrt{z_2 - 1}\right) - \operatorname{arccot}\left(\sqrt{z_2 - 1}\right)\Big] & \text{for } 1 < z_2 \le 2\,. \end{cases} \tag{6.65}$$

$$\frac{dP(z_1)}{dz_1} = \begin{cases} \frac{9\pi}{32} z_1^2 & \text{for } 0 \le z_1 \le 1\,, \\ \frac{9}{16}\Big\{2(2 - z_1)\sqrt{z_1 - 1} & \\ \quad - z_1^2 \Big[\arctan\left(\sqrt{z_1 - 1}\right) - \operatorname{arccot}\left(\sqrt{z_1 - 1}\right)\Big]\Big\} & \text{for } 1 < z_1 \le 2\,. \end{cases} \tag{6.66}$$

Here $z_i \equiv \ln(\lambda_i/\lambda_{\max})/\Delta \ln \epsilon$ where $\lambda_{\max} = (4/9\pi d^2)^{1/4} g$ and on S^1 we have $\Delta \ln \epsilon = -12(d/L)^2$. We also note the mean values of these distributions:

$$\langle z_3 \rangle = 0.2\,, \qquad \langle z_2 \rangle = 0.6\,, \qquad \langle z_1 \rangle = 1.2\,. \tag{6.67}$$

Meanwhile, the distribution functions of the CKM mixing angles follow from Eq. (6.33) and the set of approximations Eqs. (6.40–6.42). They are given by:

$$\begin{aligned} \frac{dP(t_{12})}{dt_{12}} &= 6\left(1 - \sqrt{t_{12}}\right)^3 \left(1 + 2\sqrt{t_{12}}\right) dt_{12}\,, && (6.68) \\ \frac{dP(t_{23})}{dt_{23}} &= \frac{3}{2}\left[-2\sqrt{1 - t_{23}} + \operatorname{arccosh}\left(\frac{2}{t_{23}} - 1\right)\right] \\ &\quad \times \left[2\sqrt{1 - t_{23}}(1 + 2t_{23}) - 3t_{23}\operatorname{arccosh}\left(\frac{2}{t_{23}} - 1\right)\right], && (6.69) \\ \frac{dP(t_{13})}{dt_{13}} &= 6\left(\sqrt{1 - t_{13}} + \sqrt{t_{13}} - 1\right)\left(1 - \sqrt{t_{13}}\right) \\ &\quad \times \left[2\sqrt{1 - t_{13}} - 1 - \sqrt{t_{13}} + 2t_{13} + 2\sqrt{t_{13}(1 - t_{13})}\right], && (6.70) \end{aligned}$$

where each applies over the full range $0 \le t_{ij} \le 1$, with $t_{ij} \equiv \ln|\sin\theta_{ij}|/\Delta \ln \epsilon$. The mean values of these distributions are:

$$\langle t_{12} \rangle = 0.17\,, \qquad \langle t_{23} \rangle = 0.10\,, \qquad \langle t_{13} \rangle = 0.35\,. \tag{6.71}$$

6.6.2 Gaussian Landscapes on $D = 2$ and $D = 3$ using $f_D(y)$ in Eq. (6.61)

We now turn to results from Gaussian landscapes in higher extra-dimensional spaces. The D-dimensional Gaussian landscape with density function $f_D(y)$ given by Eq. (6.61) predicts the following probability distributions of the AFS suppression factors:

$$\frac{dP(y_3)}{dy_3} = \frac{3D}{2} y_3^{D/2-1} \left(1 - y_3^{D/2}\right)^2\,, \tag{6.72}$$

$$\begin{aligned} \frac{dP(y_2)}{dy_2} &= 3D\, y_2^{D-1}\left(1 - y_2^{D/2}\right) , && (6.73) \\ \frac{dP(y_1)}{dy_1} &= \frac{3D}{2}\, y_1^{3D/2-1} . && (6.74) \end{aligned}$$

