hep-th/0201158
SU-ITP-02/02
SLAC-PUB-9112
October 2002

Signatures of Short Distance Physics in the Cosmic Microwave Background

Nemanja Kaloper[1], Matthew Kleban[1], Albion Lawrence[1,2] and Stephen Shenker[1]

[1]Department of Physics, Stanford University, Stanford, CA 94305

[2]SLAC Theory Group, MS 81, 2575 Sand Hill Road, Menlo Park, CA 94025

We systematically investigate the effect of short distance physics on the spectrum of temperature anistropies in the Cosmic Microwave Background produced during inflation. We present a general argument–assuming only low energy locality–that the size of such effects are of order H^2/M^2, where H is the Hubble parameter during inflation, and M is the scale of the high energy physics.

We evaluate the strength of such effects in a number of specific string and M theory models. In weakly coupled field theory and string theory models, the effects are far too small to be observed. In phenomenologically attractive Hořava-Witten compactifications, the effects are much larger but still unobservable. In certain M theory models, for which the fundamental Planck scale is several orders of magnitude below the conventional scale of grand unification, the effects may be on the threshold of detectability.

However, observations of both the scalar and tensor fluctuation contributions to the Cosmic Microwave Background power spectrum–with a precision near the cosmic variance limit–are necessary in order to unambiguously demonstrate the existence of these signatures of high energy physics. This is a formidable experimental challenge.

*Work supported in part by Department of Energy Contract DE-AC03-76SF00515.

Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94309, USA

1. Introduction

The enormous disparity in scales between the observed Planck mass $\sim 10^{19}$ GeV and the energy of current accelerators (10^3 GeV) stands as the main barrier to connecting theoretical work in quantum gravity to experiment. There are a few exceptions to this difficult situation. Proton decay experiments overcome this immense disparity by examining decays in kilotons of protons for millions of seconds. Investigations of coupling constant unification use the slow, logarithmic variation of couplings combined with the assumption of a desert to extract information about the nature and scale of unification. But such bright spots are few and far between.

Observational cosmology provides a window into very early times and hence, most think, into very high energy processes. This possible high energy probe has received much more attention recently because of the new data available, the experiments being done, and the experiments being planned to study the cosmic microwave backround radiation (CMBR). The benchmark theory that explains the fluctuations in the CMBR is inflation[1], which traces them to "thermal" quanta of a scalar inflaton field during a time of exponential expansion of the universe. In the simplest models of inflation the scale of vacuum energy during this period of exponential expansion was $\sim 10^{16}$ GeV and the rate of exponential expansion $H \sim 10^{13} - 10^{14}$ GeV. These enormous energies suggest that during the inflationary epoch various kinds of high energy processes were activated, and further, that they could have left their imprint on the CMBR.

Many authors have drawn attention to this exciting prospect. The first piece of high energy physics to be unraveled could well be dynamics of inflation itself. Much work has gone into how to reconstruct the potential of the inflaton field from CMBR data [4].

We should stress at this point that it is by no means necessary for the scale of inflation to be as high as $H \sim 10^{13} - 10^{14}$ GeV. Other inflationary models, e.g. hybrid models [5], exist where H can be much lower, for example, $H \sim 10^3$ GeV. Fortunately the scale of inflation can be experimentally tested. Since gravity couples to mass-energy, the amount of gravitational radiation produced during inflation is directly related to the energy available during inflation. This gravitational radiation imprints itself as a polarized component of the CMBR, whose power is proportional to $(H/m_4)^2$, where m_4 is the (reduced) four dimensional Planck length. So measurements of this power give a direct measurement of

[1] For textbook introductions see [1][2][3].

H. COBE [6] data already provide the interesting upper bound $H < 10^{14}$ GeV which corresponds to vacuum energies $\sim 10^{16}$ GeV, the supersymmetric unification scale. Intensive efforts are under way to improve this measurement.

In this paper we will concentrate on the "high scale" possibility for H since this gives the largest range for discovering new physics via inflation. There have been a number of investigations of the signature of high energy scale physics in the CMBR. Heavy particles produced by parametric resonance have been studied in [7]. Several groups [8,9,10,11,12] have studied the effect that simple phenomenological models of stringy corrections to gravity would have on the inflationary fluctuation spectrum in the CMBR. This work shares many features with the results we will present. These groups found that the size of these imprints on the CMBR is controlled by the natural dimensionless ratio $r = H^2/m_s^2$ where m_s is the string mass. For conventional weakly coupled string theories containing gravity m_s is approximately the same as the four dimensional Planck mass $\sim 10^{19}$ GeV so $r \sim 10^{-11}$. The actual size of the effects in these models depends on some delicate issues of boundary conditions at short distances that are not completely specified by the model. These groups have surveyed the range of possible long distance behaviors allowed by different boundary conditions. The authors of [10] have focused on boundary conditions that yield imprints of size $\sim r$ while the authors of [12] have focused on boundary conditions yielding effects of of size $\sim r^n, n \sim .5$. The analysis we present below shows that the effects are of size $\sim r$ in any theory that is local on momentum scales $\leq H$, an apparently sensible physical requirement. Such an effect is far too small to observe for $r \sim 10^{-11}$. In fact, the ultimate statistical limit of cosmic variance, the number of independent sky samples available, excludes it from being observed even in principle.

It is important to note, though, how great an improvement this ratio is over the suppression accelerator based physicists must confront. The energies accessible to them are of order 10^3 GeV so their suppressions are of order $(10^3/10^{19})^2 \sim 10^{-32}$. But the fact that $r \sim 10^{-11}$ is a vast improvement is cold comfort to an experimentalist waiting for counts in an apparatus.

But, as pointed out in [11], modern string and M theory models allow for the possibility of lower values of the fundamental mass scales, raising the possibility of more favorable ratios r. Much of this paper will be devoted to exploring this question in detail.

In Section 2, we will briefly review the framework of slow roll inflation, explaining the basic observable quantities in both scalar and tensor fluctuations. We emphasize that the size of inflationary perturbations is fully determined by physics at the scale H which is much below the Planck scale. Therefore the locality of effective theory used to compute the fluctuations implies that these perturbations are independent of the details of Planck scale physics.

In Section 3, we will explain the basic mechanism by which high energy physics leaves an imprint on CMBR fluctuations. We analyze this effect by assuming that string theory at energies $\sim H$ is approximately local. Therefore, by integrating out heavy degrees of freedom (of characteristic mass M), we can write a local effective action for the inflaton field at momentum scale H. We identify which terms contribute the largest effect for large M (the leading irrelevant operators) and recover the basic H^2/M^2 estimate for the imprint on the CMBR. We then show that all weakly coupled string models, and in fact all ordinary field theoretic models in the absence of fine tuning, give unobservably small effects.

In Section 4, we turn to strongly coupled string theory in a search for lower fundamental mass scales which may lead to larger effects. We analyze M theory models of the Hořava-Witten type using the phenomenologically appealing grand unified compactifications discussed in [13]. We show these models give effects of size $\leq 10^{-7}$, too small to be observed, but larger than the weakly coupled string models because the fundamental eleven dimensional Planck scale here is lower, $\sim 5 \times 10^{16}$ GeV. We go on to discuss G_2 compactifications of M theory. Here, rather than having, roughly speaking, one large dimension as in the Hořava-Witten case, we can have four large dimensions, as the singularities supporting gauge dynamics are codimension four [14,15]. If we abandon the requirement of precision grand unification and allow our compactification manifold to get as large as possible, while remaining consistent with the four dimensional character of inflation, we can make the imprint on the CMBR order one, and hence potentially observable. In these models the fundamental eleven dimensional Planck mass is $m_{11} \sim H \sim 7 \times 10^{13}$ GeV. We also consider the early cosmology of models with low unification scale $m_* \sim$ TeV. In these models the size of extra dimensions varies in the course of cosmological evolution, but the size of the imprints of the type we consider remains small.

In Section 5, we discuss in detail the requirements necessary to observe these effects and distinguish them from other phenomena. It turns out that corrections of this type over the range of wavelengths accessible in scalar CMBR observations look like a change in the power law, or "tilt" of the observed power. Such a change can be mimicked by a change in the inflationary potential. What cannot be mimicked is the differential effect in the scalar and tensor fluctuations due to short distance physics. This point was first made in [16].

Ordinary inflationary fluctuations, without new physics, obey "inflationary consistency conditions" connecting scalar and tensor quantities. New heavy physics predicts a violation of these conditions [16]. This is the unambiguous signal of new physics.

We then show that the differencing inherent in the inflationary consistency conditions means that the actual signal is not of the size $\sim r$ as discussed above, but is further suppressed by what is called an inflationary "slow roll parameter" which can range from $\sim .001 - .06$ in various models. So the size of the measurable effect is somewhat smaller than initial estimates suggest.

Finally, we point out that this unambiguous signal is very challenging to measure. First, it not only requires precision data for the scalar fluctuations, which are rapidly accumulating, but it also requires precision data for the tensor fluctuations, which have not even been observed yet. Forthcoming experiments may however be able to observe the tensor fluctuations if inflation occurred at a high scale by observing the B-mode polarization component of the CMBR. We argue that cosmic variance limited measurements over a substantial range in wavenumber of this quantity will be necessary to detect these signals. This is a formidable experimental challenge.

In Section 6, we conclude.

2. Slow Roll Inflation

We begin with a review of the basic tenets of inflation. We will parameterize the inflationary potential V by a scale M^4 and a dimensionless function $\mathcal{V}$; $V = M^4\mathcal{V}$. We will work in a spatially flat FRW universe with the metric

$$ds^2 = -dt^2 + a^2(t)d\vec{x}^2. \tag{2.1}$$

The independent background field equations then reduce to

$$\begin{aligned} 3H^2 &= \frac{1}{m_4^2}\Big[\frac{\dot{\phi}^2}{2} + M^4\mathcal{V}\Big] \\ \ddot{\phi} + 3H\dot{\phi} + M^4\frac{\partial\mathcal{V}}{\partial\phi} &= 0, \end{aligned} \tag{2.2}$$

where $H = \dot{a}/a$ is the Hubble parameter, $m_4 \sim 2.4 \times 10^{18}$GeV is the reduced four dimensional Planck mass, and dots denote time derivatives. The main feature of inflationary dynamics in the slow roll approximation is that we can ignore the acceleration of the scalar field, because the cosmological expansion has the effect of friction and nearly freezes the scalar on the potential slope. The universe is dominated by the scalar field potential energy and undergoes a period of rapid expansion. The usual parameters which characterize the validity of the slow roll approximation are

$$\begin{aligned} \eta &= \frac{\ddot{\phi}}{H\dot{\phi}} \\ \epsilon &= \frac{3\dot{\phi}^2}{2M^4\mathcal{V}}. \end{aligned} \tag{2.3}$$

The slow roll approximation is then formally defined as the regime $|\eta|, |\epsilon| \ll 1$. The relative importance of these parameters depends on the model of inflation. For example, as we will see below in the case of natural inflation [17], or modular inflation, [18], $\epsilon \ll \eta$. Thus the deviations from slow roll are mainly coded in the parameter η. In contrast, in the simplest model of chaotic inflation driven by a mass term, $\epsilon \sim m^2/H^2 >> \eta = 0$ in the slow roll regime.

In the slow roll approximation, the equations (2.2) become

$$\begin{aligned} 3H^2 &= \frac{M^4}{m_4^2}\mathcal{V} \\ 3H\dot{\phi} + M^4\frac{\partial \mathcal{V}}{\partial \phi} &= 0. \end{aligned} \tag{2.4}$$

Using these equations, one readily finds the slow roll parameters in terms of the potential function $\mathcal{V}$:

$$\begin{aligned} \eta &= \epsilon - m_4^2\frac{\partial_\phi^2\mathcal{V}}{\mathcal{V}} \\ \epsilon &= m_4^2\frac{[\partial_\phi\mathcal{V}]^2}{2\mathcal{V}^2}. \end{aligned} \tag{2.5}$$

The equations (2.4) can now be integrated; they yield

$$\frac{da}{a} = -\frac{1}{m_4^2}\frac{\mathcal{V}}{\partial_\phi\mathcal{V}}d\phi, \tag{2.6}$$

which separates variables for any potential $\mathcal{V}$. The solution is

$$a \simeq a_0 \exp\Big[\frac{1}{m_4^2}\int_\phi^{\phi_0} d\phi\frac{\mathcal{V}}{\partial_\phi\mathcal{V}}\Big] \simeq a_0 \exp\Big[\frac{\mathcal{V}_0\,(\phi_0-\phi)}{m_4^2\partial_\phi\mathcal{V}_0} + ...\Big] \tag{2.7}$$

in the slow roll regime. Hence, the universe will undergo rapid expansion while the vev of the inflaton may change only minutely. The space-time geometry is approximated by a future half of de Sitter space during this period. Eventually however the change of the inflaton vev accumulates enough for the inflaton to depart the slow roll regime, and the potential becomes steeper. The inflaton approaches the minimum of the potential, begins to oscillate around it and produce matter particles, reheating the inflated universe back to temperatures which will eventually produce the universe we inhabit.

Let us imagine that at late times the vacuum energy vanishes and inflation terminates such that there are no cosmological event horizons. This avoids conceptual difficulties with quantum gravity in spacetimes with cosmological horizons, but suffices to illustrate the main features of inflationary dynamics in the spacetime language. The causal structure of the universe is then given by the Penrose diagram of Fig. 1.

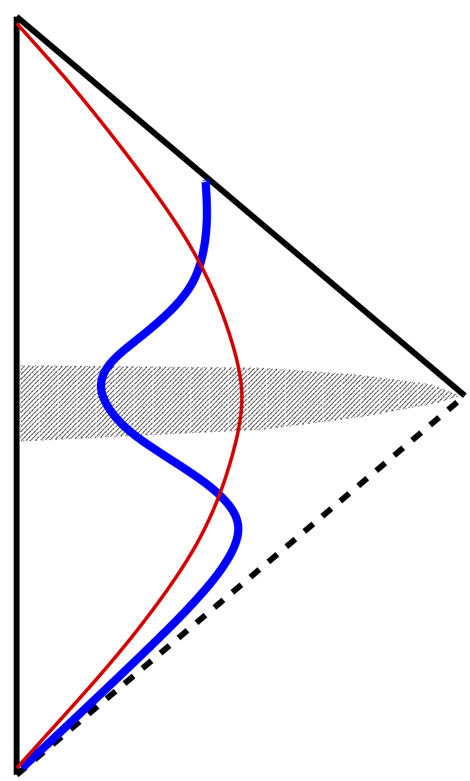

Fig. 1: Causal diagram of an inflationary model. The dashed past null line is the true particle horizon, but it could also be a null singularity. The future null line is the future infinity. The shaded area denotes the region of exit from inflation and reheating. The thin solid line is a worldline of any spacelike separated object from an observer at the center of the space. The bold solid line is the apparent horizon. Its shape is characteristic of inflation in the past, and radiation domination followed by matter domination in the future.

