

# Weak gravity strongly constrains large-field axion inflation

**Ben Heidenreich, Matthew Reece and Tom Rudelius**

*Department of Physics, Harvard University,  
Cambridge, MA, 02138 U.S.A.*

*E-mail:* [bjheiden@physics.harvard.edu](mailto:bjheiden@physics.harvard.edu), [mreece@physics.harvard.edu](mailto:mreece@physics.harvard.edu),  
[rudelius@physics.harvard.edu](mailto:rudelius@physics.harvard.edu)

**ABSTRACT:** Models of large-field inflation based on axion-like fields with shift symmetries can be simple and natural, and make a promising prediction of detectable primordial gravitational waves. The Weak Gravity Conjecture is known to constrain the simplest case in which a single compact axion descends from a gauge field in an extra dimension. We argue that the Weak Gravity Conjecture also constrains a variety of theories of multiple compact axions including N-flation and some alignment models. We show that other alignment models entail surprising consequences for how the mass spectrum of the theory varies across the axion moduli space, and hence can be excluded if further conjectures hold. In every case that we consider, plausible assumptions lead to field ranges that cannot be parametrically larger than  $M_{\text{Pl}}$ . Our results are strongly suggestive of a general inconsistency in models of large-field inflation based on compact axions, and possibly of a more general principle forbidding super-Planckian field ranges.

**KEYWORDS:** Compactification and String Models, Cosmology of Theories beyond the SM, Models of Quantum Gravity

ARXIV EPRINT: [1506.03447](https://arxiv.org/abs/1506.03447)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Preliminaries</b>	<b>6</b>
2.1	Axions from extra dimensions	6
2.2	Weak Gravity Conjecture	7
2.3	The list of ingredients	8
2.4	WGC and single-axion inflation: electric argument	9
2.5	WGC and single-axion inflation: magnetic argument	10
<b>3</b>	<b>Weak Gravity Conjecture and multi-axion models</b>	<b>10</b>
3.1	Warmup: diagonal N-flation	11
3.2	Magnetic WGC and kinetic alignment	12
<b>4</b>	<b>New conjectures on EFT over the moduli space</b>	<b>15</b>
4.1	Exploring the moduli space: masses and Kaluza-Klein reduction	15
4.2	Consistency of a single EFT across axion moduli space	16
4.3	Consequences of Single-EFT Consistency for monodromy	18
4.4	The Weak Gravity Conjecture across moduli space	19
4.5	The Lattice Weak Gravity Conjecture	22
<b>5</b>	<b>Decay constant alignment</b>	<b>23</b>
5.1	Parametrizing the fundamental domain	25
5.2	Case 1: dominant instantons satisfy electric WGC	25
5.3	Case 2: additional particles satisfy electric WGC	26
5.4	Status of decay constant alignment	27
<b>6</b>	<b>Assessing strong conjectures</b>	<b>27</b>
<b>7</b>	<b>Conclusions</b>	<b>29</b>
<b>A</b>	<b>The SECC and gauge-invariant regulators</b>	<b>31</b>
<b>B</b>	<b>Do entropy bounds exclude large-field inflation?</b>	<b>33</b>
B.1	The reheating objection	34
B.2	Bit-counting arguments	34
B.3	The classical entropy current argument	35

---

## 1 Introduction

If a significant primordial gravitational wave signal is detected in any near-future experiment, it will imply that the inflaton traversed a distance in field space larger than  $M_{\text{Pl}}$ . This is the famous Lyth bound [1] (for further refinements, see [2–5]). On first contemplating super-Planckian field ranges, an effective field theorist will tend to feel some discomfort, being inclined to write down the most general effective Lagrangian,

$$\mathcal{L} = -\frac{1}{2}m^2\phi^2 + \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{\lambda}{4!}\phi^4 + \frac{c}{M_{\text{Pl}}^2}\phi^2\partial_\mu\phi\partial^\mu\phi - \frac{\lambda_6}{M_{\text{Pl}}^2}\phi^6 - \dots \quad (1.1)$$

For order-one values of the coefficients of high-dimension operators in  $V(\phi)$ , the potential will oscillate wildly over super-Planckian field ranges and spoil inflation. However, on further reflection, one realizes that a shift symmetry  $\phi \rightarrow \phi + a$  would forbid every term in the potential, and thus a theory with a single dominant source of shift symmetry breaking can provide a technically natural approach to super-Planckian field ranges: we would predict that for some  $f$ ,  $\lambda \sim m^2/f^2$ ,  $\lambda_6/M_{\text{Pl}}^2 \sim m^2/f^4$ , and so on. We could accommodate this shift symmetry in our effective field theory by parametrizing its breaking with a spurion and building a Lagrangian out of fields that nonlinearly realize the symmetry. The good behavior of the potential is enforced by the dominance of a single spurion. In fact, many physicists readily accept the axion as a potential solution to the strong CP problem, in spite of its severe Planck-suppressed operator problem [6–8]. To explain the very tight experimental bound on the effective theta angle in our universe, a theory of a generic pseudo-Goldstone axion must forbid high-dimension Planck-suppressed Peccei-Quinn-violating operators, so that QCD instanton effects provide the leading shift-symmetry-breaking spurion by several orders of magnitude. If the axion is a compact field (or equivalently, has an exact gauged discrete shift symmetry,  $a \rightarrow a + 2\pi f$ ), then only exponentially small instanton effects can break the symmetry and it may be relatively easy to enforce single-spurion dominance. This may be realized by string theory axions, for example [9, 10]. The similarities in the need for a good shift symmetry to solve the strong CP problem and for large-field inflation have motivated concrete models of inflation which are natural from the effective field theory viewpoint [11–28]. (Among these, the idea of  $N$ -flation builds on earlier work on assisted inflation, which was first studied in the context of exponential rather than axion-like potentials and then was generalized [29–32]. Another variation is  $M$ -flation [33–35].)

While effective field theorists can find no dramatic problem with large-field inflation, they may retain some skepticism about the existence of UV completions. It has been suggested that quantum gravity will impose more severe constraints than EFT. Axion fields with super-Planckian decay constants appear to be rare in string theory vacua [36, 37], which may be a suggestive hint of a more general principle. It is generally believed that the set of quantum gravity theories is discrete (at least for nonsupersymmetric theories, where subtleties of continuous moduli spaces do not arise). This means that many apparently sensible effective field theories are actually in the swampland of theories that cannot be consistently coupled to gravity [38–41]. At this point we have relatively few guidelines for

how to judge that a theory is in the swampland, but the Weak Gravity Conjecture (WGC) is among the sharpest and most powerful and well-motivated [39, 42–45].

The WGC asserts that any theory containing both gravity and a massless abelian gauge field should have a charged particle in the spectrum whose mass is less than its charge in Planck units. To be precise, we require  $m < \sqrt{2}qeM_{\text{Pl}}$  in a four-dimensional theory in which gravitons and photons are the only massless particles. The motivation is to avoid having a plethora of exactly stable extremal black hole states, which are potentially problematic [46]. This leads to:

*The Weak Gravity Conjecture (WGC):* for any large, semiclassical, nearly-extremal black hole, there exists a state in the theory whose mass is small enough relative to its charge that the black hole can move away from extremality by emitting this state.

For a single U(1) gauge group, this implies that there is a state satisfying  $q/m \geq z_0$  for an appropriate constant  $z_0$ . In the case of multiple U(1)s, the condition is that the convex hull of the charge-to-mass vectors  $\vec{z} = \vec{q}/m$  of the kinematically available charged states must contain the ball of radius  $z_0$  [44]. This is purely a kinematic requirement: black holes should be able to decay.