The approximate distribution of Yukawa eigenvalues follow from convolution of these suppression factors. For the case of $D = 2$ these are given by:

$$\frac{dP(z_3)}{dz_3} \simeq \begin{cases} \frac{3}{10}\, z_3\left(z_3^4 - 10z_3^3 + 40z_3^2 - 60z_3 + 30\right) & \text{for } 0 \le z_3 \le 1 \\ \frac{3}{10}\,(2 - z_3)^5 & \text{for } 1 \le z_3 \le 2\,, \end{cases} \tag{6.75}$$

$$\frac{dP(z_2)}{dz_2} \simeq \begin{cases} \frac{6}{5}\, z_2^3\left(5 - 5z_2 + z_2^2\right) & \text{for } 0 \le z_2 \le 1 \\ \frac{6}{5}\,(2 - z_2)^3\left(z_2^2 + z_2 - 1\right) & \text{for } 1 \le z_2 \le 2\,, \end{cases} \tag{6.76}$$

$$\frac{dP(z_1)}{dz_1} \simeq \begin{cases} \frac{3}{10}\, z_1^5 & \text{for } 0 \le z_1 \le 1 \\ \frac{3}{10}\left(12 - 30z_1 + 20z_1^2 - z_1^5\right) & \text{for } 1 \le z_1 \le 2\,, \end{cases} \tag{6.77}$$

where again $z_i \equiv \ln(\lambda_i/\lambda_{\max})/\Delta\ln\epsilon$, $\lambda_{\max} = (4/9\pi d^2)^{1/4} g$, and on S^2 we have $\Delta\ln\epsilon = -12(d/L)^2$ while on T^2 we have $\Delta\ln\epsilon = -6(d/L)^2$. The mean values of these distributions are:

$$\langle z_3\rangle = 0.25\,,\quad \langle z_2\rangle = 0.50\,,\quad \langle z_1\rangle = 0.75\,,\quad \frac{\langle z_1 - z_2\rangle}{\langle z_2 - z_3\rangle} = 1\,. \tag{6.78}$$

Meanwhile, for $D = 2$ the distribution of mixing angles is

$$\begin{aligned} \frac{dP(t_{12})}{dt_{12}} &= 6(1 - t_{12})^5, && (6.79) \\ \frac{dP(t_{23})}{dt_{23}} &= 6(1 - t_{23})^5, && (6.80) \\ \frac{dP(t_{13})}{dt_{13}} &= 12t_{13}(1 - t_{13})^3(1 + 2t_{13})\,, && (6.81) \end{aligned}$$

again with $t_{ij} \equiv \ln|\sin\theta_{ij}|/\Delta\ln\epsilon$. The mean values of these distributions are:

$$\langle t_{12}\rangle = 0.14\,,\quad \langle t_{23}\rangle = 0.14\,,\quad \langle t_{13}\rangle = 0.37\,. \tag{6.82}$$

In the case of $D = 3$, it is also possible to obtain analytic expressions for the approximate distributions of Yukawa eigenvalues. However, these are very complicated and we do not list them here. Instead, we simply list the mean values of these distributions:

$$\langle z_3\rangle = 0.37\,,\quad \langle z_2\rangle = 0.61\,,\quad \langle z_1\rangle = 0.82\,,\quad \frac{\langle z_1 - z_2\rangle}{\langle z_2 - z_3\rangle} = \frac{5}{6}\,. \tag{6.83}$$

The distribution functions of the three mixing angles are also very complicated for $D = 3$, so we do not list these either. The mean values of these distributions are:

$$\langle t_{12} \rangle = 0.11\,, \quad \langle t_{23} \rangle = 0.14\,, \quad \langle t_{13} \rangle = 0.33\,. \tag{6.84}$$

6.7 Discussion and Conclusions

This decade has seen the emergence of a major debate: to what extent is nature fundamentally uniquely prescribed, for example by symmetries, vs. to what extent nature results from the statistics of a huge landscape of solutions to the fundamental theory, modified by cosmological and environmental selection. Within the Standard Model symmetries play a key role, but for physics beyond the Standard Model the question remains largely open. Unified gauge symmetries have striking achievements: for example a simple interpretation of the quantum numbers of a generation and a precise numerical prediction for the ratios of the measured gauge couplings. For the paradigm of flavor, however, the picture offered by approximate flavor symmetries (AFS) is much less compelling, lacking both theoretical simplicity and significant successful predictions. Although the AFS description of flavor is apparently well suited to give an understanding of the hierarchical nature of the charged fermion masses and the CKM mixing matrix, it also comes with too much flexibility: with an appropriate choice of charges and symmetries any pattern of flavor can be generated.