In the diagram, the region of geometry below the particle horizon (dashed line) is irrelevant for the future evolution as long as the period of inflation was sufficiently long. The future infinity appears as a consequence of our requirement for global exit from inflation. The spacetime below the reheating regime is the inflationary region, while that above it is the postinflationary, decelerating FRW universe. The thin solid line denotes any object spacelike separated from us, and for example is the worldline a distant galaxy follows after it forms. The bold solid line represents the apparent horizon, which plays central role for controlling the dynamics of inflation, as we will now elaborate. During inflation, it starts out almost null and "outward" directed, and then it flips "inward". This reflects that $H \sim$ const. during inflation. It ensures that the apparent horizon plays the role of the causal censor, limiting the amount of information which can fit inside an inflating region. The spacetime will therefore obey cosmic no-hair theorem, and inflation will succeed in getting rid of initial inhomogeneities. This may be viewed as another example of the cosmological variant of the holographic principle [19][20]. The structure of the spacetime is fully coded on the preferred screen, i.e. the apparent horizon. Its area is small during inflation because H must be large, and hence the Hubble region is censored from excessive outside influence, because only a limited amount of information can fit in the interior. Moreover, most of the objects inside the Hubble region are in the thermal bath of fluctuations located in the region when the apparent horizon is almost null [21], with the cosmological Hawking temperature $T_H = H/2\pi$. Since the inflaton is much lighter than the Hubble scale during inflation, the interactions with the thermal quanta cause its vev and the background metric to fluctuate.

Because of these quantum fluctuations, the inflaton is not exactly frozen to its slowly varying background vev. Instead it hops on the potential around the background value. Thus inside of some regions of the universe inflation may terminate a little later, because quantum effects push the inflaton a little farther up the plateau. These regions end up a fraction denser than their surroundings, and the matter in them begins to condense sooner, attracting additional matter from the neighborhood and eventually forming clusters and galaxies due to the classical Jeans instability. The fluctuations therefore induce small inhomogeneities on the perfectly smooth geometry left by inflation, which is measured experimentally via its imprint on the cosmic microwave background radiation, $\delta\rho/\rho \sim \delta T/T$, thanks to the Sachs-Wolfe effect. This is directly measured by the COBE [6] satellite, and by the BOOMERanG [22] and MAXIMA [23] experiments, which set the normalization for the inhomogeneities at around $\delta\rho/\rho \sim 10^{-5}$. They further observe that the spectrum of inhomogeneities is nearly scale-independent.

To determine the imprint of the fluctuations quantitatively we can use perturbation theory. In perturbation theory the fluctuations can be decomposed with respect to their transformation properties relative to the residual diffeomorphisms into scalar, vector and tensor modes. The vector modes decouple during inflation. Thus only the scalar and tensor modes are produced. The scalar modes cause density (and therefore CMBR temperature) fluctuations. The tensor modes are primordial gravitational waves produced by inflation, and affect the polarization of CMBR.

A heuristic derivation of the scalar density contrast is as follows: the RMS fluctuation of the field induced by the thermal fluctuations is $\delta\phi = \dot{\phi}\delta\tau$, and that of energy density is $\delta\rho = C\rho H\delta\tau$, where C is a numerical coefficient of order unity, whose precise value depends on the details of postinflationary cosmology. The function $\delta\tau$ is the time delay imprinted by the fluctuations on the vev in different regions of space. Combining these equations, one finds

$$\frac{\delta\rho}{\rho} = C\frac{H}{\dot{\phi}}\delta\phi \tag{2.8}$$

and then one needs to compute the RMS fluctuation of the inflaton $\delta\phi$. As we will discuss in more detail below, fluctuations of the transverse traceless modes of the graviton (which obey free scalar field equations) also contribute to the density variations.

In order to determine precisely how the quantum fluctuations of these fields evolve into temperature anisotropies in the sky today, one must first compute their effect on the curvature, and then use gauge-invariant gravitational perturbation theory to evolve the perturbation forward to the present era. One can define a gauge-invariant variable ζ, which is well approximated by the right hand side of (2.8) as modes exit from the de Sitter horizon during inflation, and which is approximately constant between this time, and when the mode re-enters the cosmological horizon later. At this later time ζ is well approximated by $\delta\rho/\rho$, establishing (2.8) [24]. This stage in the process is purely classical, because energy scales below H correspond to scales outside the causal horizon, and so coherent quantum fluctuations do not contribute at these wavelengths. The correct procedure is therefore to compute the quantum fluctuation of the inflaton field in de Sitter space, and then use it to evaluate ζ at the time the fluctuation exits the horizon; i.e., at momentum $p = H$.

As pointed out in [25], if the slow roll approximation breaks down this procedure will not be accurate (however, see [26]). For the sake of simplicity we will restrict ourselves to models where this is not a concern.

To compute the quantum fluctuation itself, one treats the fluctuating field as a perturbation around the de Sitter background and computes the mean-square variance as the (appropriately normalized) Fourier component of an equal-time two-point function evaluated at 3-momentum $p = H$:

$$(\delta\phi)^2 \sim \langle \phi(p)\phi(-p)\rangle|_{p=H}, \tag{2.9}$$

where ϕ represents either the inflaton or a physical mode of the graviton. The normalization is determined by the more detailed computation we perform below. In standard inflation, this is done assuming the inflaton is a free, minimally coupled scalar. As we will demonstrate in section 3, interactions with massive particles will modify the 2-point function and affect the spectrum of fluctuations. As long as the self-interactions of the inflaton (either in the classical potential, or induced by quantum corrections) are weak at energy scale H, so that a perturbative expansion is valid at this scale, this procedure is well defined. Of course, more general theories will involve strong coupling, but generically will also violate the observed constraints on $\delta\rho/\rho$.

Before considering such complications, we review the standard calculation. We begin by approximating the geometry by a future portion of de Sitter space. With $a = a_0 \exp(Ht)$ in (2.1), the inflaton field equation is

$$\ddot{\phi} + 3H\dot{\phi} - e^{-2Ht}\vec{\partial}^2\phi = H^2\left(\eta^2\partial_\eta^2\phi - 2\eta\partial_\eta\phi - \eta^2\vec{\partial}^2\phi\right) = 0, \tag{2.10}$$

where we have transformed to the conformal time $\eta \equiv -H^{-1}e^{-Ht}$. We can quantize ϕ by considering the general solution

$$\phi_p(\eta) = \frac{\sqrt{\pi}}{2}H\eta^{3/2}\left[A_k H^{(1)}_{3/2}(k\eta) + B_k H^{(2)}_{3/2}(k\eta)\right]. \tag{2.11}$$

Choosing the vacuum which matches the flat space case in the infinite past $\eta \to \infty$ and in the high frequency limit $k \to \infty$, we find that positive frequency modes are $A_k = 0$, $B_k = -1$. Then the mode expansion in Minkowski space is

$$\phi(\vec{x}, t) = (2\pi)^{-3/2}\int d^3k\left[a_k^\dagger \phi_k(t)e^{i\vec{k}\cdot\vec{x}} + a_k\phi_k^*(t)e^{-i\vec{k}\cdot\vec{x}}\right] \tag{2.12}$$

where

$$\phi_k(t) = \frac{iH}{k\sqrt{2k}}\left(1 + \frac{k}{iH}e^{-Ht}\right)\exp\left(\frac{ik}{H}e^{-Ht}\right). \tag{2.13}$$

The positive frequency 2-point function is

$$G^+(x,x') \equiv \langle 0|\phi(x)\phi(x')|0\rangle = \frac{1}{(2\pi)^3}\int d^3k \times$$
$$\times\, e^{-i\vec{k}(\vec{x}-\vec{x}')}\left[\frac{H^2}{2k^3} + \frac{e^{-H(t+t')}}{2k} + \frac{iH}{2k^2}\left(e^{-Ht} - e^{-Ht'}\right)\right]\exp\left(-\frac{ik}{H}\left(e^{-Ht} - e^{-Ht'}\right)\right). \tag{2.14}$$

To evaluate the fluctuations of the inflaton at lowest order, we compute the quantity

$$\langle\phi(x)\phi(x)\rangle = \frac{1}{(2\pi)^3}\int d^3k\left(\frac{e^{-2Ht}}{2k} + \frac{H^2}{2k^3}\right) = \frac{1}{(2\pi)^3}\int \frac{d^3p}{p}\left(\frac{1}{2} + \frac{H^2}{2p^2}\right), \tag{2.15}$$

where $p = e^{-Ht}k$ is the physical momentum conjugate to the proper distance $\tilde{x} = e^{Ht}x$.

The first term, which gives a UV-divergent contribution, is identical to the flat space result and should therefore be ignored. In other words, we are interested only in effects proportional to H, not in flat-space fluctuations which can be renormalized away. The second term is peculiar to de Sitter space and requires more careful treatment.

The magnitude of the fluctuations is determined by their power $P_\phi(k)$, defined by $\langle\phi(x)^2\rangle = \int \frac{dk}{k}P_\phi(k)$. Then the mean-square spectrum of fluctuations is $\langle|\delta\phi|^2\rangle = P_\phi(H)$. From (2.15) , we see that

$$\langle\phi(x)\phi(x)\rangle = \frac{1}{2\pi^2}\int \frac{dk}{k}\left(\frac{k^2}{2} + \frac{H^2}{2}\right), \tag{2.16}$$

so, neglecting the first term as explained above, we obtain

$$\langle|\delta\phi|^2\rangle = \frac{H^2}{4\pi^2}. \tag{2.17}$$

This gives $\delta\phi = H/2\pi$, finally yielding

$$\frac{\delta\rho}{\rho} = \frac{C}{2\pi}\frac{H^2}{\dot{\phi}}. \tag{2.18}$$

At this point, it is clear how to incorporate interactions into the calculation. If the theory contains a massive field (with mass $M \gg H$) which couples to the inflation, we can integrate it out using standard field theory techniques and obtain an effective potential for the inflaton. As can be seen from the two point function (2.14), such a procedure yields–in addition to the ordinary flat space terms–terms proportional to H^2/M^2, p^2H^2/M^4, etc. It is important to note that no cutoff or Planck scale comes into these corrections. The highest probe energy available in inflation and later visible in the CMB is H. As discussed in Sec. 3, it is these contributions we are primarily concerned with in this paper.

We can re-express (2.18) in terms of the inflationary potential using the slow roll equations (2.4). It is

$$\frac{\delta\rho}{\rho} = \frac{C}{2\sqrt{3}\pi}\frac{M^2}{m_4^3}\frac{\mathcal{V}^{3/2}}{\partial_\phi \mathcal{V}}. \tag{2.19}$$

This is the familiar formula for scalar fluctuations in inflation. We note that the so-called scalar power spectrum δ_S^2 is related to the density contrast by $\delta_S^2 = (2/5C)^2(\delta\rho/\rho)^2$, and using (2.19) we can express it as

$$\delta_S^2 = \frac{1}{75\pi^2}\frac{M^4}{m_4^6}\frac{\mathcal{V}^3}{[\partial_\phi \mathcal{V}]^2} \tag{2.20}$$

The causal structure of the inflationary spacetime depicted in Fig. 1. provides a straightforward understanding of the emergence of a (nearly) scale-invariant spectrum of fluctuations. A quantum fluctuation which seeds a galaxy is created just before its worldline intersects the apparent horizon. At that instant, it is as big as the Hubble horizon. Then it is expelled outside of the apparent horizon, where it freezes, and remains frozen until horizon reentry in distant future. After reentry the fluctuation becomes dynamical and evolves as dictated by gravitational instability. Scale invariance then follows from causal evolution if $H \simeq$ const., because the fluctuations of very different wavelengths are produced with the same amplitude. The evolution of the fluctuations can initially be described well by linear perturbation theory. However, nonlinearities eventually develop because of nontrivial interactions with the environment. In the matter dominated era, the fluctuations evolve differently before decoupling than after it. Before decoupling, the universe is opaque and therefore the baryonic matter is influenced by radiation pressure, which competes with gravitational collapse. This results in the emergence of acoustic oscillations, with characteristic peaks imprinted on the CMBR. The peaks appear because the perturbations whose wavelengths are half-integer divisors of the sound horizon (i.e. the largest distance sound can travel within the time of recombination) at decoupling can complete full oscillation cycles. The location and the heights of the peaks measure very accurately the cosmological parameters, in particular the Hubble parameter at decoupling.

Before turning to the specifics of modular inflation, we briefly review the mechanism for generating tensor fluctuations during inflation. These are just the gravitational waves, and correspond to the transverse-traceless metric fluctuations h_{kl}. They obey the linearized field equation $\nabla^2 h^k{}_l = 0$, where the covariant derivatives and raising and lowering of indices is defined relative to the background metric $g_{\mu\nu} = \mathrm{diag}(-1, a^2(t)\delta_{kl})$. Therefore each of the two graviton polarizations obeys the free massless scalar equation, and it is straightforward to quantize them in de Sitter space, in precisely the same way as in eqs. (2.10) - (2.17). In particular the root mean square fluctuation of the graviton is $\langle \delta h_{kl} \rangle \simeq H/2\pi$. However the formula for the tensor power spectrum is different than for the scalar. It is directly proportional to the fluctuation of the metric,

$$\delta_T^2 = \frac{1}{2\pi^2}\frac{H^2}{m_4^2} = \frac{1}{6\pi^2}\frac{M^4}{m_4^4}\mathcal{V} \tag{2.21}$$

by slow roll equations (2.4). The tensor nature of these fluctuations induce oscillations in the plasma during decoupling which polarize the CMB photons in an observable way [27].

The ratio $\mathcal{R} = \delta_T^2/\delta_S^2$ is a characteristic of the inflationary model, and is given by

$$\mathcal{R} = \frac{25}{2}\frac{m_4^2[\partial_\phi \mathcal{V}]^2}{\mathcal{V}^2} \tag{2.22}$$

It is straightforward to verify that in terms of the slow roll parameters, $\mathcal{R}$ is given as

$$\mathcal{R} = 25\epsilon \ . \tag{2.23}$$

The fluctuation spectra produced in inflation are not exactly scale-invariant. If the background inflaton vev were exactly frozen, and the geometry precisely de Sitter, the prediction for fluctuations (2.18)(2.21) would have been time-independent, and therefore exactly scale-invariant. In reality, there is weak time-dependence in (2.18) because the inflaton is sliding down the plateau. This time dependence, manifest in the variation of H and $\dot{\phi}$, translates into scale dependence of fluctuations, and produces a spectrum which is not exactly scale-invariant. This departure from scale invariance is a function of the specifics of inflationary model as defined by the potential. Below we will consider the details in the case of modular inflation.