From the very beginning, the WGC was claimed to rule out the theory of extranatural inflation [13]. Recently, there has been renewed interest in how the WGC can constrain large-field inflation [37, 47–55]. The essential idea is that theories of axions with good shift symmetries often obtain four-dimensional axion fields by dimensional reduction of higher-rank  $p$ -form fields, which are constrained by WGC arguments.

However, such arguments are not without subtlety. The WGC is a rather weak statement at low energies, since large black holes could decay to states out of the reach of low-energy effective field theory. Given an effective field theory with cutoff  $\Lambda$ , the lightest semiclassical black holes have mass of order  $M_{\text{Pl}}^2/\Lambda$ , so one could imagine that the conjecture is satisfied by states with mass between  $\Lambda$  and  $M_{\text{Pl}}^2/\Lambda$  that cannot be studied without a full theory of quantum gravity. Moreover, extremal black holes can also satisfy the WGC, provided that the subleading corrections to the extremality bound have the correct sign [39, 42]. Nonetheless, the magnetic form of the WGC does have important consequences for the low-energy effective field theory and for inflation, which we review in section 2.2.

As effective field theorists, we might want to impose the stronger constraint that black hole decay can be described in the low energy effective field theory. That is, we might want to limit our attention to theories in which we can positively assert that black holes decay, reasoning that any theory which violates this assumption lies outside theoretical control in the absence of a full, quantum gravity description. This suggests a variant of the WGC:

*The Effective Weak Gravity Conjecture (EWGC):* the state which satisfies the weak gravity conjecture should be describable in the low-energy effective field theory.

Indeed, we usually imagine black holes decaying to particles, hence the EWGC is sometimes implicit in discussions of the WGC. However, we emphasize that it amounts to

a further assumption, which however is much weaker than the “strong form” of the WGC proposed in [39], which we discuss in section 6, but which plays no role in our arguments. By contrast, the EWGC plays an important role in some — but not all — of our arguments.

(We note in passing that the EWGC is implied by the much-stronger “lattice WGC” [56] — discussed briefly in section 4.5 — whenever there is *any* charged particle that can be described in the low-energy effective field theory.)

Unlike the WGC, the EWGC is not directly motivated by the problem of remnants, since Planck-scale states satisfying the WGC can address this issue. If correct, the EWGC may stem from a dynamical version of the (kinematic) WGC. For instance, one variation on the Third Law of Thermodynamics in the black hole context could be that nearly-extremal black holes should not spontaneously move closer to extremality. To realize the implications of this statement, we note that black holes with sufficient charge do not decay predominantly by Hawking radiation. Initially, since Hawking radiation carries very little charge away, they decay *towards* extremality [57]. The temperature of a Reissner-Nordström black hole goes to zero as the black hole approaches extremality, shutting off the Hawking radiation, but charged particles satisfying the WGC can still be emitted through an effect that is similar to Schwinger pair-production near the horizon [58–60], which eventually becomes the dominant decay channel.

The effect may also be understood as a Breitenlohner-Freedman-type instability of charged particles in the near-horizon geometry [61–63]. All that is required is that at least one such particle exist; despite statements to the contrary in [43], it need not be the lightest charged particle, as lighter particles violating the WGC inequality are not emitted. Conversely, if there are no charged particles in the low-energy effective field theory satisfying WGC, we expect that the black hole continues to approach extremality, as pair production cannot occur whereas the emission of long string states or fission of the black hole into smaller black holes should be a very slow process in comparison to Hawking radiation.<sup>1</sup> Thus, the EWGC may be motivated by thermodynamic considerations.

As this paper was being finalized, we learned of work on the possibility that the WGC is satisfied by states that are not captured in the low-energy effective field theory [64]. Even if the EWGC does not hold in every consistent theory, the theories in which it is violated have the unusual property that the decay of large, semiclassical black holes cannot be described semi-classically. This suggests that the naive “low energy effective field theory” is not a completely reliable description of the full theory at low energies, and in particular any conclusions that we draw about inflation based solely on this class of theories may not be reliable.

For the same reasons of theoretical control, in this paper we focus on the case where the low-energy effective field theory is a weakly-coupled abelian gauge theory, containing electrically charged particles light enough to discharge subextremal electrically charged

---

<sup>1</sup>In principle, when the black hole is very close to extremality, the temperature of the Hawking radiation may become low enough that these strongly suppressed processes are competitive. However, we expect that for a large black hole with a weakly curved horizon, this transition occurs *exponentially* close to extremality, hence it may not even be visible in the thermodynamic limit. Moreover, it’s not clear that these processes are effective at discharging the black hole, since their rates are typically not under theoretical control.

black holes and semiclassical (solitonic) monopoles light enough to discharge subextremal magnetically charged black holes. This assumption could be circumvented in examples where a more sophisticated field theory description of the charged particles is available, e.g. in cases where the abelian gauge theory arises from Higgsing a non-Abelian gauge group and the monopoles originate from “hedgehog” configurations in the parent theory. However, we see no reason to expect that such theories will evade our constraints, hence for simplicity we defer consideration of them to future work.

Our goal in this paper is to give a critical assessment of the state of large-field inflation in light of the WGC. We focus on scenarios with compact axion fields, leaving noncompact models of axion monodromy for future consideration (though some of our remarks may extend to such models). We find that arguments against approximately isotropic models of  $N$ -flation [15] and kinetic alignment [14, 23] are robust. The most difficult scenario to rule out arises from loopholes pointed out by de la Fuente, Saraswat, and Sundrum [48, 65]. We present conjectured bounds on this scenario that depend on the way that the mass spectrum shifts as the axion VEVs are varied. In every case that we find a bound, the parametrics *precisely* compensates for any possible enhancement and determines that the field range is bounded above by  $M_{\text{Pl}}$  (times at most an order-one number).

The organization of this paper is as follows. In section 2, we review extranatural inflation [13], the Weak Gravity Conjecture and its requirement of a low UV cutoff [39], and arguments against large-field inflation with a single axion from both the electric [39] and magnetic [48] points of view. In section 3, we present arguments for how the magnetic form of the Weak Gravity Conjecture excludes the simplest models of  $N$ -flation [15] and models based on the Kim-Nilles-Peloso alignment mechanism [14] when it is realized through the structure of kinetic mixing [23]. These arguments do not address the case of decay constant alignment where the leading instanton effects arise from highly aligned electric charges in a basis in which the magnetic charges are not aligned [48]. In section 4, we discuss how the spectrum of modes for extranatural inflation varies while traversing the axion moduli space. We argue that the models that evade our earlier arguments involve spectra with surprising features that one must accept in order to realize large-field inflation. We claim that, in the presence of particles of large charge, the 5d effective field theory breaks down in some regions of the moduli space if the compactification radius is not significantly smaller than the UV cutoff. This motivates the Single-EFT Consistency Criterion (SECC), which demands that any Kaluza-Klein mode which is light in some region of moduli space should have mass below the UV cutoff throughout the moduli space. Equivalently, it requires that the axion shift symmetry  $\theta \rightarrow \theta + 2\pi$  can be understood as a large gauge transformation within the domain of validity of the UV completion. We emphasize that the SECC can be motivated within the 5d effective field theory in the presence of a Wilson line, not just the dimensionally reduced viewpoint. (In appendix A we further illustrate the SECC using a gauge-invariant lattice regulator.) A second new conjecture, the Extended Weak Gravity Conjecture, requires that the spectrum of particles should satisfy the WGC bound at every stationary point in moduli space (including local maxima). In section 5, we show that the loopholes raised by [48] violate our stronger conjectures. This completes our discussion of bounds on inflation from the WGC.