In this chapter we have studied some very simple landscapes involving extra dimensions that could account for flavor. Small symmetry breaking parameters are replaced by small overlap integrals of wavefunctions in the extra dimensions. Localization of wavefunctions is a very natural expectation, and may have a more elegant realization than Higgs potentials for flavor symmetry breaking. We claim neither precise predictions nor a particular compelling statistical model, rather we investigate what patterns of flavor emerge in the quark sector from the simplest landscapes, and what features of these extra dimensional landscapes are relevant for the flavor problem.

In the simplest toy landscape, where the quarks and the Higgs having a universal Gaussian wavefunction but with central peaks that scan randomly over a circular extra dimension, we find:

- A scale invariant probability distribution (over a finite range) for all Yukawa matrix elements.
- Generation structure: a hierarchy of quark masses.
- Pairing structure: the heaviest up- and down-type quarks are mostly paired in the same weak doublet, as are the two middle and the two lightest quarks.
- Very small CKM mixing angles, typically with $\theta_{13} \ll \theta_{12}, \theta_{23}$.

The relevant probability distributions are shown in Figures 6.4 and 6.5 and result from inputting only two parameters. One parameter sets the common normalization to the Yukawa couplings in the higher dimensional theory, and is taken of order unity, $g \approx 1$, as it can arise from higher dimensional gauge interactions. The other is the ratio of the circumference L of S^1 to the width d of the universal Gaussian. This parameter is crucial since it sets the possible range of the Yukawa couplings. We find that $L/d \approx 10$ is sufficient to account for all the observed hierarchies in the charged fermion sector.

This generation structure and pairing structure for the quarks can also be obtained from AFS using two free parameters. For example, an approximate $U(1)$ symmetry with a leading Yukawa coupling g of order unity and others suppressed by various powers of a small symmetry breaking parameter ϵ. However there is a crucial difference. In the AFS case one must carefully choose the $U(1)$ charges of each of the Standard Model quarks. A huge variety of mass patterns could be accommodated by suitable choice of charges. In the Gaussian landscape case no such choices are made, and each of the quarks is treated symmetrically. The different Standard Model particles differ only by the location of their Gaussian wavefunction, and these are scanned randomly over the extra dimension. The hierarchies arise purely from statistics; they cannot be changed as they do not involve any free parameters beyond L/d.

While the above accomplishments of the Gaussian landscape on S^1 are striking, there are certain features that are less than ideal. Although they are peaked, the probability distributions for the quark masses and the CKM mixing angles are quite broad, as can be seen in Figure 6.5. At half maximum, the deviation from the peak value is typically an order of magnitude. Thus the statistical nature of the landscape prevents us from making precise predictions. We have found that this order of magnitude width is also typical of the Gaussian landscapes in two extra dimensions that we have studied. The Gaussian landscape makes no distinction between up and down sectors. For the lightest two generations this is fine: the u/d and c/s mass ratios could arise randomly. However, for t/b this seems much less likely. That is, the probability distribution for the top and bottom quark masses is narrower than for the other generations, making it unlikely that their Yukawa couplings differ by roughly two orders of magnitude. Even a strong selection effect favoring a heavy top quark does not completely remedy this problem.

One possibility to resolve this issue is to replace the universal coupling g with two parameters: one for the up sector that is about an order of magnitude larger than the one for the down sector. It would be interesting to find a landscape origin for such a "$\tan\beta$" factor. Another possibility is that the narrowness of the t/b mass distribution is a special feature of the Gaussian landscape on S^1. We find that this is the case, to a degree. In the Gaussian landscape on T^2—where the wavefunctions are distributed at random locations over the surface of a torus—the distribution of top and bottom Yukawas is wider, as shown in Figure 6.11. A selection for a heavy quark, for example to maintain the stability of our Higgs phase after electroweak symmetry breaking, could then account for the t/b mass ratio. Indeed, for any number of extra dimensions we are able to analytically approximate the distribution of Yukawa eigenvalues near their maximal value, and we find that that the probability distribution becomes more suppressed as the dimension increases. This strengthens this interpretation of the t/b ratio.

The observed mass ratios between the second and third generation can be used to determine the parameter L/d, such that we can *predict* the expected order of magnitude of the first generation masses. Since this prediction is broadly correct, we can apparently conclude that the effects of environmental selection, for example from nuclear physics, are not very strong on the first generation masses. While this is true for the Gaussian landscape on S^1, one must be very careful making such inferences. In the actual landscape it may be that the ratio L/d itself scans, and then it is possible that the range of Yukawa couplings is environmentally selected. Even if environmental selection effects are not very powerful, they can still lead to significant modification of the distributions for masses and mixings. This is illustrated in Fig 6.7 where a cut on the top quark mass induces an asymmetry between the c and s distributions, favoring a heavier charm quark. In Figure 6.8 the same cut is seen to reduce the probability of near maximal mixing in the quark sector.