2.1. The Specifics of Modular Inflation

To proceed, we need to determine more closely the form of the inflaton potential. For definiteness, we approximate here the potential function $\mathcal{V}$ by an inverted parabola

$$\mathcal{V} = 1 - \left(\frac{\phi}{\mu}\right)^2 \tag{2.24}$$

This approximation is generically valid in at least some region of the primordial universe which begins to inflate, if the inflaton is a modulus. The modulus begins near the maximum of the potential. Then the slow roll conditions yield

$$
\begin{aligned}
H &= \frac{M^2}{\sqrt{3}m_4}\sqrt{1-(\frac{\phi}{\mu})^2} \\
\dot{\phi} &= \frac{2M^2 m_4}{\sqrt{3}\mu^2}\frac{\phi}{\sqrt{1-(\frac{\phi}{\mu})^2}}
\end{aligned}
\tag{2.25}
$$

The slow roll parameters for (2.24) are initially

$$
\begin{aligned}
\eta &\simeq \epsilon - m_4^2\frac{\partial_\phi^2\mathcal{V}}{\mathcal{V}} \simeq \frac{2m_4^2}{\mu^2} \\
\epsilon &\simeq \frac{m_4^2}{2}\frac{[\partial_\phi\mathcal{V}]^2}{\mathcal{V}^2} \simeq \frac{2m_4^2\phi^2}{\mu^4} < \eta
\end{aligned}
\tag{2.26}
$$

and hence, the parameter $1/\mu^2$ which we introduced in the potential (2.24) is equal to $1/m_4^2$ multiplied by a small parameter $\eta/2$. This guarantees that the potential is sufficiently flat to support inflation.

We could now solve these equations directly. However, rather than integrating to find the time-dependence, it is more instructive to solve the equation (2.6). We will use the number of efolds before the end of inflation, or equivalently, the value of the scale factor a, as the cosmic clock. First, we define the number of efolds that universe has expanded by to be

$$
N = \ln\Big(\frac{a}{a_0}\Big) = \int_{t_0}^{t} dtH = \int_{\phi_0}^{\phi} d\phi\frac{H}{\dot{\phi}}
\tag{2.27}
$$

Then using (2.25) we can explicitly integrate this to find

$$
N = \frac{1}{\eta}\Big[\ln\Big(\frac{\phi}{\phi_0}\Big) + \frac{1}{2\mu^2}\Big(\phi_0^2-\phi^2\Big)\Big]
\tag{2.28}
$$

Here ϕ_0 is the initial value of the inflaton. In modular inflation, ϕ_0 would typically be near the top of the potential, in this case near zero. Such initial conditions produce a huge amount of inflation, as is clear from (2.27), which diverges in the limit $\phi_0 \to 0$. Of all that expansion, we can only observe the final 60 efolds or so, during which the universe expands by a factor of about 10^{26}. Any indications of expansion beyond that would be completely outside of the current size of the universe, and hence not accessible to our observations. Because we are only interested in the last 60 efolds, we can take ϕ_0 to be near its value at the end of inflation. Inflation ends when the slow-roll conditions cease to be valid, i.e. when $\eta, \epsilon \sim 1$. This occurs when the vev of ϕ grows to $\partial_\phi^2 V \sim H^2$, which in the case of modular potential (2.24) happens when $\phi \sim \mu$. Because a small change in the value of ϕ, of order $\sim e$, yields $N \sim \eta^{-1}$ efolds of inflation, we require that $\eta \sim 1/70$, which is enough to solve the horizon and flatness problems. In that case the latter two terms in eq. (2.28) are essentially negligible compared to the logarithm, and we will drop them hereafter. We note that in this case the other slow roll parameter is $\epsilon \simeq 2m_4^2/\mu^2 e^2 \simeq \eta/e^2$.

We now define $\mathcal{N} = N_* - N$ as the number of efolds left before the end of inflation. This variable is convenient to make contact with large scale structure and CMB observations. In terms of it, we can write down the solutions as

$$\begin{aligned} \mathcal{N} &= \frac{1}{\eta}\Big[\ln\Big(\frac{\mu}{\phi}\Big) + \frac{\phi^2-\mu^2}{2\mu^2}\Big] \\ H &= \frac{M^2}{\sqrt{3}m_4}\sqrt{1-(\frac{\phi}{\mu})^2} \\ a &= a_{final}e^{-\mathcal{N}} \end{aligned} \tag{2.29}$$

Inflation now lasts from when $\mathcal{N} \sim 60$ or larger, to about $\mathcal{N} = 0$, at which point the higher-order terms in the modular potential, ignored for clarity in (2.24), become important and reverse the sign of the effective mass term of ϕ.

Because the rolling of the scalar down the potential is slow, the Hubble parameter and the scalar field change very little, and hence the amplitude of fluctuations remains nearly constant throughout inflation. Therefore the fluctuations are being incessantly produced with an almost constant value, and deployed outside of the horizon. They stay there until a long time into the future, when the Hubble horizon eventually grows large enough, and they cross back inside, and start to collapse. These are the fluctuations we observe on the sky. The weak time dependence of H and $\dot{\phi}$ implies that $\delta\rho/\rho$ is weakly scale-dependent. We trade the time dependence off for the scale dependence by the horizon crossing matching, defining the comoving momentum k of the fluctuation at horizon-crossing by

$$k = aH \tag{2.30}$$

For modular inflation solution (2.29) this enables us to explicitly evaluate (2.18) as a function of k. First, we note that

$$k \simeq k_0 e^{-\mathcal{N}} \tag{2.31}$$

where k_0 is the comoving momentum leaving the Hubble horizon at the end of inflation. In terms of it, we find

$$\frac{\delta\rho(k)}{\rho} \simeq \frac{2C\mu M^2}{8\pi\sqrt{3}m_4^3}\Big(\frac{k_0}{k}\Big)^{\eta}\Big[1-\Big(\frac{k}{k_0}\Big)^{2\eta}\Big]^{3/2} \tag{2.32}$$

At 50 efolds before the end of inflation, the COBE measurements set the overall normalization to $\delta\rho/\rho|_{50} \sim 5\times10^{-5}$. Thus $C\mu M^2/m_4^3 \simeq 8\times10^{-4}$. Taking now $C=\mathcal{O}(1)$ and $\mu \simeq \sqrt{140}m_4$ (2.26), we obtain

$$\frac{M}{m_4} \simeq 8.2\times10^{-3} \tag{2.33}$$

In this case, the Hubble scale at inflation is, using the first of (2.25),

$$H \simeq 5.2\times10^{13}\text{GeV} , \tag{2.34}$$

which is within the bound allowed by COBE and large-scale structure measurements, c.f. [28].

The scale dependence is conveniently represented by defining the spectral index n_S (or the tilt) and its gradient as

$$\begin{aligned} n_S &= 1+2\frac{d\ln\frac{\delta\rho}{\rho}}{d\ln k} \\ \nu_S &= \frac{dn_S}{d\ln k} = 2\frac{d^2\ln\frac{\delta\rho}{\rho}}{d\ln k^2} \end{aligned} \tag{2.35}$$

For the modular inflation model we find

$$n_S = 1-2\eta-\frac{6\eta}{(k_0/k)^{2\eta}-1}. \tag{2.36}$$

Numerically, using (2.36), we find the tilt (at the scales corresponding to 50 efolds before the end of inflation) to be $n_S \simeq 0.95$ and $\nu_S \simeq -24\times10^{-3}$.

The tensor power spectrum is found by substituting (2.29) and (2.31) into (2.21). We find

$$\delta_T^2 = \frac{1}{6\pi^2}\frac{M^4}{m_4^4}\Big[1-\Big(\frac{k}{k_0}\Big)^{2\eta}\Big]. \tag{2.37}$$

Therefore, the ratio of tensor to scalar spectrum of fluctuations is, using $\epsilon = \frac{\eta}{e^2}$, precisely

$$\mathcal{R} = 25\epsilon, \tag{2.38}$$

which numerically is $\mathcal{R} \simeq 4.8\times10^{-2}$.

Above we have focused on the most familiar case, where the departure from the slow roll regime is dominated by the quadratic terms in the potential. It may however happen that the inflaton mass scales are smaller than the scales set by the vev, such that the termination of inflationary conditions is controlled by higher polynomial contribution to the inflaton potential. The numerical values we have obtained for (2.33) and (2.34) clearly are sensitive to the precise form of the inflationary potential during the last 60 efolds, and it is of interest to determine the range of these parameters. To do so, one can parameterize different modular inflationary models by the potential function

$$\mathcal{V} = 1 - \left(\frac{\phi}{\mu}\right)^n, \tag{2.39}$$

which yields, in the slow roll regime, the field equations

$$\begin{aligned} H &= \frac{M^2}{\sqrt{3}m_4}\sqrt{1 - \left(\frac{\phi}{\mu}\right)^n} \\ \dot{\phi} &= \frac{nM^2 m_4}{\sqrt{3}\mu^n}\frac{\phi^{n-1}}{\sqrt{1-\left(\frac{\phi}{\mu}\right)^n}}. \end{aligned} \tag{2.40}$$

It is straightforward to find the solution,

$$\begin{aligned} a &= a_{final}e^{-\mathcal{N}} \\ \mathcal{N} &\simeq \frac{\mu^n}{n(n-2)m_4^2\phi^{n-2}} + \frac{(n-4)\mu^2}{2n(n-2)m_4^2}, \end{aligned} \tag{2.41}$$

and determine the density contrast. It is

$$\frac{\delta\rho}{\rho} = \frac{C}{2n\pi\sqrt{3}}\left(\frac{M}{m_4}\right)^2\frac{\phi}{m_4}\left(\frac{\mu}{\phi}\right)^2\left[1 - \left(\frac{M}{m_4}\right)^n\right]^{3/2}. \tag{2.42}$$

From the COBE normalization $\delta\rho/\rho|_{50} \sim 5\times10^{-5}$ we can derive an estimate of the scale of inflation. While there is some sensitivity to the initial condition, we find $M \sim 4\times10^{-2}n^{1/2}(n-2)^{1/4}m_4$, or

$$H \sim \text{few} \times 10^{14} n\sqrt{n-2}\ \text{GeV}. \tag{2.43}$$

But in light of the bound $H < 7\times10^{13}$ GeV on the Hubble scale of inflation [28], we see that the modular inflation models with $n > 2$ are excluded already, and we can ignore them henceforth.

In some cases, most of the late inflationary expansion can occur during the final approach of the inflaton to the minimum of the potential; this the the scenario of chaotic inflation [29]. In these cases, the potential is

$$V = \frac{\lambda}{n}\phi^n, \tag{2.44}$$

where V is the dimensionful quantity $V = M^4\mathcal{V}$. In the slow roll approximation the field equations reduce to

$$\begin{aligned} H &= \sqrt{\frac{\lambda}{3n}}\frac{\phi^{n/2}}{m_4} \\ \dot{\phi} &= -\sqrt{\frac{n\lambda}{3}}m_4\phi^{n/2-1}. \end{aligned} \tag{2.45}$$

The slow-roll solution is

$$\begin{aligned} a &= a_{final}e^{-\mathcal{N}} \\ \mathcal{N} &\simeq \frac{1}{2n}\Big(\frac{\phi}{m_4}\Big)^2. \end{aligned} \tag{2.46}$$

The density contrast is

$$\frac{\delta\rho}{\rho} \simeq \frac{\sqrt{\lambda}}{2\pi\sqrt{3}n^{3/2}}\frac{\phi^{n/2+1}}{m_4^3}, \tag{2.47}$$

and using the COBE normalization we can straightforwardly determine H during inflation. It is

$$H \simeq 4\pi \times 10^{-6}\sqrt{n}\; m_4, \tag{2.48}$$

which is a factor of $\sqrt{n/2}$ higher than the corresponding value in the case of quadratic (sub)leading potential. It is straightforward to determine the spectral index for scalar perturbations. It is

$$n_S = 1 - \frac{n+2}{2\mathcal{N}}\ . \tag{2.49}$$

Hence, in general, chaotic inflationary models driven by higher polynomial terms tend to yield a higher value of H during inflation, but they also give steeper potentials and therefore will yield larger values of the spectral index. The slow-roll parameters for chaotic inflation are

$$\begin{aligned} \eta &\simeq -\frac{n(n-2)}{2}\ \frac{m_4^2}{\phi^2} \\ \epsilon &\simeq \frac{n^2}{2}\ \frac{m_4^2}{\phi^2} \sim \eta. \end{aligned} \tag{2.50}$$

The ratio of tensor to scalar perturbation power obeys (2.22), $\mathcal{R} = 25\epsilon$, by virtue of (2.20),(2.21) and (2.50). For low powers n, the parameter ϵ now determines the duration of inflation, which therefore means that the ratio $\mathcal{R}$ is only weakly sensitive to the specifics of the potential, giving similar tensor power for different forms of V.

3. Imprint of Heavy States

We now turn to the heart of our work–finding the imprint of new, heavy physics on the fluctuations discussed in the previous section. We will assume that the scale of inflation H is much smaller than the Planck mass, $H \ll m_4$, so that a field theoretic treatment of gravity is appropriate. Further, we will assume that the mass scale of new physics M is much larger than H, $M \gg H$, and then assume that we can represent the effects of this new physics at the scale H by "integrating it out" and writing an effective field theory for the inflaton field. These assumptions–which rely on low energy locality and renormalization group ideas–are obviously correct in a field theoretic context, are obviously correct within string perturbation theory around supersymmetric vacua, and are correct in the known nonperturbative definitions of string and M theory in supersymmetric backgrounds. The enduring mystery of the cosmological constant, and the associated mysteries of string theory in de Sitter space, make these assumptions plausible, but not ironclad, in the present context. We make them anyway.[2]

We then can encode all the new physics by writing an effective field theory for ϕ at the scale H. The scale H is appropriate since, as we see in (2.9), that is where we evaluate inflaton correlation functions to compute the size of $\delta\rho/\rho$.

Instead of writing a fully covariant effective action for ϕ, let us just note that the curvature of de Sitter space is proportional to H^2, and so we use this as an additional dimensionful parameter in constructing terms. The interactions of the inflaton must always be very weak to give phenomenologically acceptable values of $\delta\rho/\rho$. This is usually enforced in specific models by some combination of fine tuning, dynamics and supersymmetry (broken at scale H). So we will ignore inflaton interactions. Given these considerations the most general Euclidean local action one can write down is of the form (we have assumed $p \gg H$ and used flat space notation for simplicity):

$$\begin{aligned} S_{eff}[\phi] = \int d^4p\ \phi(p)\phi(-p)\{&p^2/2 + H^2/2 + c_0 H^2(H^2/M^2) + c_1 p^2(H^2/M^2) + c_2 p^4/M^2 \\ &+ c_3 p^4/M^2(H^2/M^2) + c_4 p^6/M^4 + \ldots\}. \end{aligned} \tag{3.1}$$

[2] For an example of a speculation on how locality might break down in de Sitter space string theory see [30].

This structure follows from the fact that only even powers of momenta are allowed, and that the curvature is $\sim H^2$. Therefore, no odd powers of M can appear.