Some prior claims exist in the literature of strong constraints on large-field inflation models. Section 6 addresses such claims based on hypothetical strong forms of the WGC [51]. Appendix B considers claims based on entropy bounds [66–68]. These sections exist to place our paper in context and to convince the reader that we are not simply providing weaker proofs of previously known facts; they can be freely skipped. We claim that the entropy bound arguments in the literature rest on overly strong assumptions, and that the “strong form” of the WGC is ambiguous when applied to multiple U(1)s. We emphasize that we have never used such a strong form in this paper, and that we view the issues of Kaluza-Klein mode monodromy to be the most important new requirements we have used.

We conclude in section 7 with a discussion of our view of what remains to be done to place our conjectured requirements on a sounder footing. We believe that further progress along these lines can either thoroughly exclude parametrically large field inflation or identify special theories that satisfy all consistency requirements of quantum gravity.

## 2 Preliminaries

In this section we review the concept of extranatural inflation, the Weak Gravity Conjecture and its implications for the UV cutoff of a theory, and how the WGC rules out models of inflation driven by a single axion field. Along the way we make a few small remarks not present in the existing literature, but readers thoroughly familiar with the WGC can skip to the next section for our new results.

### 2.1 Axions from extra dimensions

In this paper we will focus on extranatural inflation models [13] in which the axion field arises by reducing a  $p$ -form gauge field ( $p \geq 1$ ) on a  $p$ -dimensional cycle within a compactification manifold. String theory axions [9, 10] share this feature with simple phenomenological models. It seems plausible that any consistent theory of a compact axion field coupled to quantum gravity can be viewed, in some duality frame, as a member of this class, so we view the restriction to extranatural models as a mild assumption.

For example, consider the case of an ordinary (1-form) gauge symmetry. If the compactification manifold contains a circle, we can define an axion-like field in four dimensions from the integral around this circle:

$$\theta(x) = \oint_0^R dx^5 A_5(x, x^5). \quad (2.1)$$

Large gauge transformations of  $A_5$  lead to  $\theta \rightarrow \theta + 2\pi n$  (where  $n \in \mathbb{Z}$ ) representing a (gauged) discrete shift symmetry and signifying that the axion field is compact. The same statement holds for axions obtained from higher-rank  $p$ -forms, where there is always the freedom to perform a large gauge transformation shifting the field by an integral multiple of the volume form of the  $p$ -cycle we integrate over.

Beginning from a 5d action  $\int d^5x (\frac{1}{2}M_5^3 \mathcal{R} - \frac{1}{4e_5^2} F_{\mu\nu}^2)$ , we obtain a 4d Planck scale set by  $M_{\text{Pl}}^2 \equiv M_4^2 = 2\pi R M_5^3$ , gauge coupling  $1/e^2 = 2\pi R/e_5^2$ , and an axion decay constant set by

$$\mathcal{L}_{\text{axion}} = \int d^4x \frac{1}{2} f^2 \partial_\mu \theta \partial^\mu \theta, \quad \text{where } f^2 = \frac{1}{2\pi R e_5^2} = \left( \frac{1}{2\pi R e} \right)^2. \quad (2.2)$$

In much of the paper, we will focus on this case, where a four-dimensional axion arises from the dimensional reduction of a one-form gauge field in five dimensions. Nonetheless, most of our arguments generalize straightforwardly to axions originating from  $p$ -form gauge fields in  $D$  dimensions.

## 2.2 Weak Gravity Conjecture

A useful formulation of the Weak Gravity Conjecture is that any near-extremal charged black hole should be able to move away from extremality by emitting a charged particle. For a  $D$ -dimensional theory containing a  $p$ -form field with coupling  $g_p$  also coupled to a massless canonically normalized dilaton  $\varphi$  through its kinetic term  $\propto e^{-\alpha\varphi\sqrt{16\pi G}} F_p^2$ , the WGC asserts that there should exist a  $(p-1)$ -brane with charge  $q$  under the  $p$ -form and tension  $T$  satisfying the inequality

$$8\pi G \left[ \frac{\alpha^2}{2} + \frac{p(D-p-2)}{D-2} \right] T^2 \leq g_p^2 q^2. \quad (2.3)$$

We comment on the detailed  $\alpha$ ,  $p$ , and  $D$  dependence of this expression in a separate paper [56].

For a U(1) gauge theory in four dimensions with coupling constant  $e$ , this has the immediate implication that there should exist an electrically charged particle of mass  $m$  and charge  $q$  satisfying  $m < \sqrt{2}eqM_{\text{Pl}}$  and a magnetic monopole of mass  $m_{\text{mon}}$  and charge  $q_m$  satisfying  $m_{\text{mon}} < \sqrt{2}q_m M_{\text{Pl}}/e$ . As pointed out in [39], the existence of a monopole satisfying such a bound implies that weakly-coupled gauge theories coupled to gravity must have a UV cutoff that is parametrically below the Planck scale. The self-energy of the monopole from integrating its classical magnetic field down to a distance  $r_{\text{min}} \sim \Lambda^{-1}$  is linearly divergent:  $\delta m_{\text{mon}} \sim \frac{q_m}{e^2} \Lambda$ . The precise coefficient is not essential because we make a naturalness argument, that  $m_{\text{mon}} \gtrsim \delta m_{\text{mon}}$  (in the absence of fine-tuning), which like all naturalness arguments leaves the precise order-one coefficient in the bound as a matter of taste. Combining the naturalness bound with the magnetic WGC, we learn that

$$\Lambda \lesssim \frac{e}{q_m} M_{\text{Pl}}. \quad (2.4)$$

In other words, the UV cutoff of the theory should be below  $eM_{\text{Pl}}$ , and if the magnetic WGC is satisfied by a monopole of large charge the UV cutoff will be even lower. We emphasize that this magnetic form of the WGC, from [39], does not depend on the EWGC, and our arguments based on it are the most robust.

Notice that although one could attempt to apply the self-energy argument also to the electric field of electrically charged particles, such an integral must be cut off at the Compton wavelength  $m^{-1}$  of the field. As is familiar from QED, the self-energy of the electron



is a small correction to its mass. The argument that monopoles can bound the cutoff  $\Lambda$  relies on the fact that the magnetic coupling is strong, the monopole is a solitonic object with mass above the cutoff, and that electric and magnetic charges are mutually nonlocal, so a theory containing both necessarily has a fundamental cutoff. The interpretation of the quantity  $\Lambda$  that we are bounding is the scale at which it is no longer appropriate to treat the theory as a weakly-coupled, local effective field theory of a U(1) gauge boson. This may be because U(1) embeds into a nonabelian group which has  $W$  bosons at the scale  $\Lambda$ , or may be due to more fundamental new physics such as the string scale or the higher-dimensional Planck scale.

### 2.3 The list of ingredients

In this paper there are four ingredients that will play a role in our arguments:

1. *The instanton effects.* We are dealing with a set of axions that have a potential of the form  $\sum_i c_i \cos(\sum_j Q_{ij}\theta_j)$ . The charge matrix  $Q$  is typically determined by a set of wrapped worldvolumes of electrically charged particles in extranatural inflation theories, but could originate from other physics.
2. *The electric charges satisfying the WGC.* A set of electrically charged particles exist that satisfy the constraints demanded by Weak Gravity. That these may not be the *same* charged particles contributing the dominant instanton effects is one source of loopholes to the simplest arguments.
3. *The magnetic charges satisfying the WGC.* A set of magnetically charged particles exist that satisfy the constraints of Weak Gravity. These particles give us crucial access to information about the UV cutoff of the theory and are the central elements of many of our arguments.
4. *The kinetic mixing matrix.* Our axions (or the higher-dimensional gauge fields they originated as) in general mix with each other through a matrix  $K$ .