Finally, we have made a first attempt to understand the robustness of the landscape results on S^1 by studying the probability distributions for the quark sector flavor observables that result from two two-dimensional spaces: a torus T^2 with equal lengths L around each circle, and the two-sphere S^2 with radius R. The four itemized accomplishments listed above remain intact, except that the probability distribution for Yukawa matrix elements deviates somewhat from the scale invariant form. That is, over a range determined by L/d or R/d the distribution of Yukawa matrix elements $dP/d\ln\lambda$ becomes a polynomial in $\ln\lambda$, rather than being constant. Yet this is a mild change; apart from the width of the distribution for the top and bottom Yukawa couplings mentioned above, the gross structure is unchanged. Specifically, the family pairings, mass and mixing angle hierarchies and order of magnitude widths of the distributions are all preserved, as seen for T^2 in Figures 6.11 and 6.12. Thus we conclude that these are robust features of quark sector flavor in Gaussian landscapes on extra dimensions.

Acknowledgments

This work was supported by the U.S. Department of Energy under contract Nos. DE-FG03-92ER40701 and DE-AC03-76SF00098 and by the U.S. National Science Foundation under Grant No. PHY-04-57315.

Bibliography

[1] C. W. Bauer, S. Fleming and M. E. Luke, Phys. Rev. D **63**, 014006 (2001).

[2] C. W. Bauer, S. Fleming, D. Pirjol and I. W. Stewart, Phys. Rev. D **63**, 114020 (2001).

[3] C. W. Bauer and I. W. Stewart, Phys. Lett. B **516**, 134 (2001).

[4] C. W. Bauer, D. Pirjol and I. W. Stewart, Phys. Rev. D **65**, 054022 (2002).

[5] C. W. Bauer, D. Pirjol and I. W. Stewart, Phys. Rev. D **67**, 071502 (2003).

[6] R. J. Hill and M. Neubert, Nucl. Phys. B **657**, 229 (2003).

[7] A. V. Manohar and M. B. Wise, *Heavy Quark Physics*, Cambridge University Press (2000).

[8] C. W. Bauer, M. P. Dorsten and M. P. Salem, Phys. Rev. D **69**, 114011 (2004).

[9] E. W. Kolb and M. S. Turner, *The Early Universe*, Perseus Publishing, Cambridge (1990).

[10] A. D. Linde, *Particle Physics and Inflationary Cosmology*, Harwood Academic, Chur (1990).

[11] D. H. Lyth and A. Riotto, Phys. Rept. **314**, 1 (1999).

[12] A. A. Starobinsky, Phys. Lett. B **117**, 175 (1982).

[13] M. Sasaki and E. D. Stewart, Prog. Theor. Phys. **95**, 71 (1996).

[14] A. A. Starobinsky, JETP Lett. **42**, 152 (1985) [Pisma Zh. Eksp. Teor. Fiz. **42**, 124 (1985)].

[15] D. H. Lyth, K. A. Malik and M. Sasaki, JCAP **0505**, 004 (2005).

[16] J. M. Bardeen, Phys. Rev. D **22**, 1882 (1980).

[17] J. M. Bardeen, P. J. Steinhardt and M. S. Turner, Phys. Rev. D **28**, 679 (1983).

[18] G. Dvali, A. Gruzinov and M. Zaldarriaga, Phys. Rev. D **69**, 023505 (2004).

[19] G. Dvali, A. Gruzinov and M. Zaldarriaga, Phys. Rev. D **69**, 083505 (2004).

[20] L. Kofman, arXiv:astro-ph/0303614.

[21] J. Maldacena, JHEP **0305**, 013 (2003).

[22] D. H. Lyth and D. Wands, Phys. Lett. B **524**, 5 (2002).

[23] T. Moroi and T. Takahashi, Phys. Lett. B **522**, 215 (2001) [Erratum-ibid. B **539**, 303 (2002)].

[24] K. Enqvist and M. S. Sloth, Nucl. Phys. B **626**, 395 (2002).