Information about new physics is contained in the coefficients c_i and in the scale M. From (2.9) we see that measurements of $\delta\rho/\rho$ help determine the value of $\langle\phi(p)\phi(-p)\rangle$ at $p = H$. From (3.1) it follows that

$$\begin{aligned}\langle\phi(p)\phi(-p)\rangle|_{p=H} = H^2 + c_0 H^2(H^2/M^2) + c_1 H^2(H^2/M^2) + c_2 H^4/M^2 \\ + c_3 H^2(H^2/M^2)^2 + c_4 H^2(H^2/M^2)^2 + \ldots\end{aligned} \tag{3.2}$$

The large M corrections to $\langle\phi(p)\phi(-(p))\rangle|_{p=H}$ organize themselves into a power series in the dimensionless ratio $r = H^2/M^2$. We have assumed that this ratio is small, so the only terms that are potentially observable are the ones with coefficients c_1 and c_2. The term with coefficient c_0 is just a renormalization of the potential which we can ignore.

On very general grounds the effect of new physics, whether field theoretic, string theoretic, M theoretic, etc. is proportional to $r = H^2/M^2$. The coefficients c_i must be computed, however, and can be much smaller than one, giving effects much smaller than the naive expectation.

Several groups [8,9,10,11,12] have previously analyzed a special case of this situation. They have added an irrelevant operator to Einstein gravity and directly computed its effect on inflationary fluctuations by solving the linearized wave equations. This requires specifying new boundary conditions at high momentum on the higher order differential equation. These boundary conditions are not determined by the model itself. The authors in [10] impose the constraint that the solutions rapidly relax to the "adiabatic" vacuum shortly after they are created. They find imprints of size $r \sim H^2/M^2$, consistent with our general result. In [11,12] the authors study the general boundary condition and then focus on a special, different[3] boundary condition that results in effects of size $\sim (r)^n,\ n \simeq .5$. This effect is inconsistent with our effective action result and so presumably this boundary condition violates locality in some way. We should also note that an effect of this functional form would imply a nonanalytic dependence on g_s^2 and on α' which would signal the breakdown of perturbation theory at weak coupling. All in all it seems likely to us that the special boundary condition chosen in [11,12] is unphysical. The subtleties mentioned

[3] The authors in [10] speculated that this boundary condition was the same as their adiabatic condition. The results of [12] show this is not the case.

at the beginning of this section make it impossible to definitively rule out such a result, however.

The above illustrates the virtue of the effective action approach we are using here. Equation (2.9) shows that the relevant momentum scale for these processes is H, not m_4. If there is a large hierarchy between these scales–which is the situation we are envisioning– then there should be no reason to consider Planckian dynamics at all, e.g., short distance boundary conditions, in studying the fluctuation problem. One simply encapsulates all the unknown short distance physics in an effective action. All the subtlety of choice of boundary condition is buried in the assumption of the existence of a local effective action. Given that our world appears to be local, this seems an excellent assumption.

Perhaps we should phrase things in a more optimistic way. If experiments detect imprints in the CMBR of strength $r^{.5}$ as predicted in [11,12] it would imply a breakdown of locality in low energy string theory, which might be a crucial clue in solving the cosmological constant problem! Alternatively, such results could also indicate that physics other than inflation may be responsible for the origin of structure in the universe.

We now turn to the evaluation of the parameters in the effective action (3.1) in some specific physical situations. First imagine a heavy fermion field ψ coupled to the inflaton via a Yukawa interaction $\lambda\phi\bar{\psi}\psi$. A one loop graph in de Sitter space of ψ particles clearly induces interactions of the form in (3.1) . These produce effects in the propagator (3.2) of size $\sim \lambda^2 H^2/M^2$ with $M = m_\psi$ the fermion mass. Typically $m_\psi \sim \lambda\langle\phi\rangle \sim \lambda m_4$. (We ignore slow roll parameters here.) So these effects are $\sim H^2/m_4^2 \sim 10^{-11}$ and hence unobservable. This result is quite general. A particle renormalizably coupled to the inflaton will *generically* have a mass $\sim \langle\phi\rangle \sim m_4$ and so the virtual effects of this particle will be of order H^2/m_4^2: unobservably small.[4] Exceptions to this result can occur if counterterms are fine tuned to make the particle masses unnaturally small. Then the virtual effects can be very large and certainly observable. An extreme case of this limit has been studied in [7] where a fermion becomes massless for a certain value of the inflaton field vev. When this vev is reached during the slow roll, fermions are produced copiously, sharply reducing $\dot{\phi}$ and so, by (2.18), creating a sharp increase in $\delta\rho/\rho$ for a short time. This translates into a sharp peak in momentum space in the fluctuation spectrum.

These phenomena require an additional level of fine tuning on top of any fine tuning required to make the inflaton potential well behaved.

[4] We thank S. Thomas for pointing this out to us.

Next we turn to weakly coupled heterotic string theory models of the "traditional" type: $g_s^2 \sim .1$, $m_s \sim m_4 \sim 10^{19}$ GeV. The Calabi-Yau compactification radii are hence also of order $1/m_s$. It will be useful to be able to vary this scale independently, so we will denote it $1/m_{CY}$. We do not understand inflation in string theory, or string theory in de Sitter space. If we assume the existence of an effective action in these environments, though, we can compute by evaluating terms in the effective action from the string theory S matrix in flat space ($H = 0$).

These models have four real supercharges and hence have no constraints on the kinetic term in the four dimensional effective action. On the other hand, sixteen or more supercharges would require no renormalization of the kinetic term. So as $m_{CY} \to 0$ and flat ten dimensional space is recovered, the higher derivative terms in the effective action must vanish. So we expect effects in the propagator (3.2) of size $m_{CY}^2 H^2/m_s^4$. For $m_{CY} \sim m_s$ this becomes $H^2/m_s^2 \sim 10^{-11}$. This is unobservable.

To find effects closer to the threshold of detectability we must enter the realm of strongly coupled string theory, where the fundamental mass scale can be much less than m_4.

4. Large effects in string and M-theory

We have shown that new physics at a scale M leads to the following expression for quantum fluctuations of the inflaton:

$$\langle \delta\phi^2 \rangle = \frac{H^2}{4\pi^2}\left(1 + \mathcal{X}\frac{H^2}{M^2} + \cdots\right) \ . \tag{4.1}$$

The second term in brackets is the leading correction to the standard, free-field expression used in inflationary cosmology. $\mathcal{X}$ is a model-dependent, dimensionless number related to the coefficients in the effective action (3.1). It may get contributions from phase space factors in loop integrals, sums over heavy particles coupling to the inflaton, and so on.

As we will argue in the next section, this correction is potentially observable as a correction to a well-known consistency condition on the tensor and scalar fluctuations of the CMBR. We believe such an effect is measurable in principle if

$$\mathcal{X}\frac{H^2}{M^2} \sim 0.1 - 1 \ . \tag{4.2}$$

It is hard to be more precise with this number, as it depends on the measurability of the B-mode polarization, which is not yet well understood.

The Hubble constant H^2 can be calculated, or hopefully measured in polarization experiments. The current upper bound from COBE, current degree-scale anisotropy experiments, and large-scale structure data is $H = 7 \times 10^{13}$ GeV [28]. In 4D GUT models, $M = m_4$, $\mathcal{X} \ll 1$, and the correction is unobservable.[5] However, in most phenomenologically viable string and M-theory models, the fundamental scale M_f – either the higher-dimensional Planck scale or the string scale – is lower than the $4D$ Planck scale m_4 by up to two orders of magnitude [31,32,13]. If we compactify a d-dimensional theory with Planck scale M_f on a $(d-4)$-dimensional manifold X_{d-4} with volume V_{d-4}, then $m_4^2 = M_f^{d-2} V_{d-4}$. The high scale m_4 is not a dynamical scale, but rather an artifact of the large volume of the compactification manifold.

In these models we might expect $M = M_f$. However, so long as the Hubble scale is lower than the compactification scale, $4D$ effective field theory still applies. The effect on (4.1) of integrating out a given four-dimensional field still gives $M = m_4, \mathcal{X} < 1$.

But higher-dimensional models have several new features which can significantly enhance the corrections to (4.1). First, the corrections in (4.1) arise from nonrenormalizable gravitational couplings which become large at high energies. Thus high-scale physics–in particular the large numbers of particles above the Kaluza-Klein threshold–contributes significantly in loops. Secondly, the existence of tensor fields in 10- and 11-dimensional models leads to a large factor $\mathcal{X}$ from summing over polarizations of these fields.

In almost all of the models we are interested in, the dominant effects arise from supergravity modes. The loop integrals appear highly divergent; but for the effects we are calculating they are cut off by either the restoration of maximal supersymmetry (to 16 or 32 unbroken supercharges), or by the soft ultraviolet behavior of the fundamental theory. The result is highly model-dependent, and the numbers we arrive at by no means constitute a precise prediction. Nonetheless we can estimate whether the correction in (4.1) is observable. To that end, we will begin this section by estimating $\mathcal{X} H^2/M^2$ as a function of the compactification radii and the cutoff. We will then analyze a variety of supersymmetric $N = 1, d = 4$ models in string and M-theory and estimate the size of

[5] Models for which the particle coupling to the inflaton becomes light some $\phi = \phi_0$ during the inflationary epoch [7] lead to an observable effect at a particular angular scale on the sky; this will be observationally distinct from the effects we discuss here.

the one-loop contribution to (4.1) in each. Readers who are less theoretically inclined (or simply impatient) will find the results of this section in the following paragraphs above §4.1; they may then skip to §5 where the experimental consequences are discussed.

Summary of results

Our strategy will be as follows. The first two models we analyze – M-theory on $X_7 = X_6 \times S^1/\mathbb{Z}_2$ [33], also known as Hořava-Witten theory, and M-theory on a manifold of G_2 holonomy – can be made consistent with the unification prediction of [34], by keeping all scales including the eleven-dimensional Planck scale to within an order of magnitude of the unification scale, $M_{GUT} = 2\times 10^{16}$ GeV. In this case we will find that $\mathcal{X}H^2/M^2 \sim 10^{-7}$, which is unobservably small.

However, if we give up perturbative unification, we can increase the volume of the compactification manifold and decrease the fundamental Planck scale. We examine such models under three constraints. First, the inflationary dynamics must remain four-dimensional. This puts an upper limit on the size of the compactification manifold, on the order of $1/H$. Secondly, if we assume that the energy density during inflation is constant in d dimensions, then it must be lower than m_d^d, where m_d is the d-dimensional (reduced) Planck scale. We will find that this also places an upper limit on the compactification volume.

Finally, the four-dimensional gauge coupling must remain $\alpha \sim 1/25$ in order that the standard model couplings are roughly correct at a TeV. The origin of gauge dynamics in a given model, combined with the constraint on α, affects how many dimensions can be made large. In M- and F-theory models, gauge dynamics arises on singularities or on branes, both at finite codimension in the compactification manifold. If the singularity or brane lies on a k-dimensional submanifold $\Sigma \subset X$, $\alpha = V_\Sigma M_f^k$ and V_Σ is fixed. The number of dimensions which may be made large is then the codimension $d-k$ of the brane or singularity, so the models with the greatest chance of giving rise to observable corrections in (4.1) are those with the highest $d-k$.

In M-theory on $X_6 \times S^1/\mathbb{Z}_2$ the gauge dynamics occurs on the boundaries of $S^1/\mathbb{Z}_2$ which have codimension one. The volume of X_6 is constrained, and we cannot decrease the size of the interval $S^1/\mathbb{Z}_2$ low enough to make the correction term in (4.1) observable. Manifolds with G_2 holonomy are in much better shape. The gauge dynamics lies on singularities of codimension four [15]. We can increase the volume of the transverse manifold such that the eleven-dimensional Planck scale is $m_{11} \sim H$.

We then move to ten-dimensional type I models. In the simplest such models the gauge degrees of freedom propagate in ten dimensions. The compactification manifold can be made large consistently with $\alpha = 1/25$ by adjusting the string coupling. But this coupling is weak, so that ten-dimensional physics is controlled by very soft string physics and the correction in (4.1) is unobservable.

One may also study models for which the gauge degrees of freedom propagate on branes. Two such models consistent with $\mathcal{N} = 1$ supersymmetry in four dimensions are Hořava-Witten models with the gauge dynamics arising on M5-branes wrapped on Riemann surfaces, and F-theory models with the gauge dynamics arising on D3-branes. For both of these models, the strongest constraint is that imposed by sub-Planckian energy densities. Up to the model-dependent factor $\mathcal{X}$, the constraints on the corrections in (4.1) lead to estimates for $\mathcal{X}H^2/M^2$ that are within a factor of a few of the estimate for manifolds of G_2 holonomy, so we will not discuss these other models further.

All of these estimates are model-dependent and imprecise. In particular, we will argue below that the loop expansion is starting to break down as the corrections in (4.1) start to become observable. It is easy to imagine these effects changing our estimates by an order of magnitude in more precise calculations. For this reason we parameterize our results in terms of H^2/M^2 and $\mathcal{X}$ separately.

4.1. Notation

First we specify our notation: the dimensionful gravitational coupling κ_d is the coefficient of the Einstein term:

$$\mathcal{L} = \frac{1}{2\kappa_d^2}\int_{X_d} d^dx\sqrt{g}R\ . \tag{4.3}$$

We define two versions of the Planck mass (differing by a numerical factor):

$$2\kappa_d^2 = (2\pi)^8\ell_d^9\ = \tilde{\ell}_d^9; \qquad M_d = 1/\ell_d\ , m_d = 1/\tilde{\ell}_d\ . \tag{4.4}$$

When we compactify on a manifold X_{d-4} with volume V_{d-4}, the four-dimensional Planck scale is:

$$m_4^2 = 2m_d^{d-2}V_{d-4}\ . \tag{4.5}$$

In ten-dimensional string theories the gravitational coupling can be written via the string scale as:

$$2\kappa_{10}^2 = g_s^2\alpha'^4 = \frac{g_s^2}{m_s^8} = g_s^2(2\pi)^7\ell_s^8 = g_s^2\frac{(2\pi)^7}{M_s^8}\ . \tag{4.6}$$

The string tension is $T = 1/2\pi\alpha'$ and a string oscillator mode carries energy $1/\sqrt{\alpha'}$.

4.2. Corrections to the propagator in higher-dimensional theories

Ideally we could choose a string model and simply calculate the one-loop corrections to (4.1) in perturbative string theory. However, string theory in approximately de Sitter backgrounds is poorly understood. Furthermore, we will find that the effects of high scale physics are closest to observability in M- and F-theory models.

However, supergravity remains a good approximation in the calculations we are interested in. We will begin by simply studying a scalar field coupled to a d-dimensional graviton on a $(d-4)$-dimensional torus. This may seem nonsensical as the loop integrals will be badly divergent. However, the corrections to the p^2 and p^4 terms in the propagator vanish in supersymmetric theories when 16 or 32 supercharges are unbroken, at energies above the compactification scale. Therefore, supersymmetry cuts off the otherwise highly divergent amplitudes without our needing to appeal to the ultraviolet physics of M- or F-theory. The scale of the cutoff will be set by the scale at which the full supersymmetry of the underlying theory is restored.[6]

Because the loop integrals are dominated by energies near the cutoff, well above the compactification scale, we can ignore the effects of the curvature and topology of X_d. We will therefore estimate the correction in (4.1) by coupling the inflaton to the d-dimensional graviton on a rectangular $(d-4)$-dimensional torus. For our purposes, the effects of the actual geometry can be summarized in terms of the model-dependence of $\mathcal{X}$ in (4.1).