A number of models in the literature rely on “alignment,” following the work of Kim, Nilles, and Peloso [14]. A general alignment model is one in which all four of our ingredients can freely vary. In this paper we address two *physically distinct* special cases of alignment, which are not simply the same idea in a different basis. We will use the term *kinetic alignment* to refer to a scenario in which the electric charges contributing the dominant instanton effects (our first ingredient) and the magnetic charges satisfying the WGC (our third ingredient) are *both* simple (small integers in the charge lattice) in a basis where  $K$  is arbitrary. This addresses models like [23], though we make the additional assumption of simple magnetic charges. On the other hand, we could consider a model in which the kinetic term is simple but the charge matrix of the instanton effects and the magnetic charges satisfying the WGC are not (so that, for instance, there are large numbers appearing in the instanton matrix while the magnetic charges are simple). This is the scenario of [48], which we refer to as *decay constant alignment*. Given this scenario, we could always redefine our basis to make the electric charges simple while scrambling the kinetic matrix. However, we

would then have unusual charge assignments for the magnetic monopoles, and we require a different physical argument than the one we applied to the scenario we referred to as kinetic alignment. Although we only obtain bounds on these two special cases, we expect that a combination of the ideas used in these bounds can exclude the general case.

To fix conventions for the electric charge in the extranatural case, we write the coupling of a charged particle to the gauge potential as

$$S = Q_a \int_P A^a, \tag{2.5}$$

where the integral is over the worldline of the particle, and  $a$  indexes the different gauge fields in the case of multiple  $U(1)$ 's. This leads to an axion potential of the form

$$V = V_0 \sum_n c_n \cos(nQ_a \theta^a), \tag{2.6}$$

in the dimensionally reduced theory, where  $\theta^a = \oint A^a$ . Likewise, we define the magnetic charge enclosed in a spatial region  $\Sigma$  as

$$\tilde{Q}^a \equiv \frac{1}{2\pi} \int_{\partial\Sigma} F^a, \tag{2.7}$$

where  $F^a = dA^a$ . The Dirac quantization condition is

$$\tilde{Q}^a Q_a \in \mathbb{Z}. \tag{2.8}$$

Unless otherwise specified, we always work in a basis where  $\tilde{Q}^a$  and  $Q_a$  are integrally quantized.

#### 2.4 WGC and single-axion inflation: electric argument

An argument that the WGC excludes extranatural inflation with super-Planckian decay constants was given already in the original paper [39]. The electric WGC in 5 dimensions tells us that a charged particle exists with mass  $m < \sqrt{3/2} e_5 q M_5^{3/2}$ . Upon reducing to four dimensions, this implies that the charged particle contributes an instanton action

$$S_{\text{inst}} = 2\pi R m < \sqrt{\frac{3}{2}} \frac{q M_{\text{Pl}}}{f}. \tag{2.9}$$

Let us assume that the same instanton generates the axion potential. If we want to focus on large instanton actions so that higher-order corrections to the potential are exponentially suppressed, this shows that we require  $f/q \lesssim M_{\text{Pl}}$ , where  $f/q$  is the field range determined by this potential.

As recently emphasized in [48, 65], this argument is not completely convincing because there is a “small action loophole”: the prefactor in front of higher-instanton terms can suppress them even if the instanton action  $S \ll 1$ . Direct calculation confirms that the potential is a sum of terms proportional to  $\exp(-2\pi n m R) \cos(na/f)$  with coefficients decreasing with powers of  $n$  [13, 69, 70]. Even at  $m = 0$  where the instanton action gives no suppression, the  $n^{\text{th}}$  term in the sum has a  $1/n^5$  suppression which is enough to safely

allow inflation. A useful way to understand this power law suppression is to write the contribution of winding number  $n$  in terms of the 5d Green's function for charged particle propagation  $n$  times around the circle (see appendix A of [71]), in which case the power law is just the usual cost of propagating a massless field over the long distance  $2\pi Rn$ .

One can partially constrain this small action loophole by demanding that the convex hull condition should be satisfied for a 4d theory that includes both the usual U(1) and a Kaluza-Klein U(1). As shown in [56], this convex hull condition yields the inequality:

$$m_0 R \geq \frac{1}{2z_0 (z_0^2 - 1)^{1/2}}, \tag{2.10}$$

where

$$z_0 = \sqrt{\frac{3}{2}} \frac{eqM_{\text{Pl}}}{m_0} = \sqrt{\frac{3}{2}} \frac{qM_{\text{Pl}}}{2\pi f m_0 R}. \tag{2.11}$$

Here  $m_0$  is the mass of the particle in 5d, and we have turned off the dilaton in the 5d theory. The convex hull condition for the 5d theory just enforces  $z_0 \geq 1$ , which is equivalent to (2.9). For  $q = 1$ , we see that  $f \sim (m_0 R)^{-1}$ , so the maximal value of the decay constant grows inversely with the instanton action  $S = 2\pi m_0 R$  in the limit  $S \rightarrow 0$ . However, imposing the stronger condition (2.10) and looking at the  $m_0 R \rightarrow 0$  limit, we find

$$f^2 \leq \frac{3M_{\text{Pl}}^2}{(2\pi)^2 m_0 R}. \tag{2.12}$$

This tells us that the maximal allowed value of  $f$  grows like  $S^{-1/2}$ , rather than the naively expected  $S^{-1}$ . Thus, the weak gravity bound on axion decay constants in the context of extranatural inflation is stronger than expected, but it is not strong enough to close the small action loophole — one may achieve a super-Planckian decay constant by taking  $S$  small without violating the electric form of the WGC.

### 2.5 WGC and single-axion inflation: magnetic argument

The small action loophole led [48] to propose a different argument for how the WGC can exclude single-axion extranatural inflation. Starting with the UV cutoff  $\Lambda \lesssim eM_{\text{Pl}}$ , they demand that the size of the compactification manifold be larger than  $\Lambda^{-1}$ . Then

$$1 \lesssim 2\pi R \Lambda \lesssim 2\pi R e M_{\text{Pl}} = \frac{M_{\text{Pl}}}{f}. \tag{2.13}$$

In this way we obtain the constraint  $f \lesssim M_{\text{Pl}}$ , completely independent of the size of the leading instanton effect or which charged particle generates it. Notice that because the cutoff is  $\Lambda < eM_{\text{Pl}}/q_m$  when the monopole satisfying magnetic WGC has charge  $q_m$  and a particle of electric charge  $q$  leads to an instanton proportional to  $\cos(qa/f)$ , choosing nonminimal charge assignments anywhere in the argument only makes the bound stronger.

## 3 Weak Gravity Conjecture and multi-axion models

In this section we will extend the magnetic argument against single-axion inflation from section 2.5 to more general scenarios with multiple axion fields. We will first tackle the

simplest case of  $N$ -flation with the strong assumption of diagonal kinetic terms and minimal charge assignments. Then we relax our assumptions to consider alignment models.