[25] D. H. Lyth, JCAP **0511**, 006 (2005).

[26] C. W. Bauer, M. L. Graesser and M. P. Salem, Phys. Rev. D **72**, 023512 (2005).

[27] M. P. Salem, Phys. Rev. D **72**, 123516 (2005).

[28] M. R. Douglas and S. Kachru, arXiv:hep-th/0610102.

[29] A. Vilenkin, Phys. Rev. D **27**, 2848 (1983).

[30] A. D. Linde, Mod. Phys. Lett. A **1**, 81 (1986).

[31] A. D. Linde, Phys. Lett. B **175**, 395 (1986).

[32] S. Weinberg, Phys. Rev. Lett. **59**, 2607 (1987).

[33] H. Martel, P. R. Shapiro and S. Weinberg, Astrophys. J. **492**, 29 (1998).

[34] B. Carter, in *International Astronomical Union Symposium 63: Confrontation of Cosmological Theories with Observational Data*, 291 (1974).

[35] T. Banks, Nucl. Phys. B **249**, 332 (1985).

[36] J. Barrow and F. Tipler, *The Anthropic Cosmological Principle*, Oxford Univ. Press (1986).

[37] W. H. Press and P. Schechter, Astrophys. J. **187**, 425 (1974).

[38] C. D. Froggatt and H. B. Nielsen, Nucl. Phys. B **147**, 277 (1979).

[39] See C. D. Froggatt, arXiv:hep-ph/9504323, and references therein.

[40] N. Arkani-Hamed and M. Schmaltz, Phys. Rev. D **61**, 033005 (2000).

[41] M. L. Graesser and M. P. Salem, arXiv:astro-ph/0611694.

[42] L. J. Hall, M. P. Salem, and T. Watari, in preparation.

[43] T. Becher, R. J. Hill and M. Neubert, Phys. Rev. D **69**, 054017 (2004).

[44] T. Becher, R. J. Hill, B. O. Lange and M. Neubert, Phys. Rev. D **69**, 034013 (2004).

[45] C. W. Bauer, D. Pirjol and I. W. Stewart, Phys. Rev. D **68**, 034021 (2003).

[46] R. J. Hill, private communication.

[47] For a recent discussion on this point see I. Z. Rothstein, arXiv:hep-ph/0308266.

[48] M. Beneke and T. Feldmann, Nucl. Phys. B **685**, 249 (2004).

[49] C. L. Bennett *et al*, Astrophys. J. **464**, L1 (1996).

[50] C. L. Bennett *et al*, Astrophys. J. Suppl. **148**, 1 (2003).

[51] D. N. Spergel *et al* [WMAP Collaboration], Astrophys. J. Suppl. **148**, 175 (2003).

[52] S. Dodelson, *Modern Cosmology*, Academic Press, San Diego (2003).

[53] H. V. Peiris *et al*, Astrophys. J. Suppl. **148**, 213 (2003).

[54] S. Matarrese and A. Riotto, JCAP **0308**, 007 (2003).

[55] A. Mazumdar and M. Postma, Phys. Lett. B **573**, 5 (2003) [Erratum-ibid. B **585**, 295 (2004)].

[56] F. Vernizzi, Phys. Rev. D **69**, 083526 (2004).

[57] S. Tsujikawa, Phys. Rev. D **68**, 083510 (2003).

[58] M. Zaldarriaga, Phys. Rev. D **69**, 043508 (2004).

[59] L. Ackerman, C. W. Bauer, M. L. Graesser and M. B. Wise, Phys. Lett. B **611**, 53 (2005).

[60] K. Enqvist, A. Mazumdar and M. Postma, Phys. Rev. D **67**, 121303 (2003).

[61] M. Postma, JCAP **0403**, 006 (2004).

[62] R. Allahverdi, Phys. Rev. D **70**, 043507 (2004).

[63] E. Komatsu *et al*, Astrophys. J. Suppl. **148** 119 (2003).

[64] C. P. Ma and E. Bertschinger, Astrophys. J. **455**, 7 (1995).

[65] A. H. Guth, Phys. Rev. D **23**, 347 (1981).

[66] A. D. Linde, Phys. Lett. B **108**, 389 (1982).

[67] A. H. Guth and S. Y. Pi, Phys. Rev. Lett. **49**, 1110 (1982).

[68] S. W. Hawking, Phys. Lett. B **115**, 295 (1982).