The Lagrangian for a massive scalar minimally coupled to the d-dimensional graviton is

$$S = \frac{1}{2}\int d^d x\sqrt{g}\left(g^{ab}\partial_a\phi\partial_b\phi + m^2\phi^2\right) \ . \tag{4.7}$$

The metric g can be written in terms of the background metric η (which we take to be flat) and a small fluctuation:

$$g_{ab} = \eta_{ab} + h_{ab} \ . \tag{4.8}$$

S can be expanded in powers of h using the formulae

$$\begin{aligned} \delta\sqrt{g} &= \frac{1}{2}\sqrt{g}g^{ab}\delta g_{ab} \\ \delta g^{ab} &= -g^{ac}g^{bd}\delta g_{bd} \ . \end{aligned} \tag{4.9}$$

[6] Of course supersymmetry is also broken by the vacuum energy. However it is restored for momenta $k > H$. The corrections we will discuss will arise from momenta much larger than H.

This will lead to nonrenormalizable couplings of the form $h(\partial\phi)^2$ and $h^2(\partial\phi)^2$:

$$\delta S = \frac{1}{2}\int d^dx\,\big(T_3^{ab,mn}h_{ab}\partial_m\phi\partial_n\phi + T_4^{abcd,mn}h_{ab}h_{cd}\partial_m\phi\partial_n\phi \\ +M_3^{ab}h_{ab}\phi^2 + M_4^{abcd}h_{ab}h_{cd}\phi^2\big)\ , \tag{4.10}$$

where

$$\begin{aligned} T_3^{ab,mn} &= \frac{1}{2}\eta^{ab}\eta^{mn} - \eta^{ma}\eta^{nb} \\ T_4^{abcd,mn} &= -\frac{1}{2}\eta^{ab}\eta^{mc}\eta^{nd} + \frac{1}{8}\eta^{ab}\eta^{cd}\eta^{mn} - \frac{1}{4}\eta^{ac}\eta^{bd}\eta^{mn} + \frac{1}{2}\eta^{mc}\eta^{ad}\eta^{nb} + \frac{1}{2}\eta^{ma}\eta^{nc}\eta^{bd} \\ M_3^{ab} &= \frac{1}{2}\ m^2\eta^{ab} \\ M_4^{abcd} &= \frac{1}{8}m^2\eta^{ab}\eta^{cd} - \frac{1}{4}\ m^2\eta^{ac}\eta^{bd}\ . \end{aligned} \tag{4.11}$$

The propagator for h is, in de Donder gauge:

$$\langle h_{ab}h_{cd}\rangle = \frac{1}{m_4^2k^2}\left(\eta_{ac}\eta_{bd} + \eta_{ad}\eta_{bc} - \frac{2}{d-2}\eta_{ab}\eta_{cd}\right). \tag{4.12}$$

The two one-loop diagrams are shown in Fig. 2. The diagram on the left will contribute both wavefunction renormalization terms and p^4/Λ^2 corrections to the propagator, while the right-hand diagram will give further wavefunction renormalization corrections.

Fig. 2: Two one-loop diagrams important for the computation.

We are interested in the divergent part of the loops with loop momenta of order $k \gg H$. Therefore, we can approximate the H-dependence of the propagators at tree level via the first two terms in (3.1), which amounts to shifting all of the masses by $m^2 \to m^2 - 2H^2$. At the end we will take the leading correction in H^2/Λ^2.

The left-hand diagram leads to the correction:

$$D_1(p) = \frac{1}{m_4^2} \sum_n \int \frac{d^4k}{(2\pi)^4} \frac{p^2(p-k_n)^2 + p\cdot(p-k_n)m^2 + m^4}{(k_n^2 + m^2 - 2H^2)\,((p-k_n)^2 - 2H^2)} \tag{4.13}$$

and the right-hand diagram to the correction:

$$D_2(p) = -\frac{1}{4}(d^2+d-8)\frac{1}{m_4^2} \sum_n \int \frac{d^4k}{(2\pi)^4} \frac{(p^2+m^2)}{k_n^2 - 2H^2} \ . \tag{4.14}$$

The sum over n is over Kaluza-Klein momenta, and k_n denotes the full eleven-dimensional momentum of the internal graviton propagator.

The four-dimensional integrals in (4.13),(4.14) are already quadratically divergent, and the Kaluza-Klein sum only increases the degree of divergence. So long as the cutoff is more than a few times the Kaluza-Klein scale, we can approximate this sum by an integral:

$$\sum_n = \frac{V_{d-4}}{(2\pi)^{d-4}} \int d^{d-4}k \ , \tag{4.15}$$

where V_{d-4} is the radius of the T^{d-4}. The eleven-dimensional momentum integrals are highly divergent and dominated by the UV end of the integral, near the cutoff $\Lambda \gg H$. We can therefore expand the integrand in powers of H^2/k^2. After subtracting the H-independent wavefunction renormalization correction, the most divergent terms in this expansion are:

$$D(p) = -\frac{n_d V_{d-4} H^2(p^2+m^2)}{m_4^2} \int \frac{d^dk}{(2\pi)^d}\frac{1}{k^4} + \frac{V_{d-4}p^4}{m_4^2} \int \frac{d^dk}{(2\pi)^d}\frac{1}{k^4} \ . \tag{4.16}$$

The first term leads to an H-dependent wavefunction renormalization, and the second to a p^4 term in the propagator. The factor n_d arises from the sum over graviton polarizations. Since de Donder gauge is not complete, we must subtract off the ghosts. The result should be the number of physical graviton polarizations, $(d-1)(d-2)/2-1$. In theories with 32 unbroken supercharges before compactification, the graviton supermultiplet contains additional scalars, gauge fields, and tensors, so that when we include all of the bosonic degrees of freedom we will find $n_d = 128$.

The size of (4.16) depends strongly on the cutoff. Naïvely one expects this cutoff to be M_f. If $M_f = m_d$, then

$$V_{d-4} m_d^{d-4} = \frac{m_4^2}{M_d^2} \tag{4.17}$$

and so we can write $M = m_d$ in (4.1). However, the relation between Λ and M_f is model-dependent, and may involve factors of 2π and other dimensionless numbers. Because these factors are raised to high powers, they can have a significant effect on the size of the correction in (4.1). For now we will set

$$\Lambda = c m_d \ , \tag{4.18}$$

with $c \sim \mathcal{O}(1)$ parameterizing the model-dependence.

After performing the momentum integrals in (4.16), we find:

$$D(p) = -\frac{2 n_d \pi^{d/2} c^{d-4}}{(d-4)(2\pi)^d \Gamma\left(\frac{d}{2}\right)} \frac{H^2}{M_f^2}\left(p^2+m^2\right) + \frac{2\pi^{d/2} c^{d-4}}{(d-4)(2\pi)^d \Gamma\left(\frac{d}{2}\right)} \frac{p^4}{m_d^2} \ . \tag{4.19}$$

Since $n_d \sim 100$, the wavefunction renormalization term will dominate, and we will find that the coefficient $\mathcal{X}$ in (4.1) will take the value:

$$\mathcal{X} = \frac{2 n_d \pi^{d/2} c^{d-4}}{(d-4)(2\pi)^d \Gamma\left(\frac{d}{2}\right)}, \tag{4.20}$$

while $M = M_f$.

These estimates are hardly precise. In addition to the model-dependence we have discussed, the loop expansion will begin to break down in models where the corrections in (4.1) are close to observability. For these models $c \geq 1$ in (4.18). If the fundamental scale is $M_f = m_d$, where m_d is the d-dimensional Planck scale, then the dimensionless gravitational coupling governing loop corrections will be:

$$g_{grav}^2 = \left(\frac{\Lambda}{m_d}\right)^{d-2} \ . \tag{4.21}$$

Once $\Lambda \sim m_d$, $g_{grav}^2 \sim 1$. Nonetheless we will assume that the one-loop answer gives a rough estimate of the size of the corrections in (4.1).

We will still try to be careful about numerical factors. This may seem perverse given the above discussion. However, these factors are often raised to high powers, so that they contribute appreciably to our order-of-magnitude estimates.

The remainder of this section will amount to estimates of the magnitude of (4.1) in a variety of string and M-theory models, with these caveats firmly in mind.

4.3. Physical constraints on compactifications

In the minimal supersymmetric standard model, the running strong, weak and electromagnetic couplings unify at

$$\alpha_{GUT} = \frac{g^2}{4\pi} \sim \frac{1}{25} \tag{4.22}$$

at a scale of order $M_{GUT} \sim 2 \times 10^{16}$ GeV [34]. This is strong evidence for grand unification at that scale. Nonetheless it still indicates a small hierarchy between M_{GUT} and m_4.

In traditional string phenomenology, one starts with ten-dimensional type I or heterotic string theories, which have 16 unbroken supercharges. One then chooses a six-dimensional Calabi-Yau manifold X with volume $V_X \sim M_{GUT}^{-6}$, which preserves $N = 1$ SUSY at the compactification scale. M_{GUT}, α_{GUT} and m_4 are computable functions of V_X, the string scale m_s and the string coupling g_s, and one may adjust the compactification parameters in order to match the unification predictions of [34].

For type I models, the measured values of α_{GUT},M_{GUT}, and m_4 can be achieved in models with weak string coupling. For heterotic models, the observed couplings and scales are incompatible with weak string coupling [13]. One may try to work at strong heterotic coupling, but it is not clear that the expressions for the gauge couplings are correct.

Instead we can appeal to string duality [13]. The strong coupling limit of the $SO(32)$ string is weakly coupled type I string theory [35]. The strong coupling limit of the $E_8 \times E_8$ heterotic string compactified on X is M-theory on $X \times S^1/Z_2$ [33]. In this latter limit, gauge coupling unification is compatible with a background well described by 11-dimensional supergravity [13].

We will also study M-theory on a manifold of G_2 holonomy, and weakly coupled type I string models. We will find that Hořava-Witten theory (with the standard model as a subgroup of $E_8 \times E_8$) and weakly coupled type I models do not give rise to observable corrections in (4.1), in any reasonable regimes of parameter space. It appears that for such corrections to be observable, the dynamics must be strongly coupled and the standard model should live on a brane or singularity with high codimension.

In the remainder of this section we will discuss a variety of models which have low-energy gauge dynamics and a fundamental scale lower than m_4, and estimate the size of corrections to (4.1). We will spend the most time on Hořava-Witten models. We will then discuss M theory on manifolds of G_2 holonomy and Hořava-Witten theory type I models.

In models consistent with coupling unification, the correction to (4.1) will turn out to be too small to be observed. We will therefore examine a wider class of models under the following constraints. First, the four-dimensional Planck scale must be that given by experiment. Secondly, although we have given up coupling unification, the gauge coupling at the fundamental scale must be on the order of $\alpha = 1/25$, to get roughly the correct standard model couplings at a TeV. Thirdly, we will demand that inflationary dynamics be truly four-dimensional. The upper limit on the compactification volume is set by demanding that the Kaluza-Klein momenta be larger than the deSitter temperature, $T_{dS} = H/2\pi$, so that the dynamics of quantum inflaton fluctuations remains four-dimensional. If we imagine compactification on a circle with circumference L, this condition means that $L < (2\pi)^2/H$. For a manifold X_k with volume V_k, we take this to mean that

$$V_k \leq \frac{(2\pi)^{2k}}{H^k} \; . \tag{4.23}$$

Finally, we demand that the d-dimensional energy density be sub-Planckian. Let us assume that the energy density responsible for inflation is constant over the compactification manifold X_{d-4}. Denoting the k-dimensional energy density by $E^k_{(k)}$,

$$E^4_{(4)} = 3H^2 m_4^2 = E^d_{(d)} V_{d-4} \; , \tag{4.24}$$

which implies

$$\left(\frac{H}{m_d}\right)^2 = \frac{1}{3}\left(\frac{E_{(d)}}{m_d}\right)^d \; . \tag{4.25}$$

Therefore we demand[7]

$$\left(\frac{H}{m_d}\right)^2 \leq 1 \; . \tag{4.26}$$

Since m_4 is fixed, Eq. (4.5) ties a lower limit on m_d to an upper limit on V_{d-4}. Depending on the model at hand, this bound may be more or less stringent than (4.23).

In our study of perturbative type I models, we will also demand that the energy density $E^{10}_{(10)} \leq m_s^{10}$. At higher energy densities stringy physics is not understood.

We will find that for models which give measureable correction terms in (4.1), the fundamental scale $m_d \sim H \sim 7 \times 10^{13}$ GeV. With such a low scale we have to worry again about proton decay. In GUT models dimension-six operators suppressed by $1/M^2_{GUT}$ lead to proton lifetimes close to the experimental lower bound, close enough to model-dependent factors to rule out models. Dimension-six operators suppressed by $1/H^2$ will

[7] We ignore the factor of 1/3; it disappears if we allow e.g. $E_{(d)} = 1.2 m_d$ which we cannot rule out at this crude level.

lead to proton decay which is 10 or 11 orders of magnitude more rapid than if they were suppressed by $1/M_{GUT}^2$. If one is able to forbid operators below dimension seven, then higher-dimensional operators suppressed by powers of $1/H$ will lead to phenomenologically acceptable lifetimes. One could achieve this, for example, if some discrete subgroup of the $U(1)$ baryon number symmetry was gauged, along the lines of [36,37]. Since we are not studying our models in detail we will leave this issue aside.

4.4. The Hořava-Witten model

Compactifications of M-theory on $X_{11} = X_{10} \times S^1/Z_2$ were the first known M-theory models with chiral gauge dynamics [33]. These models can be described relatively explicitly, so we will spend the greatest amount of time on them. In addition, in models which realize $N = 1$ supersymmetry, the explicit pattern of supersymmetry breaking means that moduli of the compactification manifold are good inflaton candidates [38,39] as we will review below.

If $X_{10} = \mathbb{R}^4 \times X_6$ and X_6 is Calabi-Yau, the theory has four unbroken supercharges in four dimensions. One E_8 gauge multiplet is localized on each end of the interval. The gauge couplings are:

$$\sum_{i=1}^{2} \frac{1}{8\pi \left(4\pi\kappa_{11}^2\right)^{2/3}} \int_{M_{10,i}} \sqrt{g} F_i^2 \tag{4.27}$$

where the sum is over the two boundary components. Upon compactification on X, anomaly cancellation will require gauge field configurations which break this gauge group further; generally one breaks one of the E_8 groups to the GUT group and then to the standard model gauge group, while the other E_8 is the gauge symmetry of a hidden sector.

Without going into great detail, we can see that these models can match the predicted coupling unification in a regime where all scales, including the fundamental scale, are close to M_{GUT} and supergravity is valid.