### 3.1 Warmup: diagonal $N$ -flation

Suppose that we have  $N$   $U(1)$  gauge fields  $A_i$  with couplings  $e_i$  and that the kinetic mixing among them is negligible. We will also assume that in this basis the charge lattice simply consists of integer electric or magnetic charges under each  $U(1)$ . These are strong simplifying assumptions, but provide a useful starting point. The Weak Gravity Conjecture applied to each gauge field separately implies the existence of electrically and magnetically charged particles satisfying certain bounds. But the constraint for the set of  $N$  fields is actually stronger than for any individual field: if we marginally saturate the bound for each  $U(1)$  by postulating a magnetic monopole with charge  $q_m = 1$  and mass  $\sqrt{2}M_{\text{Pl}}/e_i$ , for example, then a nearly-extremal black hole with large and equal charges under every  $U(1)$  will not be able to decay. This is because the extremality bound for a black hole charged under multiple groups depends not on the sum of the charges but on the charges added in quadrature: for a magnetically charged black hole in four dimensions, the bound is  $Q_{\text{eff}} \equiv \sqrt{Q_1^2/e_1^2 + \dots + Q_N^2/e_N^2} < M_{\text{BH}}/(\sqrt{2}M_{\text{Pl}})$ .

Consider extranatural  $N$ -flation that moves along the diagonal in each axion direction, attempting to obtain an effective decay constant

$$f_{\text{eff}}^2 = f_1^2 + f_2^2 + \dots + f_N^2 = \left(\frac{1}{2\pi R}\right)^2 \left(\frac{1}{e_1^2} + \dots + \frac{1}{e_N^2}\right). \quad (3.1)$$

Notice that this is the appropriate expression under the assumption of no kinetic mixing *and* the further assumption that the dominant instanton effects give rise to a potential of the form  $\sum c_i \cos(a_i/f_i)$ , as would be generated for example from wrapped worldlines of electrically charged particles of charge 1 under each gauge group. More general instantons can lead to alignment phenomena in which inflation winds around one direction in axion space multiple times. We will return to such a possibility later.

The linear combination of  $1/e_i^2$  factors appearing on the right-hand side of (3.1) is precisely what appears in the extremality bound for a magnetically charged black hole with equal charge  $Q$  under all  $N$  gauge groups. Let us build some intuition by considering ways that such a diagonally magnetically charged black hole could decay:

- It could emit a monopole of diagonal charge  $(q, q, \dots, q)$ , so that its charge-to-mass vector points in the same direction after the emission but is now shorter (we take  $Q > q > 0$ ). In this case, the problem essentially reduces to the one-field case. The self-energy of the monopole imposes  $m_{\text{mon}} \gtrsim q^2 \sum_i \frac{1}{e_i^2} \Lambda$ , while the condition that the black hole moves away from extremality imposes that  $m_{\text{mon}}^2 \lesssim q^2 \sum_i \frac{1}{e_i^2} M_{\text{Pl}}^2$ . These conditions together with  $2\pi R\Lambda \gtrsim 1$  require that  $f_{\text{eff}} \lesssim M_{\text{Pl}}/q$ .
- It could emit a monopole charged under a single gauge group. Suppose it emits a particle with mass  $m_1$  and charges  $(q_1, 0, \dots, 0)$ . The self-energy constraint leads to  $m_1 \gtrsim q_1^2 \Lambda / e_1^2$ . If the diagonally-charged black hole emits this particle, its effective

charge decreases only by (expanding the square root)  $-\Delta Q_{\text{eff}} \approx Qq_1/(e_1^2 Q_{\text{eff}}) = (q_1/e_1)(f_1/f_{\text{eff}})$ . As a result, the condition that the monopole can be emitted is no longer  $m_1 < \sqrt{2}q_1/e_1 M_{\text{Pl}}$  but the stronger condition  $m_1 < \sqrt{2}q_1/e_1 (f_1/f_{\text{eff}}) M_{\text{Pl}}$ . This leads to  $f_{\text{eff}} \lesssim M_{\text{Pl}}/q_1$ .

- Now consider the general case in which the monopole emitted has mass  $m$  and charges  $(q_1, \dots, q_N)$ . For the black hole to move away from extremality we first require that  $Q_{\text{eff}}$  decreases, so that  $\sum q_i/e_i^2 > 0$ . A straightforward generalization of the previous argument leads to a bound

$$f_{\text{eff}} \lesssim M_{\text{Pl}} \frac{\sum_i q_i/e_i^2}{\sum_i q_i^2/e_i^2} \lesssim M_{\text{Pl}}. \tag{3.2}$$

The last step follows because charge quantization demands that  $q_i^2 > |q_i|$ .

These arguments give a suggestive hint of how scenarios with multiple axions can be more strongly constrained by the Weak Gravity Conjecture in a manner that precisely compensates for the expected gain in field range. However, we have made a strong simplifying assumption that the electric charges leading to dominant instanton effects are simple in the same basis that the gauge field kinetic term is diagonal. We will now explore the constraints imposed by the WGC if we relax this assumption.

### 3.2 Magnetic WGC and kinetic alignment

Consider the case of a general kinetic matrix for the gauge fields:

$$-\frac{1}{4} K_{ij} F_{\mu\nu}^i F^{j\mu\nu}. \tag{3.3}$$

Assume that we are working in a basis in which there are  $N$  magnetic monopoles that satisfy the magnetic WGC and have unit charges  $(1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, 0, \dots, 1)$ . We can choose a different basis to diagonalize the kinetic terms:

$$K = O D O^T, \tag{3.4}$$

where  $O$  is an orthogonal matrix and  $D = \text{diag}(1/g_1^2, \dots, 1/g_N^2)$  is a diagonal matrix. Without loss of generality we can choose

$$g_{\text{min}}^2 \equiv g_1^2 \leq g_2^2 \leq \dots \leq g_N^2 \equiv g_{\text{max}}^2. \tag{3.5}$$

In this basis, the  $i^{\text{th}}$  monopole has charge assignments  $\vec{o}_i$  that can be read off from the matrix  $O$ . The self-energy of this monopole gives us an inequality relating the UV cutoff  $\Lambda$  and the monopole mass  $m_i$ :

$$\left( \sum_j \frac{o_{ij}^2}{g_j^2} \right) \Lambda \lesssim m_i. \tag{3.6}$$

In particular, if we sum over all  $i$  and exploit orthogonality, we learn that

$$\sum_j \frac{1}{g_j^2} \Lambda \lesssim m_1 + \dots + m_N \equiv m_{\text{tot}}. \tag{3.7}$$

The magnetic WGC tells us that the convex hull of the charge-to-mass vectors  $\pm \vec{z}_i$  of the  $N$  monopoles contains the unit ball. We have

$$(\vec{z}_i)_j = \frac{o_{ij} M_{\text{Pl}}}{g_j m_i}. \quad (3.8)$$

Since the  $\vec{z}_i$  form a basis, the convex hull condition can be restated as the requirement that for any coefficients  $\alpha_1, \dots, \alpha_N$ ,

$$\left| \sum_i \alpha_i \vec{z}_i \right| \geq \sum_i |\alpha_i|. \quad (3.9)$$

Consider the choice  $\alpha_i = \sigma_i m_i$ , where  $\sigma_i = \pm 1$  is a choice of sign. In this case the convex hull condition tells us that

$$m_1 + \dots + m_N \leq \sqrt{\sum_j \left( \sum_i \frac{\sigma_i o_{ij}}{g_j} \right)^2} M_{\text{Pl}}. \quad (3.10)$$

Combining the convex hull condition with the constraint on the cutoff, we learn that for any set of sign choices  $\sigma_i$

$$\sum_j \frac{1}{g_j^2} \Lambda \lesssim \sqrt{\sum_j \left[ \frac{1}{g_j^2} \left( \sum_i \sigma_i o_{ij} \right)^2 \right]} M_{\text{Pl}}. \quad (3.11)$$

There are  $2^N$  choices of sign  $\sigma_i$ , some of which potentially provide much stronger bounds than others.