[69] S. Mollerach, Phys. Rev. D **42**, 313 (1990).

[70] A. D. Linde and V. Mukhanov, Phys. Rev. D **56**, 535 (1997).

[71] T. Matsuda, Phys. Rev. D **72**, 123508 (2005).

[72] B. A. Bassett, S. Tsujikawa and D. Wands, Rev. Mod. Phys. **78**, 537 (2006).

[73] F. Bernardeau, L. Kofman and J. P. Uzan, Phys. Rev. D **70**, 083004 (2004).

[74] V. F. Mukhanov, H. A. Feldman and R. H. Brandenberger, Phys. Rept. **215**, 203 (1992).

[75] D. Langlois, arXiv:hep-th/0405053.

[76] D. H. Lyth and Y. Rodriguez, Phys. Rev. Lett. **95**, 121302 (2005).

[77] H. C. Lee, M. Sasaki, E. D. Stewart, T. Tanaka and S. Yokoyama, JCAP **0510**, 004 (2005).

[78] A. Riotto, arXiv:hep-ph/0210162.

[79] A. D. Linde, Phys. Lett. B **249**, 18 (1990).

[80] A. D. Linde, Phys. Rev. D **49**, 748 (1994).

[81] A. D. Linde, Phys. Lett. B **129**, 177 (1983).

[82] T. S. Bunch and P. C. W. Davies, Proc. Roy. Soc. Lond. A **360**, 117 (1978).

[83] A. D. Linde, Phys. Lett. B **116**, 335 (1982).

[84] A. A. Starobinsky and J. Yokoyama, Phys. Rev. D **50**, 6357 (1994).

[85] K. Dimopoulos, D. H. Lyth, A. Notari and A. Riotto, JHEP **0307**, 053 (2003).

[86] E. Komatsu and D. N. Spergel, Phys. Rev. D **63**, 063002 (2001).

[87] A. S. Goncharov, A. D. Linde and V. F. Mukhanov, Int. J. Mod. Phys. A **2**, 561 (1987).

[88] R. Bousso and J. Polchinski, JHEP **0006**, 006 (2000).

[89] L. Susskind, arXiv:hep-th/0302219.

[90] M. R. Douglas, JHEP **0305**, 046 (2003).

[91] A. Giryavets, S. Kachru, P. K. Tripathy and S. P. Trivedi, JHEP **0404**, 003 (2004).

[92] A. Giryavets, S. Kachru and P. K. Tripathy, JHEP **0408**, 002 (2004).

[93] E. Baum, Phys. Lett. B **133**, 185 (1983).

[94] S. W. Hawking, Phys. Lett. B **134**, 403 (1984).

[95] S. R. Coleman, Nucl. Phys. B **307**, 867 (1988).

[96] T. Banks, Nucl. Phys. B **309**, 493 (1988).

[97] S. B. Giddings and A. Strominger, Nucl. Phys. B **307**, 854 (1988).

[98] V. Agrawal, S. M. Barr, J. F. Donoghue and D. Seckel, Phys. Rev. Lett. **80**, 1822 (1998).

[99] V. Agrawal, S. M. Barr, J. F. Donoghue and D. Seckel, Phys. Rev. D **57**, 5480 (1998).

[100] B. Feldstein, L. J. Hall and T. Watari, Phys. Rev. D **74**, 095011 (2006).

[101] A. D. Linde, Rept. Prog. Phys. **47**, 925 (1984).

[102] A. D. Sakharov, Sov. Phys. JETP **60**, 214 (1984).

[103] A. D. Linde, Print-86-0888 (June 1986), in *Three Hundred Years of Gravitation*, ed. by S. W. Hawking and W. Israel, Cambridge University Press (1987).

[104] M. Tegmark, A. Vilenkin and L. Pogosian, Phys. Rev. D **71**, 103523 (2005).

[105] T. Banks, M. Dine and E. Gorbatov, JHEP **0408**, 058 (2004).

[106] M. Dine, E. Gorbatov and S. D. Thomas, arXiv:hep-th/0407043.

[107] R. Kallosh and A. Linde, JHEP **0412**, 004 (2004).

[108] I. Antoniadis and S. Dimopoulos, Nucl. Phys. B **715**, 120 (2005).

[109] M. Dine, D. O'Neil and Z. Sun, JHEP **0507**, 014 (2005).

[110] M. Dine, JHEP **0601**, 162 (2006).