The GUT group is broken to the standard model gauge group by visible sector gauge field configurations on X – *c.f.* [40] for a discussion. Therefore we let $L_{CY} = V_X^{1/6} = M_{GUT}^{-1}$. Newton's constant G_N and the GUT coupling α_{GUT} can be written as [13]:

$$\begin{aligned} \frac{1}{8\pi G_N} &= \frac{V_X L_{11}}{\kappa_{11}^2} \\ \alpha_{GUT} &= \frac{g_{GUT}^2}{4\pi} = \frac{\left(4\pi\kappa_{11}^2\right)^{2/3}}{2V_X} \end{aligned} \tag{4.28}$$

With the above values of α_{GUT} and M_{GUT}, we find:

$$\begin{aligned} m_{11} &\sim 2M_{GUT} \\ M_{11} &\sim 10M_{GUT} \\ \frac{1}{L_{11}} &\sim 0.01M_{GUT} \ . \end{aligned} \tag{4.29}$$

Therefore although the heterotic coupling is strong, this compactification is well described by $11d$ supergravity. Note that we do not have to postulate large hierarchies between the GUT and fundamental scales. The largest hierarchy is between M_{11} and $1/L_{11}$. If we take the ratio between M_{11} and the mass gap of the Kaluza-Klein excitations with momentum along $S^1/\mathbb{Z}_2$:

$$m_{11}/m_{KK} = m_{11}L_{11}/\pi \sim 60 \ , \tag{4.30}$$

so the Kaluza-Klein scale is about an order of magnitude off from the GUT scale.

The expansion parameter in these models is $(2\kappa_{11}^2)^{2/3}/V_X$. The assumption that the geometry is a simple product $X \times S^1/\mathbb{Z}_2$ holds only at lowest order. To next order in our expansion parameter the product is warped; V_X depends on the coordinate x_{11} along $S^1/\mathbb{Z}_2$ [13]. A natural size for L_{11} is that for which the volume vanishes at the end of the interval where the hidden sector gauge group resides. One can then imagine strong coupling effects leading to supersymmetry breaking and the stabilization of moduli [13,41]. L_{11} determined this way depends on the topology of the $E_8 \times E_8$ gauge field configurations on X, and on the Kähler moduli of X. For reasonable choices of both, L_{11} is consistent with (4.29).

In this model maximal supersymmetry is broken to $N = 1$ in ten dimensions at the fixed points of $S^1/\mathbb{Z}_2$, and then to $N = 1$ in $d = 4$ by the compactification on X_6. The cutoff in (4.15) should be roughly $V_X^{-1/6}$, so we might expect the cutoff to be on the order of $1/V_X^{1/6} \sim M_{GUT}$. Again, the precise value of Λ is highly model-dependent. In a sufficiently anisotropic Calabi-Yau we can raise this scale. We will take it to be the fundamental UV cutoff that quantum-mechanical M-theory is expected to provide.

This cutoff can be estimated by studying four-graviton scattering at one loop [42]; since the amplitude is protected by supersymmetry, it can be calculated in string theory and extrapolated to strong coupling. The computation is cutoff dependent in supergravity. If we define the cutoff Λ_{11} by matching the supergravity result to the finite M-theory result, then [42]:

$$\Lambda_{11} = 2^{4/9}\pi^{11/9}m_{11} \sim 5m_{11} \ . \tag{4.31}$$

Inflaton dynamics in the Hořava-Witten model

In the Hořava-Witten models, the moduli of X_6 are natural inflaton candidates. A simple argument due to Banks [38,39][8] shows that the pattern of supersymmetry breaking in Hořava-Witten models can lead to an inflaton potential with the right properties. M-theory compactified on a Calabi-Yau threefold X has eight supercharges, and the moduli of X are exactly flat directions, protected by supersymmetry. Upon further compactification on $S^1/\mathbb{Z}_2$, supersymmetry is broken to $N = 1$ in four dimensions *at the boundaries of the interval.* Superpotentials for the moduli of X can arise only on the boundaries.

Let $\tilde{\phi}^A$ be the (complex) moduli of X in M-theory, describing sizes of various cycles of X in units of M_{11}. In four dimensions the kinetic term is:

$$\mathcal{L}_{kin} = \frac{1}{2\kappa_{11}^2} V_X L_{11} \int d^4x G_{A\bar{B}}(\tilde{\phi}) \partial\tilde{\phi}^A \partial\bar{\tilde{\phi}}^{\bar{B}} \ . \tag{4.32}$$

where G_{AB} is the dimensionless metric on the moduli space of X. The factor in front of the integral also multiplies the 4-dimensional Einstein term, which is expected since the moduli are simply components of the metric in X. This is just the "reduced" $4d$ Planck mass m_4.

The canonically normalized scalar fields in four dimensions are:

$$\phi^A = m_4 \tilde{\phi}^A \ . \tag{4.33}$$

In $N = 1$ language, G is the derivative of the Kähler potential

$$G_{A\bar{B}} = \partial_A \partial_{\bar{B}} K \tag{4.34}$$

where the derivatives are with respect to the canonically normalized fields. The fact that G is dimensionless and of order one means that we can write K in terms of a dimensionless order one potential $\tilde{K}$:

$$K = m_4^2 \tilde{K} \tag{4.35}$$

so that

$$G_{A\bar{B}} = \tilde{\partial}_A \tilde{\partial}_B \tilde{K} \tag{4.36}$$

where $\tilde{\partial}$ is a the derivative with respect to $\tilde{\phi}$.

[8] With many caveats, extensively discussed in [39].

$N=2$ SUSY is broken to $N=1$ by the boundaries of $S^1/\mathbb{Z}_2$. Fundamental physics on these boundaries is still controlled by m_{11}, so that the superpotential will have the form:

$$\mathcal{L}_{super} = m_{11}^3 \int d^2\theta d^4 x w(\tilde{\phi}) + \text{h.c.} \tag{4.37}$$

The bosonic potential in $N=1$ supergravity arising from this superpotential is:

$$V(\phi) = e^{\tilde{K}} \frac{m_{11}^6}{m_4^2} \left(G^{A\bar{B}} \tilde{D}_A w \tilde{D}_{\bar{B}} \bar{w} - 3|w|^2 \right) = \frac{m_{11}^6}{m_4^2} \mathcal{V}\left(\frac{\phi}{m_4}\right) \equiv M^4 \mathcal{V}, \tag{4.38}$$

where

$$\tilde{D}_A w = \tilde{\partial}_A w + \tilde{\partial}_A \tilde{K} w \ .$$

For a successful model of inflation, ϕ must roll slowly for approximately 60 e-foldings, and the fluctuations in ϕ must generate the density perturbations measured by COBE, $\delta\rho/\rho \sim 5\times 10^{-5}$. We can use this requirement to compute M [18]. We will choose our coordinates so that a single coordinate $\tilde{\phi}$ parameterizes the trajectory in the moduli space travelled during the inflationary epoch.

We rewrite (2.27) in the present context:

$$N_e = \frac{1}{2m_4^2} \int_{\phi_e}^{\phi} \frac{V}{\partial_\phi V} \sim 60 \ , \tag{4.39}$$

using the slow-roll expression

$$H^2 = \frac{V}{3m_4^2} \ .$$

Here ϕ_e is the vev of the inflaton at the end of inflation, and ϕ the vev 60 efoldings prior to that. Assuming V does not change much during inflation we can approximate (4.39) by:

$$\frac{\phi-\phi_e}{m_4^2} \frac{\mathcal{V}}{\partial_\phi \mathcal{V}} \sim 60 \ . \tag{4.40}$$

If we let $(\phi-\phi_e)\sim m_4$ and solve for $\mathcal{V}/\partial_\phi\mathcal{V}$, we can use (2.18) to solve for M and M_{11}:

$$\begin{aligned} M &\sim 3\times 10^{-3}\ m_4 = 7\times 10^{15}\text{GeV} \\ m_{11} &\sim 5\times 10^{16}\text{GeV} \ . \end{aligned} \tag{4.41}$$

M is close to the unification scale M_{GUT}, and the value of m_{11} predicted here is close to that in (4.29). Within our crude set of approximations we can take this as an estimate of m_{11} independent of (4.29).

Since these numbers are rough estimates, we will use the experimentally determined upper bound H in our estimates of H^2/M^2.

Corrections to the inflaton propagator

Eq. (4.31) implies that $c = 2^{4/9}\pi^{11/9}$ in (4.20). The correction in (4.1) is determined by: (here $d = 11$, $n_b = 128$)

$$M^2 \sim m_{11}^2 \ , \qquad \mathcal{X} \sim 0.1 \ . \tag{4.42}$$

Using the experimental upper bound on H, and m_{11} as given in (4.29),

$$\frac{H^2}{M^2} \sim 10^{-6} \tag{4.43}$$

While this is better than the result expected from four-dimensional GUT models, it is not close to observable. If we were willing to give up unification at M_{GUT} and require only that the gauge couplings satisfy $\alpha \sim 1/25$ and that the inflationary dynamics be four-dimensional, we can have a smaller value of m_{11} and the corrections in (4.1) will be larger. (If we push these constraints to their limits, the eleven-dimensional energy density is still sub-Planckian). Note that for such models the arguments in [38,39] will cease to generate inflaton potentials with $\mathcal{M} \sim M_{GUT}$, as we must push $m_{11} < M_{GUT}$ for corrections in (4.1) to be observable. We will have to assume that such potentials are generated by four-dimensional gauge theory effects.

The correct four-dimensional Planck scale,

$$m_4^2 = \alpha^{-1} \left(4\pi\right)^{2/3} L_{11} m_{11}^3 \ , \tag{4.44}$$

The constraint that the inflaton fluctuations remain four-dimensional is:

$$L_{11} < \frac{1}{\gamma H}; \qquad \gamma > \frac{1}{(2\pi)^2} \ , \tag{4.45}$$

while $\alpha \sim 1/25$ constrains the Calabi-Yau volume. Then

$$\frac{H^2}{m_{11}^2} = \frac{(25)^{2/3}(4\pi)^{4/9}}{\gamma^{2/3}} \left(\frac{H}{m_4}\right)^{4/3} \ , \tag{4.46}$$

Assuming also $\gamma = 1/(2\pi)^2$ and $H = 7 \times 10^{13}$ GeV, we find that:

$$m_{11} \sim 6 \times 10^{15} \mathrm{GeV} \ , \qquad 1/L_{11} \sim 10^{12} \ \mathrm{GeV} \ , \tag{4.47}$$

so that

$$\frac{H^2}{M^2} \sim 10^{-4} \ , \qquad \mathcal{X} \sim 0.1 \ . \tag{4.48}$$

This is a considerable improvement, but it is still unobservable. We will find below that if the gauge dynamics are restricted to a lower-dimensional brane, more directions transverse to the brane may be made large, and the fundamental scale can be lowered further still, while keeping the four-dimensional Planck scale fixed.

4.5. G_2 manifolds

M-theory compactified on seven-manifolds with G_2 holonomy also provide $d = 4$, $N = 1$ vacua. Few compact examples are known but one may appeal to heterotic-M theory duality in seven dimensions to make some arguments about their structure [14,15]. For another related construction see [43].

Calabi-Yau threefolds with geometric mirror partners are believed to be T^3 fibrations over an S^3 base [44]. Now heterotic string theory on T^3 is dual to M theory on K_3, so if the base is large and we stay away from the singular fibers, we can claim that the heterotic string on a Calabi-Yau threefold is dual to M-theory on some K_3-fibered manifold with an S^3 base, and hope that the story continues when the singular T^3 fibers are included [14,15]. Indeed, noncompact examples which realize gauge theory with chiral matter take the form of an ALE space (a noncompact K_3) fibered over S^3 or over $S^3/\mathbb{Z}_n$ [15]. We will assume that sensible compact G_2s exist which are K_3 fibrations over $S^3/\mathbb{Z}_n$.

Begin with M theory on a singular K_3 surface with volume V_{K3}. The GUT group in such models arises from the singularities in the K_3 fiber, and so one begins with a seven-dimensional gauge theory with dimensionful gauge coupling $g^2 = \tilde{\ell}_{11}^3$. If we fiber this over $S_3/\mathbb{Z}_p$ with volume V_{S^3} then discrete Wilson lines can break the GUT group to the standard model at the scale $M_{GUT} = V_{S^3}^{-1/3}$.

The four-dimensional GUT coupling is

$$\alpha_{GUT} = \frac{1}{25} = \frac{1}{4\pi V_{S_3} m_{11}^3} \ , \tag{4.49}$$

while the four-dimensional Planck mass is:

$$m_4^2 = \frac{V_{K_3} V_{S_3}}{2\kappa_{11}^2} \ . \tag{4.50}$$

Again we can use these to fix the eleven-dimensional Planck mass and the volume of the K_3 fiber:

$$\begin{aligned} m_{11} &\sim M_{GUT} \\ M_{11} &\sim 6\ M_{GUT} \\ V_{K3}^{-1/4} &\sim 0.1\ M_{GUT} \sim \frac{1}{7} M_{GUT} \ . \end{aligned} \tag{4.51}$$

The eleven-dimensional Planck scale and the compactification scales are within an order of magnitude of each other.

These compactifications look like "brane world" models; the gauge dynamics are localized on singularities with codimension four. One can have several singular regions in the K_3 fibers separate by a length of order $V_{K_3}^{1/4} > 1/M_{GUT}, 1/M_{11}$. The singularities give rise to $7d$ gauge dynamics and the different gauge sectors will be "hidden" from each other, communicating via 11d gravity. Furthermore, the chiral matter also resides on singularities which are points on S^3 [15]. In the $K3$ directions, maximal supersymmetry will be most strongly broken at the singularities on which the gauge dynamics reside. One can imagine an argument similar to that in [38,39] for the existence of inflaton candidates. We leave this for future work.

Corrections to the propagator

For G_2 manifolds, the correction in (4.1) is still given by (4.42). Using the value of m_{11} given by (4.51) the effect is only slightly larger, roughly by a factor of 2. Again, we can ask what happens if we give up grand unification. Here the constraint on m_{11} is simply:

$$m_{11}^6 = \frac{m_4^2}{50\pi V_{K_3}} \ . \tag{4.52}$$

The volume of the $S^3/\mathbb{Z}_n$ base is restricted by $\alpha \sim 1/25$. The constraint

$$V_{K_3} \sim \frac{(2\pi)^8}{H^4} \ , \tag{4.53}$$

and the constraint that the eleven-dimensional energy density be sub-Planckian, lead to the same lower limit on m_{11} to within a factor of 2/3. Using the (tighter) constraint (4.53) , we find, using $H = 7 \times 10^{13}$GeV:

$$\begin{aligned} M \sim m_{11} &\sim 8 \times 10^{13} \text{ GeV} \\ \frac{1}{V_{K3}^{1/4}} &\sim 2 \times 10^{12} \text{ GeV} \\ \frac{H^2}{M^2} &\sim 1 \ , \qquad \mathcal{X} \sim 0.1 \ . \end{aligned} \tag{4.54}$$

M-theory in this limit could have an observable effect on CMBR anisotropies, via the corrections in (4.1). We emphasize again the imprecision of our estimate of $\mathcal{X}$; it is easy to imagine gaining or losing an order of magnitude in an explicit model.