Consider the case where the largest eigenvalue completely dominates, so that we can drop all terms in the sum not proportional to  $\frac{1}{g_1^2}$ . In [23], it was pointed out that the eigenvector with largest eigenvalue  $1/g_1^2$  of a randomly chosen kinetic matrix  $K_{ij}$  will almost certainly point close to a diagonal direction of the fundamental cube, e.g.  $\sim (1, 1, \dots, 1)/\sqrt{N}$ . Since the diagonal of an  $N$ -dimensional cube has length  $\sqrt{N}$ , this implies (in the extranatural context with minimal instanton charges) an effective decay constant of  $f_{\text{eff}} \approx \sqrt{N}/(2\pi R g_1)$  in the direction of largest eigenvalue. However, in this case, we can choose the signs  $\sigma_i$  to alternate and nearly cancel so that the sum  $\sum_i \sigma_i o_{i1} \sim \frac{1}{\sqrt{N}}$ . This leads to an estimated bound

$$\frac{1}{g_1^2} \Lambda \lesssim \sqrt{\frac{1}{g_1^2 N}} M_{\text{Pl}} \Rightarrow \Lambda \lesssim \frac{g_1 M_{\text{Pl}}}{\sqrt{N}}. \quad (3.12)$$

This bound is larger than the naive one-field version of the magnetic WGC by a factor of  $\sqrt{N}$ , so imposing that  $\Lambda R \lesssim 1$  precisely produces  $f_{\text{eff}} \lesssim M_{\text{Pl}}$ .

It is possible to make a more general version of the argument that excludes any case in which one eigenvalue  $g_1$  dominates the sums (3.11), without making an assumption about the eigenvector. The cutoff  $\Lambda$  in this case obeys

$$\Lambda \lesssim g_1 \left| \sum_i \sigma_i o_{i1} \right| M_{\text{Pl}}. \quad (3.13)$$

On the other hand, the 4d Lagrangian for the axions is given by

$$\frac{1}{2(2\pi R)^2} D_{ij} \partial_\mu \theta_i \partial^\mu \theta_j - \sum_i A_i e^{-S_i} \cos \left( \sum_j o_{ij} \theta_j \right), \quad (3.14)$$

where  $D$  is the diagonal matrix of (3.4). In the limit in which  $1/g_1^2$  is much larger than the other eigenvalues of  $D$ , we may approximate the axion moduli space radius by considering only the field displacement in the direction of largest eigenvalue  $|\Delta\theta_1|$ . The maximal displacement is given by the largest value of  $|\theta_1|$  satisfying the conditions

$$|o_{i1} \Delta\theta_1| \leq \pi,$$

for all  $i$ . Thus,

$$|\Delta\theta_1|_{\max} = \frac{\pi}{\text{Max}_i |o_{i1}|}. \quad (3.15)$$

Combining this with our previous bound (3.13) and setting  $\Lambda R \gtrsim 1$ , we get a bound on the radius of axion moduli space,

$$r \lesssim \frac{\Lambda}{2\pi g_1} |\Delta\theta_1|_{\max} \lesssim \frac{1}{2} \left| \frac{\sum_i \sigma_i o_{i1}}{\text{Max}_i |o_{i1}|} \right| M_{\text{Pl}} \quad (3.16)$$

Finally, it is not hard to see that we can choose the signs  $\sigma_i$  so that this fraction is smaller than 1. Order the  $o_{i1}$ 's in descending order of their magnitude. Set  $\sigma_1 = +1$ . Then, recursively define,

$$\sigma_k = \begin{cases} \text{sgn}(o_{k1}) & : \quad \sum_{i=1}^{k-1} \sigma_i o_{i1} < 0 \\ -\text{sgn}(o_{k1}) & : \quad \sum_{i=1}^{k-1} \sigma_i o_{i1} \geq 0 \end{cases} \quad (3.17)$$

Since the  $o_{k1}$ s are decreasing in magnitude, the partial sum  $|\sum_{i=1}^{k-1} \sigma_i o_{i1}|$  can never jump more than  $|o_{11}|$  by adding a new term. By picking the signs in this way, we ensure that we are always moving towards the origin, so the magnitude of the partial sums is necessarily decreasing with  $k$ . Since the magnitude of the first partial sum is just  $|o_{11}|$ , we see that the full sum must be smaller in magnitude than  $|o_{11}|$ . Thus, (3.16) gives

$$r \lesssim \frac{1}{2} M_{\text{Pl}}. \quad (3.18)$$

This excludes any kinetic alignment model with a single dominant large eigenvalue, again under the assumption that the instanton effects are controlled by minimal electric charges in the same basis for which the magnetic monopoles satisfying the WGC have minimal charge. A *parametric* violation of this assumption, such as instanton effects that are highly aligned, can evade our arguments. We will discuss such a case in section 5. The assumption of single-eigenvalue dominance, on the other hand, is made only for simplicity. It is straightforward to check in the two-axion case that the bound holds for completely arbitrary eigenvalues. Furthermore, simple numerical studies in which the kinetic matrix is chosen from a Wishart distribution,  $K_{ij} \sim W_N(\sigma^2, N)$ , reveal that indeed the radius of moduli space decreases with increasing  $N$ .

## 4 New conjectures on EFT over the moduli space

### 4.1 Exploring the moduli space: masses and Kaluza-Klein reduction

In this section we develop a new tool for constraining large-field axion models arising from extra dimensions, which opens an opportunity to obtain powerful constraints on models of axion monodromy. This approach relies, in part, on the nontrivial manner in which shift symmetries are realized in the effective theory. The potential energy in extranatural inflation (including string axion models) is a sum of cosine terms from instantons of various winding numbers, respecting an exact discrete shift symmetry. However, other terms in the effective theory preserve the shift symmetry in a less transparent way. Consider the case of a 4d axion obtained by dimensional reduction of a 5d 1-form gauge field, and suppose that in five dimensions there is a fermion  $\Psi$  with charge  $q$  under the gauge field. (The case of a charged scalar field is similar.) Its action is

$$\int d^5x \sqrt{-g} \left( i\bar{\Psi}\Gamma^M D_M \Psi + m_5 \bar{\Psi}\Psi + \frac{c}{\Lambda} D_M \bar{\Psi} D^M \Psi + \dots \right), \quad (4.1)$$

where  $D_M = \partial_M - iqA_M$ ,  $\Lambda$  is the UV cutoff of the theory, and the dots represent various higher-dimension operators. The five Dirac matrices  $\Gamma^M$  correspond to the usual 4d Dirac matrices together with  $-i\gamma^5$ . We emphasize that  $\Lambda$  is the scale at which the local, 5d abelian gauge theory breaks down. In particular, we have no guarantee of five-dimensional locality holding at distances shorter than  $\Lambda^{-1}$ .