[111] M. Dine and Z. Sun, JHEP **0601**, 129 (2006).

[112] A. Vilenkin, Phys. Rev. Lett. **74**, 846 (1995).

[113] M. Tegmark, JCAP **0504**, 001 (2005).

[114] A. D. Linde, Phys. Lett. B **160**, 243 (1985).

[115] A. D. Linde, Phys. Lett. B **201**, 437 (1988).

[116] S. Hellerman and J. Walcher, Phys. Rev. D **72**, 123520 (2005).

[117] M. Tegmark, A. Aguirre, M. Rees and F. Wilczek, Phys. Rev. D **73**, 023505 (2006).

[118] B. J. Carr and M. J. Rees, Nature **278**, 605 (1979).

[119] R. Harnik, G. D. Kribs and G. Perez, Phys. Rev. D **74**, 035006 (2006).

[120] L. Clavelli and R. E. W. III, arXiv:hep-ph/0609050.

[121] M. Tegmark and M. J. Rees, Astrophys. J. **499**, 526 (1998).

[122] M. J. Rees, arXiv:astro-ph/0401424.

[123] M. L. Graesser, S. D. H. Hsu, A. Jenkins and M. B. Wise, Phys. Lett. B **600**, 15 (2004).

[124] A. Aguirre, Phys. Rev. D **64**, 083508 (2001).

[125] A. D. Linde and A. Mezhlumian, Phys. Lett. B **307**, 25 (1993).

[126] A. D. Linde, D. A. Linde and A. Mezhlumian, Phys. Lett. B **345**, 203 (1995).

[127] A. D. Linde and A. Mezhlumian, Phys. Rev. D **53**, 4267 (1996).

[128] J. Garcia-Bellido, A. D. Linde and D. A. Linde, Phys. Rev. D **50**, 730 (1994).

[129] J. Garcia-Bellido, Nucl. Phys. B **423**, 221 (1994).

[130] J. Garcia-Bellido and A. D. Linde, Phys. Rev. D **52**, 6730 (1995).

[131] S. Winitzki and A. Vilenkin, Phys. Rev. D **53**, 4298 (1996).

[132] A. Vilenkin, Phys. Rev. Lett. **81**, 5501 (1998).

[133] V. Vanchurin, A. Vilenkin and S. Winitzki, Phys. Rev. D **61**, 083507 (2000).

[134] J. Garriga and A. Vilenkin, Phys. Rev. D **64**, 023507 (2001).

[135] J. Garriga, D. Schwartz-Perlov, A. Vilenkin and S. Winitzki, JCAP **0601**, 017 (2006).

[136] V. Vanchurin and A. Vilenkin, Phys. Rev. D **74**, 043520 (2006).

[137] R. Easther, E. A. Lim and M. R. Martin, JCAP **0603**, 016 (2006).

[138] R. Bousso, Phys. Rev. Lett. **97**, 191302 (2006).

[139] A. Aguirre, S. Gratton and M. C. Johnson, arXiv:hep-th/0611221.

[140] A. Linde, JCAP **0701**, 022 (2007).

[141] N. Arkani-Hamed, S. Dimopoulos and S. Kachru, arXiv:hep-th/0501082.

[142] A. D. Linde, Phys. Lett. B **238**, 160 (1990).

[143] T. Biswas and A. Notari, Phys. Rev. D **74**, 043508 (2006).

[144] B. Feldstein, L. J. Hall and T. Watari, Phys. Rev. D **72**, 123506 (2005).

[145] J. Garriga and A. Vilenkin, Prog. Theor. Phys. Suppl. **163**, 245 (2006).

[146] K. Freese, J. A. Frieman and A. V. Olinto, Phys. Rev. Lett. **65**, 3233 (1990).

[147] N. Arkani-Hamed, P. Creminelli, S. Mukohyama and M. Zaldarriaga, JCAP **0404**, 001 (2004).

[148] M. Dine and A. Kusenko, Rev. Mod. Phys. **76**, 1 (2004).

[149] M. Trodden, Rev. Mod. Phys. **71**, 1463 (1999).

[150] A. Strumia, arXiv:hep-ph/0608347.

[151] M. Fukugita and T. Yanagida, Phys. Lett. B **174**, 45 (1986).

[152] H. Murayama and T. Yanagida, Phys. Lett. B **322**, 349 (1994).