4.6. Type I models

In type I models, supersymmetric coupling unification is consistent with weak string coupling. As discussed in §3, the corrections in (4.1) should be computable via string perturbation theory. These corrections are unobservable as long as string theory is in a computable regime. To see this, we will estimate the maximum size of tree-level and one-loop contributions to (4.1) regardless of unification constraints.

A four-dimensional $N = 1$ supersymmetric model arises in type I string theory from compactification on a six-dimensional Calabi-Yau manifold X. Tree-level corrections to $c_{1,2}$ in (3.1) are the result of compactification. No such terms exist in theories with sixteen supercharges. However, higher-derivative terms such as R^4 terms do exist, suppressed by powers of α'. Upon compactification, such higher-derivative terms will lead to

$$c_{1,2} \sim \frac{m_{CY}^2}{m_s^2} + \mathcal{O}\left(\frac{m_{CY}^4}{m_s^4}\right) \ , \qquad M \sim m_s \tag{4.55}$$

in (3.1), where $m_{CY} \sim V_X^{-1/6}$ is the radius of curvature of X. These terms lead to corrections in (4.1) with $M = m_s$ and $\mathcal{X}$ a function of m_{CY}^2/m_s^2.

The scales and couplings are constrained by:

$$\begin{aligned} \alpha &= \frac{g_s}{4\pi m_s^6 V_6} \\ m_4^2 &= \frac{2 m_s^8 V_6}{g_s^2} \ . \end{aligned} \tag{4.56}$$

Combined, these imply:

$$\begin{aligned} m_s^2 &= \frac{g_s m_4^2}{8\pi\alpha} \\ g_s &= 4\pi\alpha \left(\frac{m_s}{m_{CY}}\right)^6 \end{aligned} \tag{4.57}$$

If the ten-dimensional coupling is weak, then $m_{CY} \gg m_s$ and the α' expansion breaks down [31]. If the α' expansion is good, the ten-dimensional string coupling is strong and (4.57) implies $m_s \geq m_4$. In the scenario which is closest to computable, $m_{CY} \sim m_s \sim m_4$. The correction terms in (4.1) will appear as:

$$M \sim m_s \sim m_4 \ ; \qquad \frac{H^2}{M^2} \sim 3 \times 10^{-8} \ , \tag{4.58}$$

which is unobservable; a large $\mathcal{X}$ in (4.1) would be unnatural. For some type I compactifications with $m_{CY} \sim m_s$, such as orbifolds or Gepner models, or marginal perturbations of them, there is hope of doing a controllable calculation; indeed if $m_{CY} = 2m_s$, $g_s \sim 10^{-2}$. However, unless such models deliver an extremely large value of $\mathcal{X}$, which seems unlikely, the constraints in (4.57) require m_{CY} to be an order of magnitude larger than m_s before $\mathcal{X}H^2/m_s^2$ is observable.

We conclude that for corrections in (4.1) to be observable in a string model, either the $2d$ σ-model coupling or g_s must be large.

4.7. Models with TeV scale gravity

We can take the Hořava-Witten philosophy regarding the four-dimensional Planck scale to a more extreme conclusion. If we assume fewer extra dimensions, with the standard model particles still confined to a $3+1$-dimensional submanifold, we may substantially reduce the fundamental scale of quantum gravity, as low as $m_* = 1$ TeV, while keeping the four-dimensional Planck scale at its known value [36]. In particular, if there are two extra dimensions, the compactification volume could be as large as $(1mm)^2$.

In this class of models the vacuum energy cannot exceed the fundamental scale m_*. Hence after the extra dimensions are stabilized, and the effective 4D Planck scale is given by its low energy value $m_4 \simeq 2 \times 10^{18}$ GeV, if the vacuum energy is localized to the branes the Hubble scale $H = \frac{m_*^2}{3m_{Pl,4}}$ is incredibly small [45], and the mass of the inflaton must be tiny [45],[46], sixteen orders of magnitude below m_*. In addition the effect, which is a correction on the order of $H^2/m_*^2 \sim m_*^2/m_{Pl,4}^2$ to the inflaton fluctuation $\delta\phi$, is unobservable. However, it is inconsistent to search for inflation after such large extra dimensions are stabilized, because if the fundamental scale is low, inflationary dynamics after the stabilization of extra dimensions fails to solve the age problem and does not reproduce the spectrum of primordial fluctuations [46].

These problems are ameliorated if the extra dimensions play an active dynamical role in the early universe. Specifically, if the compactification volume was much smaller at the time of inflation [47], the instantaneous Planck scale at the time of inflation was much smaller than its later value after the stabilization, implying that inflation at times before the extra dimensions are stabilized can address both the age and the fluctuation problems. The details of the pre-stabilization inflationary dynamics are given in [47]. The simplest realization of the scenario proposed in [47] is if the modulus parameterizing the size of the extra dimensions itself is the inflaton. In that case the slow roll condition can be restated as a bound on the ratio of the expansion rate of the dimensions transverse to the brane (extra dimensions) to the expansion rate of the dimensions longitudinal to the brane (macroscopic dimensions). Representing the former by a scale factor b and the latter by a, it is convenient to quantify the slow roll condition $H_b/H_a \ll 1$ (where $H_a = \dot{a}/a$ etc) by the parameters S, T, defined by $H_b/H_a \simeq S + T(b/b_I - 1)^2 + \dots$. Here b_I is the initial size of extra dimensions. Then the slow roll conditions (i.e. the requirement to get sufficient number of efoldings ≥ 70) and the scale-invariance of the spectrum of fluctuations require $T \ll S < 0.002$ [47]. Since the Planck scale at the time of inflation is $m_{4,early}^2 \simeq m_*^{n+2} b_I^n$, the Hubble rate can be expressed as

$$H_a^2 \simeq \frac{V}{3 b_I^n m_*^{n+2}} \tag{4.59}$$

where n is the number of extra dimensions and V the inflationary potential. The COBE normalization of the density contrast at 50 efolds before the end of inflation requires

$$b_I^n m_*^n \simeq \frac{10^3}{S} \frac{\sqrt{V}}{m_*^2} \tag{4.60}$$

and we find after simple algebra

$$\frac{H^2}{m_*^2} \simeq \frac{S}{3000} \frac{\sqrt{V}}{m_*^2} \tag{4.61}$$

independently of the number of extra dimensions. The precise value of V and S is clearly model-dependent; in principle, V which is supported by the branes can be as high as m_*^4 and $S < 0.02$. Thus the maximal value of the imprint of large extra dimensions in the sky is

$$\frac{H^2}{m_*^2} \leq 6.6 \times 10^{-6} \tag{4.62}$$

This is several orders of magnitude too small to be detectable. We should stress that this formula is quite general, because it does not depend on the number of extra dimensions nor the details of the potential, but holds merely as a consequence of a very basic slow roll requirement. The only assumption which this is based on is that the radius modulus is the inflaton. In those cases rapid asymmetric inflation erases any short distance physics imprints on the sky very efficiently. These conclusions might be altered by the construction of more complex scenarios where the inflaton is different from a radius modulus, or where the potential is distributed throughout the bulk. However in the case of TeV gravity models, direct searches for signatures of the new physics in colliders would be much more promising than the surveys of the sky anyway.

5. Modification of Inflationary Consistency

A useful test of inflationary dynamics is the so-called "consistency condition", which relates the ratio of amplitudes of the tensor and scalar modes to the tensor tilt (for a review of potential reconstruction and the consistency condition, see [4]). In standard inflation models (assuming Einstein gravity) the spectrum of scalar fluctuations A_S determines the inflaton potential, and one can then, in principle, use the potential reconstructed from this data to predict the tensor spectrum A_T. In practice, if one expands $\ln(A_S)$ and $\ln(A_T)$ in a power series in the momentum $\ln(k)$, one can only determine the first few coefficients in the series. However, these are enough to provide at least a lowest-order (in the slow-roll parameters) check of consistency. As we will demonstrate, if the effects of high-scale physics are included, the usual relations for inflation in Einstein gravity will not be satisfied.

5.1. Consistency in standard inflation

We define the (unmodified by high scale physics) scalar and tensor spectra (2.20)(2.21)

$$A_{S_0}(k) \equiv \frac{H^2}{10\pi m_4^2 H'} \tag{5.1}$$

and

$$A_{T_0}(k) \equiv \frac{H}{\sqrt{20}\pi m_4}. \tag{5.2}$$

Recall that the k dependence is implicit in H. Then

$$\left(A_{T_0}/A_{S_0}\right)^2 = 2m_4^2 \frac{H'^2}{H^2} \equiv \epsilon_0. \tag{5.3}$$

The tensor tilt is

$$n_{T_0} \equiv \frac{\partial(\ln A_{T_0}^2)}{\partial(\ln k)} = 2\frac{\partial\phi}{\partial(\ln k)}\frac{\partial H}{\partial\phi} = -\frac{4m_4^2 H'^2}{H^2} = -2\epsilon_0, \tag{5.4}$$

to lowest order in ϵ_0. Hence the lowest order prediction of inflationary consistency is:

$$n_{T_0} + 2\left(A_{T_0}/A_{S_0}\right)^2 = 0. \tag{5.5}$$

5.2. High-scale modifications to consistency

When we include the effects of high-scale physics, the observed scalar and tensor spectra will be modified:

$$\begin{aligned} A_S &= A_{S_0}(1 + \mathcal{X}_S H^2/M^2) \\ A_T &= A_{T_0}(1 + \mathcal{X}_T H^2/M^2), \end{aligned} \tag{5.6}$$

where $\mathcal{X}_S$ and $\mathcal{X}_T$ are numerical constants in the effective action, and M is the scale of the new physics. The ratio of the observed tensor and scalar power spectra is:

$$(A_T/A_S)^2 \cong \epsilon_0(1 + 2(\mathcal{X}_T - \mathcal{X}_S)H^2/M^2) \equiv \epsilon. \tag{5.7}$$

However, the tensor tilt n_T is now

$$\partial(\ln A_T^2)/\partial(\ln k) = -2\epsilon_0 \left(1 + 2\mathcal{X}_T H^2/M^2\right), \tag{5.8}$$

and therefore

$$n_T + 2\left(A_T/A_S\right)^2 = -2\epsilon_0 \mathcal{X}_S H^2/M^2 \neq 0. \tag{5.9}$$

Hence we can parameterize the predicted effect of the high-scale physics in a way that is independent of the modification to the graviton kinetic term.

Of course, in standard inflationary models the consistency relation (5.5) is modified at higher order in the slow roll parameters:

$$n_{T_0} + 2\frac{A_{T_0}^2}{A_{S_0}^2} = 2\frac{A_{T_0}^2}{A_{S_0}^2}\left(\frac{A_{T_0}^2}{A_{S_0}^2} - (1 - n_{S_0})\right) = -2\epsilon_0\left(2\eta_0 + 3\epsilon_0\right), \tag{5.10}$$

where η was defined in (2.3) and n_{S_0} is the unmodified scalar tilt. As is manifest, this correction is determined by the slow roll parameters, which in turn can be determined via the measured scalar and tensor power and tilt.

It will therefore be possible to observe the violation of the consistency condition due to high scale physics if the measurements of the scalar and tensor power and tilts are precise enough. This accuracy is limited by cosmic variance, instrumental noise, and backgrounds. Since the tensor fluctuations have not yet been detected, it is not known what the backgrounds will be, and therefore how many independent data points will ultimately be available.

Let us assume optimistically that in a region where the signal is within a factor of three of the maximum we will be able to measure the B-mode of the CMB polarization to an accuracy limited only by cosmic variance. This gives a baseline in the spherical harmonic l from, say, $l = 50$ to $l = 150$ (see *e.g.* [48]). Given that the cosmic variance error in each point is $\sim \sqrt{l} \sim 10$, and we have ~ 100 points, we expect a precision of $\sim \pm 1\%$. We should emphasize again that this estimate is close to a best case scenario. Many other factors could stand in the way of making cosmic variance limited measurements of this quantity, which we must bear in mind, has not even been detected yet.

Assuming one percent precision we can continue our discussion of observability. The violation of the consistency condition is $2\epsilon_0 \mathcal{X}_S H^2/M^2$ (5.9). If we assume $\epsilon_0 \sim 1/15$, and $\mathcal{X}_S H^2/M^2 \sim .1$, the effect will be on the edge of observability. Note that this value of ϵ_0 is about the largest allowed by current experiment. From the standpoint of chaotic inflation models, it requires that the potential near the minimum be controlled by a fairly high order monomial ϕ^n, which requires significant fine tuning. A large discrete symmetry group at the minimum would be required to make this technically natural.

While the corrections in (4.1) lead to violations of (5.10), they are not the only possible source of such an effect. In particular, they are possible in theories with multiple scalar fields in addition to the inflaton [49,50]. Upon examining these models we find that appreciable effects require fine tuning of the scalar potential and the initial conditions for the scalar fields. In general these additional effects are probably unobservable even with our generous estimate of the accuracy of future experiments.

6. Conclusions

Let us summarize the main points we have made. We reviewed the basics of slow roll inflation, emphasizing in particular that the size of CMB fluctuations is determined by inflaton fluctuations at momentum scales $\sim H$. The experimentally important regime is $H \ll m_4$. It is very plausible, then, that one can express the inflaton dynamics at scale H in an effective local field theoretic action where all unknown short distance physics will be encapsulated in the coefficients of irrelevant operators. There are just two leading irrelevant operators (see (3.1)) which produce corrections to $\delta\rho/\rho$ of size $\mathcal{X} H^2/M^2$ where M is the mass scale of the short distance physics and $\mathcal{X}$ is a numerical constant that is calculable if the short distance physics is under calculational control. This is one of the major results of our work.

We then turn to evaluating the size of these corrections in various contexts. For all renormalizable field theories these corrections are generically of size $H^2/m_4^2 \sim 10^{-11}$, too small to observe (although fine tuning can make them much larger). Weakly coupled string theories of conventional type display corrections of similar size. Regions of parameter space that display larger corrections clearly must involve smaller fundamental mass scales. Horava–Witten theory compactified with scales appropriate to grand unification [13] has a fundamental mass scale–the eleven dimensional Planck mass–that is smaller, $m_{11} \sim$

5×10^{16} GeV. Nonetheless for this theory $\mathcal{X} H^2/M^2 \leq 10^{-7}$, still far too small to be observed.

If we allow ourselves to give up precision grand unification we found in section 4. that by studying G_2 compactifications of M theory with large K3 fiber we could lower m_{11} until it was almost comparable to H, giving effects of order one that are potentially observable. We must stress again that these models are not particularly attractive phenomenologically. Precision grand unification must be abandoned, although a reasonably large desert (up to $\sim 10^{13}$ GeV) can be maintained. To maintain the height of the inflationary potential at 10^{16} GeV as m_{11} is lowered we must abandon m_{11} on the branes as the source of V [38] and invoke a more recondite four dimensional mechanism that is invariant under increases in compactification scale. To avoid large proton decay rates we must invoke additional discrete symmetries.