We study this theory on a background of  $\mathbb{R}^{3,1} \times S^1$  with the fifth dimension having a periodic identification  $y \sim y + 2\pi R$  with a background gauge field  $A_5 = \frac{\theta}{2\pi R}$ . Although fixing  $A_5$  to be constant is a gauge choice, there is a gauge-invariant Wilson loop determined by  $\theta$  which is well-defined modulo  $2\pi$ . The compactified theory contains a term

$$\int d^5x \sqrt{-g} \frac{q\theta}{2\pi R} \bar{\Psi}\Gamma^5\Psi, \quad (4.2)$$

that we may think of as an effective mass (albeit one that depends on the spontaneous breaking of 5d Lorentz symmetry) which potentially decouples  $\Psi$  from the effective theory if  $\theta$  is large enough. This 5d term gives rise to a (CP-odd) mass term  $\propto i\theta\bar{\psi}\gamma^5\psi$  in the 4d theory, which can have important dynamical consequences when  $\theta$  is large. At first glance, such a mass term constitutes a hard breaking of the shift symmetry for  $\theta$  —even of the *gauged*  $\theta \rightarrow \theta + 2\pi$  symmetry! The resolution of this puzzle is that there is a monodromy in the Kaluza-Klein spectrum. Using the Kaluza-Klein decomposition  $\Psi(x, y) = \sum_{n=-\infty}^{\infty} \exp(iny/R)\psi_n(x)/\sqrt{2\pi R}$ , this action leads to a 4d effective theory

$$\mathcal{L}_{\text{eff}} = \sum_{n=-\infty}^{\infty} \left( i\bar{\psi}_n \gamma^\mu D_\mu \psi_n + m_5 \bar{\psi}_n \psi_n + i \frac{n - \frac{q\theta}{2\pi}}{R} \bar{\psi}_n \gamma^5 \psi_n + \frac{c}{\Lambda} \left| \frac{n - \frac{q\theta}{2\pi}}{R} \right|^2 \bar{\psi}_n \psi_n + \dots \right). \quad (4.3)$$

If we were to truncate this theory to a few low-lying modes, we would find a violation of the shift symmetry  $\theta \rightarrow \theta + 2\pi$ . But this symmetry is a large gauge transformation in the higher-dimensional UV completion, so it cannot be violated. Writing the EFT for



all Kaluza-Klein modes makes the answer manifest. There is a monodromy effect that rearranges the spectrum; when  $\theta \rightarrow \theta + 2\pi$ , the mode with label  $n$  acquires the same mass spectrum that the mode with label  $n - q$  previously had. Because the derivative  $\partial_5$  and the contribution of  $A_5$  are always packaged together in a covariant derivative, this will be true of arbitrary higher-dimension operators as well.

Recall that for a Dirac fermion with mass term  $m\bar{\psi}\psi + i\mu\bar{\psi}\gamma^5\psi$ , the physical mass is  $\sqrt{m^2 + \mu^2}$ . In particular, all 4d fields have mass larger than the 5d mass  $m_5$ .

## 4.2 Consistency of a single EFT across axion moduli space

We have seen that the 5d theory compactified on a circle with a Wilson loop  $\theta = \oint A_5 dx^5$  turned on has a spectrum that depends nontrivially on the value of  $\theta$ . Let us ask what happens when we move a large distance in moduli space. Tracking a single KK mode adiabatically as  $\theta$  varies, we find that its CP-odd mass is shifted by

$$\Delta m = \frac{q\Delta\theta}{2\pi R} \tag{4.4}$$

In particular, if  $\Delta\theta \gtrsim 2\pi R\Lambda/q$ , then a KK mode which is initially light acquires a large mass of order the cutoff  $\Lambda$ , and exits the effective theory. In fact, when we move this far in moduli space, the entire KK spectrum is shifted, so that the modes which were initially light are heavy, and modes initially above the cutoff are light. Since our description of the five-dimensional theory breaks down at  $\Lambda$  (and in particular 5d locality may not hold above this scale), it is possible that in the process new physics can emerge from the cutoff and become light, ruining our effective description. Thus, if we wish to retain control of the KK spectrum, we should impose:

$$q \Delta\theta \lesssim 2\pi R\Lambda. \tag{4.5}$$

We emphasize that this is *not* a statement about the 4d effective theory cut off at the compactification scale, which obviously does not include Kaluza-Klein modes that may be important elsewhere in the moduli space. It is a statement about the 4d theory including a tower of weakly coupled modes all the way up to the cutoff  $\Lambda$ , which is fully equivalent to the 5d theory on the Wilson loop background. One point that we should emphasize is that the breakdown of effective field theory that we are discussing does *not* correspond to a violation of perturbative unitarity in high-energy scattering in 5d. The Wilson loop is gauge-invariant only when integrated over the full circle, so short-distance 5d scattering experiments do not detect it. Local scattering experiments are not the only way to detect a failure of EFT, however, and the KK mode spectrum is a physical observable that does so.

One clear instance in which a subtlety of this kind *does not* arise is when the path in moduli space that we have taken winds many times around a small periodic circle (without monodromy). In this case, the exact shift symmetry of the axion ensures that nothing dramatic can occur. However, we emphasize that inflation requires a motion in moduli space which is not periodic, either due to monodromy or because the size of the circle is large. In this case, the shift symmetry does not help.

This suggests an alternate perspective on the problem. The periodicity of  $\theta$  arises because we can do a large gauge transformation  $A_M \rightarrow A_M - \partial_M \chi$  for which  $\chi$  is not single-valued on the circle but  $e^{i\chi}$  is. In particular, we can identify  $\theta = 2\pi$  with  $\theta = 0$  by performing the transformation

$$A_5 \rightarrow A_5 - 1/R, \quad \chi = y/R, \quad \Psi \rightarrow e^{-iqy/R} \Psi. \tag{4.6}$$

While this appears at first glance to be a completely innocent operation, notice that if we are working within an effective field theory with UV cutoff  $\Lambda$ , this large gauge transformation can bring in modes that are outside the validity of our effective field theory. In particular, if we do not require

$$\frac{q}{R} \lesssim \Lambda, \tag{4.7}$$

then the low-frequency modes of the gauge-transformed  $\Psi$  field involve very high-frequency modes of the original field, and vice versa. If we do not require (4.7), then even the periodicity of  $\theta$  becomes a subtle question in the low-energy theory!

Heuristically, another way to see a problem with these large field ranges is to consider the effective mass for  $\Psi$  (4.2):

$$\int d^5x \sqrt{-g} \frac{q\theta}{2\pi R} \bar{\Psi} \Gamma^5 \Psi. \tag{4.8}$$

If (4.5) is violated then  $\Psi$  receives an effective mass which removes it from the low-energy effective field theory. Of course, the full term involves  $\partial_5 - iqA_5$ , so the large mass obtained from the Wilson loop can be compensated by high-frequency oscillations in  $y$ , but these high-frequency modes are not part of the EFT that we started with at the origin of moduli space. We elaborate on this point in appendix A, using a manifestly gauge-invariant lattice regulator to explore how physical quantities can depend on the cutoff if  $\Lambda R$  is not large compared to  $q$ . Large effective masses far out on the moduli space are particularly suspect in cases where  $\Psi$  plays an important dynamical role. For instance, if  $\Psi$  provides one of the dominant instanton contributions to the potential, what does it mean to compute  $V(\theta)$  for a value of  $\theta$  for which it is inconsistent to keep track of the particle generating the potential? If  $\Psi$  is a field that is necessary to satisfy the electric WGC, decoupling it from the effective theory is inconsistent with the EWGC.

We propose a new constraint on theories of extranatural inflation based on this consistency requirement. One statement of the constraint is the following:

*Single-EFT Consistency Criterion (SECC):* in order to have a controlled description of a portion of the moduli space within a single effective field theory, we demand that any field which is part of the EFT at one point of the moduli space is not decoupled by terms like (4.2) in a different region of the moduli space. Equivalently, if a Kaluza-Klein mode is light somewhere in the moduli space, this mode should exist within the effective theory at the origin of moduli space. This constrains  $R\Lambda$  to satisfy (4.5).