[153] V. A. Kuzmin, V. A. Rubakov and M. E. Shaposhnikov, Phys. Lett. B **155**, 36 (1985).

[154] I. Affleck and M. Dine, Nucl. Phys. B **249**, 361 (1985).

[155] M. Dine, L. Randall and S. D. Thomas, Nucl. Phys. B **458**, 291 (1996).

[156] J. Preskill, M. B. Wise and F. Wilczek, Phys. Lett. B **120**, 127 (1983).

[157] S. Hawking, Mon. Not. Roy. Astron. Soc. **152**, 75 (1971).

[158] B. J. Carr and J. E. Lidsey, Phys. Rev. D **48**, 543 (1993).

[159] J. C. Niemeyer and K. Jedamzik, Phys. Rev. D **59**, 124013 (1999).

[160] D. N. Spergel *et al*, arXiv:astro-ph/0603449.

[161] J. M. Bardeen, J. R. Bond, N. Kaiser and A. S. Szalay, Astrophys. J. **304**, 15 (1986).

[162] J. Binney and S. Tremaine, *Galactic Dynamics*, Princeton University Press (1987).

[163] M.J. Rees and J.P. Ostriker, Mon. Not. R. Astron. Soc. **179**, 541 (1977).

[164] T. Abel, P. Anninos, Y. Zhang and M. L. Norman, New Astron. **2**, 181 (1997).

[165] P. Anninos, Y. Zhang, T. Abel and M. L. Norman, New Astron. **2**, 209 (1997).

[166] S. M. Fall and G. Efstathiou, Mon. Not. Roy. Astron. Soc. **193**, 189 (1980).

[167] P. J. E. Peebles, Astrophys. J. **155**, 393 (1969).

[168] P. J. E. Peebles, Astron. Astrophys. **11**, 377 (1971).

[169] G. Efstathiou and B. J. T. Jones, Mon. Not. R. Astron. Soc. **186**, 133 (1979).

[170] M. Steinmetz and M. Bartelmann, Mon. R. Astron. Soc. **272**, 570 (1995).

[171] M.J. Rees, Mon. Not. Roy. Astron. Soc. **176**, 483 (1976).

[172] J.P. Cox and R.T. Giuli, *Principles of Stellar Structure*, Gordon and Breach (1968).

[173] R. Kippenhahn and A. Weigert, *Stellar Structure and Evolution*, Springer-Verlag (1994).

[174] D.N. Limber, Astrophys. J., **127**, 363 (1958).

[175] D.N. Limber, Astrophys. J., **127**, 387 (1958).

[176] G. Chabrier and I. Baraffe, Astron. Astrophys. **327**, 1039 (1997).

[177] S. E. Woosley and T. A. Weaver, Ann. Rev. Astron. Astrophys. **24** 205 (1986).

[178] H. A. Bethe, Rev. Mod. Phys. **62** 801 (1990).

[179] K. Nomoto *et al*, Nucl. Phys. A **621**, 467c (1997).

[180] K. Iwamoto *et al*, Astrophys. J. Suppl. **125**, 439 (2000).

[181] V. F. Cardone and M. Sereno, Astron. Astrophys. **438**, 545 (2005).

[182] A. Linde and V. Mukhanov, JCAP **0604**, 009 (2006).

[183] L. J. Hall, T. Watari and T. T. Yanagida, Phys. Rev. D **73**, 103502 (2006).

[184] J. S. Bullock *et al*, Astrophys. J. **555**, 240 (2001).

[185] W.-M. Yao *et al* (Particle Data Group), J. Phys. **G 33**, 1 (2006).

[186] C. J. Hogan, Phys. Rev. D **74**, 123514 (2006), and references therein.

[187] E. A. Mirabelli and M. Schmaltz, Phys. Rev. D **61**, 113011 (2000).

[188] L. J. Hall, H. Murayama and N. Weiner, Phys. Rev. Lett. **84**, 2572 (2000).

[189] J. F. Donoghue, K. Dutta and A. Ross, Phys. Rev. D **73**, 113002 (2006).

[190] N. Haba and H. Murayama, Phys. Rev. D **63**, 053010 (2001).

[191] J. F. Donoghue, Phys. Rev. D **69**, 106012 (2004) [Erratum-ibid. D **69**, 129901 (2004)].

[192] N. Arkani-Hamed, S. Dimopoulos and S. Kachru, arXiv:hep-th/0501082.