Nonetheless we believe that there are a set of viable M theoretic models that produce potentially observable signals. The comments in Section 5 indicate that we will need cosmic variance limited observations of tensor fluctuations–a very challenging, long term experimental goal–to see such signals. However, there may be other ways to probe these effects in the future. Direct detection of relic gravitational waves (by more sensitive successors to LISA, for example) would probe short wavelengths and so would not be limited by cosmic variance (see *e.g.* [51]). Millisecond pulsar timing measurements would have the same advantage [52]. [9]

Perhaps the most important lesson we have drawn from our work is a qualitative one: the idea of probing short distance physics using cosmological observations looks feasible, possibly even at energies as high as $10^{13} - 10^{14}$ GeV. The challenge now is to open the window wider.

Acknowledgements

It is a pleasure to thank C. Armendariz-Picon, R. Brandenberger, S. Carroll, S. Church, S. Dimopoulos, W. Fischler, M. Green, B. Greene, S. Kachru, D.E. Kaplan, A. Kosowsky, D. Langlois, A. Linde, J. McGreevy, S. Sethi, G. Smoot, A. Strominger, L. Susskind, S. Thomas, E. Witten and J. Yokoyama for valuable discussions. A.L. would like to thank the Aspen Center for Physics and the High Energy Theory Group at the

[9] We thank E. Witten for bringing these methods to our attention.

University of Chicago for their hospitality at various times during the course of this work. This work is supported in part by NSF grant PHY-9870115 and by the Stanford Institute for Theoretical Physics. Albion Lawrence is also supported by the DOE under contract DE-AC03-76SF00515. Matthew Kleban is the Mellam Family Foundation Graduate Fellow.

References

[1] A. Linde, *Particle Physics and Inflationary Cosmology* Harwood Academic, (1990).

[2] E.W. Kolb and M.S. Turner, *The Early Universe* Addison-Wesley (1990).

[3] A. Liddle and D. Lyth, *Cosmological Inflation and Large-Scale Structure* Cambridge University Press (2000).

[4] For a review see J. E. Lidsey *et. al.*, "Reconstructing the Inflaton Potential–An Overview," Reviews of Modern Physics, **69** (1997) 373.

[5] A. Linde, "Axions in Inflationary Cosmology," *Phys. Lett.***B259** (1991) 38; *ibid.*, "Hybrid Inflation," *Phys. Rev.***D49** (1994) 748, astro-ph/9307002.

[6] J. C. Mather *et al.*, "A Preliminary Measurement of the Cosmic Microwave Background Spectrum by the Cosmic Microwave Background Explorer (COBE) Satellite," Astrophys. J. **354** (1990) L37.

[7] D.J.H. Chung, E.W. Kolb, A. Riotto and I. Tkachev, "Probing Planckian Physics: Resonant Production of Particles During Inflation and Features in the Primordial Power Spectrum," Phys.Rev. **D62** (2000) 043508, hep-ph/9910437.

[8] J. Martin and R. H. Brandenberger, "The Transplanckian Problem of Inflationary Cosmology," Phys. Rev. **D63**(2001) 123501, hep-th/0005209; R. H. Branbenberger and J. Martin, "The Robustness of Inflation to Changes in Superplanck Scale Physics," Mod. Phys. Lett. **A16** (2001) 999, astro-ph/0005432; R. H. Brandenberger, S. E. Joras, and J. Martin, Trans-Planckian Physics and the Spectrum of Fluctuations in a Bouncing Universe, hep-th/0112122 .

[9] J. C. Niemeyer, "Inflation With a Planck Scale Frequency Cutoff," Phys. Rev. **D63** (2001) 123502, astro-ph/0005533; J. C. Niemeyer and R. Parentani, "Transplanckian Dispersion and Scale invariance of Inflationary Perturbations," Phys. Rev. **D64** (2001) 101301, astro-ph/0101451.

[10] A. Kempf and J. C. Niemeyer, "Perturbation Spectrum in Inflation With Cutoff," Phys. Rev. **D64** (2001) 103501, astro-ph/0103225.

[11] R. Easther, B. R. Greene, W. H. Kinney, and G. Shiu, "Inflation as a Probe of Short Distance Physics, " Phys. Rev. **D64** (2001) 103502, hep-th/01044102.

[12] R. Easther, B. R. Greene, W. H. Kinney, and G. Shiu, "Imprints of Short Distance Physics on Inflationary Cosmology," hep-th/0110226 .

[13] E. Witten, "Strong Coupling Expansion of Calabi-Yau Compactification," *Nucl. Phys.* **B471** (1996) 135, hep-th/9602070

[14] J. A. Harvey, D. A. Lowe and A. Strominger, "N=1 String Duality," *Phys. Lett.* **B362** (1995) 65, hep-th/9507168; B. S. Acharya, "N=1 Heterotic/M-theory Duality and Joyce Manifolds," *Nucl. Phys.* **B475** (1996) 579, hep-th/9603033.

[15] E. Witten, "Anomaly Cancellation on G_2 Manifolds," hep-th/0108165; B. S. Acharya and E. Witten, "Chiral Fermions from Manifolds of G_2 Holonomy," hep-th/0109152;

E. Witten, "Deconstruction, G_2 Holonomy, and Doublet-Triplet Splitting," hep-ph/0201018.

[16] L. Hui and W. H. Kinney, "Short Distance Physics and the Consistency Relation for Scalar and Tensor Fluctuations in the Inflationary Universe," astro-ph/0109107. For examples of other models that discuss modifications of the consistency conditions, see *e.g.*, J. Garriga and V. F. Mukhanov, "Perturbations in K-Inflation," Phys. Lett. **B458** (1999) 219, hep-th/9904176; and G. Shiu and S. H. H. Tye, "Some Aspects of Brane Inflation," Phys. Lett. **B516** (2001) 421, hep-th/0106274.

[17] K. Freese, J.A. Frieman and A.V. Olinto, "Natural Inflation with Pseudo-Nambu-Goldstone Bosons", Phys. Rev. Lett. **65** (1990) 3233; F.C. Adams, J.R. Bond, K. Freese, J.A. Frieman and A.V. Olinto, "Natural Inflation: Particle Physics Models, Power Law Spectra for Large Scale Structure, and Constraints from COBE", Phys. Rev. **D47** (1993) 426, hep-ph/9207245

[18] P. Binetruy and M.K. Gaillard, "Candidates for the Inflaton Field in Superstring Models," Phys. Rev. D34 (1986) 3069; L. Randall and S. Thomas , "Solving the Cosmological Moduli Problem with Weak Scale Inflation," Nucl. Phys. **B449** (1995) 229, hep-ph/9407248; T. Banks, M. Berkooz, G. Moore, S. Shenker and P. Steinhardt, "Modular cosmology," *Phys. Rev.* **D52** (1995) 3548, hep-th/9503114; S. Thomas, "Moduli Inflation from Dynamical Supersymmetry Breaking," Phys.Lett. **B351** (1995) 424, hep-th/9503113.

[19] W. Fischler and L. Susskind, "Holography and Cosmology," hep-th/9806039.

[20] R. Bousso, "A Covariant Entropy Conjecture," JHEP **9907** (1999) 004, hep-th/9905177; *ibid.*, "Holography in general space-times," JHEP **9906** (1999) 028, hep-th/9906022.

[21] S. Hellerman, N. Kaloper and L. Susskind, "String theory and quintessence," JHEP **0106**, (2001) 003, hep-th/0104180; W. Fischler, A. Kashani-Poor, R. McNees and S. Paban, "The Acceleration of the Universe, a Challenge for String Theory", JHEP **0107** (2001) 003, hep-th/0104181.

[22] A. E. Lange *et al.*, "Cosmological Parameters from the First Results of BOOMERANG," Phys. Rev. **D63** (2001) 042001, astro-ph/0005004.

[23] Balbi, A. *et al.*, "Constraints on Cosmological Parameters from MAXIMA-1", Astrophys. J. **545** (2000) L5, astro-ph/0005124.

[24] J. Bardeen, P. Steinhardt, and M. Turner, "Spontaneous Creation of Almost Scale-Free Density Perturbations in an Inflationary Universe," Phys. Rev **D28** (1983) 679.

[25] L. Wang, V. Mukhanov, and P. Steinhardt, "On the Problem of Predicting Inflationary Perturbations," Phys. Lett. **B414** (1997) 18, astro-ph/9709032.

[26] E. Copeland, I. Girvell, E. Kolb, and A. Liddle, "On the Reliability of Inflaton Potential Reconstruction," Phys. Rev. **D58** (1998) 043002, astro-ph/9802209.

[27] For a pedagogical review see W. Hu and M. J. White, "A CMB Polarization Primer," *New Astron.* **2** (1997) 323, astro-ph/9706147.

[28] M. Kamionkowski and A. Kosowsky, "The Cosmic Microwave Background and Particle Physics," *Ann. Rev. Nucl. Part. Sci.* **49** (1999) 77, astro-ph/9904108.

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

[30] T. Banks, "Cosmological Breaking of Supersymmetry? or Little Lambda Goes Back to the Future 2. ," hep-th/0007146 .

[31] M. Dine and N. Seiberg, "Couplings and Scales in Superstring Models," *Phys. Rev. Lett.* **55** (1985) 366.

[32] V. Kaplunovsky, "Mass Scales of the String Unification," *Phys. Rev. Lett.* **55** (1985) 1036.

[33] P. Hořava and E. Witten, "Heterotic and Type I string dynamics form eleven dimensions", *Nucl. Phys.* **B460** (1996) 506, hep-th/9510209; *ibid.*, "Eleven-dimensional supergravity on a manifold with boundary", *Nucl. Phys.* **B475** (1996) 94, hep-th/9603142.

[34] S. Dimopoulos and H. Georgi, "Softly Broken Supersymmetry and SU(5)", Nucl. Phys. **B192** (1981) 150; S. Dimopoulos, S. Raby and F. Wilczek, "Supersymmetry and the Scale of Unification", Phys. Rev. **D24** (1981) 1681; U. Amaldi, W. de Boer and H. Furstenau, "Comparison of Grand Unified Theories with Electroweak and Strong Coupling Constants Measured at LEP", Phys. Lett. **B260** (1991) 447; P. Langacker and N. Polonsky, "Uncertainties in Coupling Constant Unification", Phys. Rev. **D47** (1993) 4028, hep-ph/9210235.

[35] J. Polchinski and E. Witten, "Evidence for Heterotic - Type I String Duality," *Nucl. Phys.* **B460** (1996) 525, hep-th/9510169.

[36] N. Arkani-Hamed, S. Dimopoulos and G.R. Dvali, "The Hierarchy Problem and New Dimensions at a Millimeter," *Phys. Lett.* **B429** (1998) 263, hep-ph/9803315; *ibid.*, "Phenomenology, Astrophysics and Cosmology of Theories with Sub-millimeter Dimensions and TeV Scale Quantum Gravity," *Phys. Rev.* **D59** (1999) 086004, hep-ph/9807344.

[37] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. Dvali, "New Dimensions at a Millimeter to a Fermi and Superstrings at a TeV," *Phys. Lett.* **B436** (1998) 257; hep-ph/9804398.

[38] T. Banks, "Remarks on M Theoretic Cosmology," hep-th/9906126.

[39] T. Banks, "M-theory and cosmology," hep-th/9911067.

[40] M.B. Green, J. Schwarz and E. Witten, *Superstring Theory*, vols. I and II, Cambridge Univ. Press (1987).

[41] P. Hořava, "Gluino Condensation in Strongly Coupled Heterotic String Theory," *Phys. Rev.* **D54** (1996) 7561, hep-th/9608019.

[42] M. Green and P. Vanhove, "D Instantons, Strings, and M-Theory," Phys. Lett. **B 408** (1997) 122, hep-th/9704145; M. B. Green, M. Gutperle, and P. Vanhove, "One Loop in Eleven–Dimensions," Phys. Lett. **B409** (1997) 177, hep-th/9706175; and J. Russo and A. Tseytlin, "One-loop Four-graviton Amplitude in Eleven-Dimensional Supergravity," Nuc. Phys. **B508** (1997) 245, hep-th/9707134.

[43] M. Cvetic, G. Shiu, and A. Uranga, "Three Family Supersymmetric Standard–Like Models from Intersecting Brane Worlds," Phys. Rev. Lett. **87**, (2001) 201801, hep-th/0107143; and M. Cvetic, G. Shiu, and A. Uranga, "Chiral Four-Dimensional N = 1 Supersymmetric Type IIA Orientifolds from Intersecting D6-Branes," Nucl. Phys. B **615**,(2001) 3, hep-th/0107166.

[44] A. Strominger, S.-T. Yau and E. Zaslow, "Mirror Symmetry is T-duality," *Nucl. Phys.* **B479** (1996) 243, hep-th/9606040.

[45] K. Benakli and S. Davidson, "Baryogenesis in Models with a Low Quantum Gravity Scale," *Phys. Rev.* **D60** (1999) 025004, hep-ph/9810280; D. Lyth, "Inflation With TeV-scale Gravity Needs Supersymmetry," *Phys. Lett.* **B448** (1999) 191, hep-ph/9810320.

[46] N. Kaloper and A. Linde, "Inflation and Large Internal Dimensions," *Phys. Rev.* **D59** (1999) 101303, hep-th/9811141.

[47] N. Arkani-Hamed, S. Dimopoulos, N. Kaloper and J. March-Russell, "Rapid Asymmetric Inflation and Early Cosmology in Theories with Sub-millimeter Dimensions," *Nucl. Phys.* **B567** (2000) 189, hep-ph/9903224.

[48] U. Seljak and M. Zaldarriaga, "Signature of Gravity Waves in Polarization of the Microwave Background," *Phys.Rev.Lett.* **78** (1997) 2054.

[49] D. Polarski and A.A. Starobinsky, "Structure of Primordial Gravitational Waves Spectrum in a Double Inflationary Model," *Phys. Lett.* **B356** (1995) 196, astro-ph/9505125; J. Garcia-Bellido and D. Wands, *Phys. Lett.* **D52** (1995) 6739; M. Sasaki and E. Stewart, "A General Analytic Formula for the Spectral Index of the Density Perturbations Produced During Inflation," *Prog. Theor. Phys.* **95** (1996) 71, astro-ph/9507001.

[50] N. Bartolo, S. Matarrese and A. Riotto, "Adiabatic and Isocurvature Perturbations from Inflation: Power Spectra and Consistency Relations," *Phys. Rev.* **D64** (2001) 123504.

[51] M. Turner, "Detectability of Inflation Produced Gravitational Waves," *Phys. Rev.* **D55** (1997) 435, astro-ph/9607066

[52] D.R. Stinebring, M.F. Ryba, J. H. Taylor, and R.W. Romani, "The Cosmic Gravitational Wave Background: Limits from Millisecond Pulsar Timing," Phys. Rev. Lett. 65 (1990) 285.