Loosely, in a controlled theory a mode cannot appear “out of the blue.” This seems to us to be a sufficiently well-motivated criterion that it is worthwhile to explore its consequences. An equivalent statement, if we want to describe the *entire* moduli space in a single EFT, is:

*Single-EFT axion periodicity criterion:* the periodic identification of 4d axions arising from an underlying higher-dimensional gauge theory should arise from large gauge transformations that are well-defined within the higher-dimensional EFT. Specifically, if the theory has a UV cutoff  $\Lambda$ , then fields which are smooth on scales much larger than  $\Lambda^{-1}$  should not oscillate on length scales shorter than  $\Lambda^{-1}$  after the gauge transformation.

The SECC assumes that we should be able to work with a single well-defined 5d effective field theory. One might imagine a patchwork of effective field theories, each valid over a limited range of  $\theta$ , which are matched onto each other in overlapping regimes. Nothing intrinsically seems to prevent us from considering the 5d theory on a Wilson line background with any particular value of  $\theta$ ; what we have seen is that connecting the theories at different values of  $\theta$  may be difficult. One might consider the case of Seiberg-Witten theory [72], in which vacua with weakly coupled electrons and with weakly coupled monopoles cannot coexist in the same EFT from the IR point of view but are guaranteed to be smoothly joined together due to well-understood UV physics. Our claim is that because our 5d theory came with a built-in cutoff at  $\Lambda$ , we do not actually have such a sharp understanding of the UV physics in this case. It may exist if we embed the 5d theory in a more complete UV setting.

If the large gauge transformations that guarantee an identification  $\theta \sim \theta + 2\pi$  in the four-dimensional effective field theory are not actually valid operations in the UV completion that we started with, this suggests that we do not truly have a controlled theory of axions. In such a case it is unclear what a computation of the axion potential as a periodic function of the  $\theta$ 's even means. Nonetheless, we cannot give any fully rigorous argument in favor of the Single-EFT Consistency Criterion. In this paper, we will explore the consequences of the SECC, while welcoming debate on its merits.

### 4.3 Consequences of Single-EFT Consistency for monodromy

The Single-EFT Consistency Criterion, in the form of the bound (4.5), is a significant potential obstacle to any model based on axion monodromy. To see why, consider any model in which an axion field winds around the circle  $N$  times in the presence of monodromy. We have the constraint

$$N \lesssim R\Lambda \tag{4.9}$$

from the SECC. But we also have, from the magnetic form of the WGC, the additional constraint

$$\Lambda \lesssim eM_{\text{Pl}}. \tag{4.10}$$

These conditions together with  $f = 1/(2\pi R e)$  imply

$$Nf \lesssim M_{\text{Pl}}, \tag{4.11}$$

so the effective *total* field range from winding  $N$  times around the circle is *still* bounded above by the Planck scale.

Monodromy was important in this argument. Without monodromy, if the physical state were *exactly* the same after each trip around the circle, we could get away with only requiring that a gauge transformation  $\theta \rightarrow \theta + 2\pi$  is well-defined (and then repeat it  $N$  times) rather than that a larger field range  $\Delta\theta \sim 2\pi N$  is accessible within the effective theory.

Although we have phrased our argument in terms of 1-form gauge fields in five dimensions, a similar constraint will arise from the SECC for the more general  $p$ -form models. Just as our charged field  $\Psi$  obtained an effective Lorentz-violating mass  $\sim A_5 \bar{\Psi} \Gamma^5 \Psi$  in the presence of a background gauge field, the presence of a background  $p$ -form will add a Lorentz-violating tension term to the worldvolume effective theory of a  $(p - 1)$ -brane, potentially decoupling it from the effective field theory.

The SECC argument against monodromy is not airtight. In section 5 we will consider a two-axion model of inflation in which one axion winds  $N$  times around the circle, but we will see in section 5.1 that this does not necessarily violate the SECC. The reason is that there is a compensating contribution to the mass of the charged fields coming from the second axion. This gives some insight into how a monodromy model might successfully escape the SECC. However, in the model we will discuss, the existence of any charged fields with *different* charges than those producing the dominant instanton effects will restore the power of the SECC, whereas the one case that evades the SECC is constrained by a different requirement that we will formulate in section 4.4.

#### 4.4 The Weak Gravity Conjecture across moduli space

There is one other conjecture, in a similar spirit to the SECC but differing in its details and its implications, that is worth considering:

*Extended Weak Gravity Conjecture (XWGC):* the weak gravity conjecture should be satisfied at any stationary point of the potential.

In fact, we will only use this condition applied at extrema of the potential, rather than generic stationary points. We could also have imposed a stronger condition that the WGC holds *everywhere* in the moduli space, though at least for the cases we consider the results would be equivalent.

This provides a new viewpoint on the small-action loophole. A charged particle with 5d mass  $m_5 = 0$  obviously satisfies the (electric) WGC in five dimensions, and we saw that such a particle can generate a potential compatible with inflation despite the lack of exponential suppression of higher harmonics. However, the XWGC demands that the electric WGC is satisfied also at stationary points of the potential away from  $\theta = 0$ . These are classically stable states, but those that are not local minima will eventually tunnel away from the critical point. Because tunneling can be a slow process, and charged black

holes discharge quickly when the WGC is satisfied [58–60], it seems plausible that the WGC should hold even in these unstable states. As mentioned in the introduction, we suspect that the requirement that black holes decay is actually a dynamical requirement that they shed charge often enough relative to uncharged Hawking quanta, rather than a simple kinematic statement that they can decay at all. Further work on black hole thermodynamics may help to justify or refute the XWGC by quantifying the timescale on which we require charge to be lost.

The Kaluza-Klein modes have masses spaced by  $1/R$ , so at the maximum of the potential  $\theta = \pi/q$  the masses are maximally shifted and the lightest electrically charged particle has  $m = 1/(2R)$ , or larger if we begin with  $m_5 \neq 0$ . Let us assume, as in section 2.4, that the same particle which generates the leading contribution to the axion potential satisfies the XWGC. Let us first give a simple heuristic argument for why the XWGC could close the small-action loophole. For a particle of charge  $q$ , we obtain the bound:

$$\frac{1}{2R} < \sqrt{2}qeM_{\text{Pl}} = \frac{qM_{\text{Pl}}}{\sqrt{2\pi Rf}} \Rightarrow \frac{f}{q} < \frac{\sqrt{2}M_{\text{Pl}}}{\pi}. \quad (4.12)$$

Assuming that the same particle generates the axion potential,  $f/q$  is precisely the effective axion field range, and the small action loophole appears to have been closed without invoking the magnetic form of the conjecture.<sup>2</sup>

The argument we have just given ignores an important effect. The compactification on the circle produces a second U(1) gauge field, namely the KK U(1) arising from graviton modes with one leg on the circle. At nonzero values of  $\theta$ , the two U(1) gauge fields mix and the correct WGC to consider is the convex hull condition applied to our original U(1) gauge theory and the Kaluza-Klein U(1). We present a detailed derivation of this statement and discussion of the mixing effect in [56]. The weak gravity bound becomes

$$m^2 \leq \gamma e^2 q^2 M_{\text{Pl}}^2 + \frac{g_{\text{KK}}}{R^2} \left( n - \frac{q\theta}{2\pi} \right)^2. \quad (4.13)$$

The constant  $\gamma$  is 2 as above if the radion mode is stabilized, but is 3/2 if the radion is unstabilized. Similarly, the constant  $g_{\text{KK}}$  is 1 for an unstabilized radion and 2 for a stabilized radion. If a 5d particle obeys the WGC, then any of its KK modes in 4d will in fact obey this inequality for any value of  $\theta$ , undermining the heuristic argument we gave above. However, our conclusion survives once we take the convex hull condition into account.

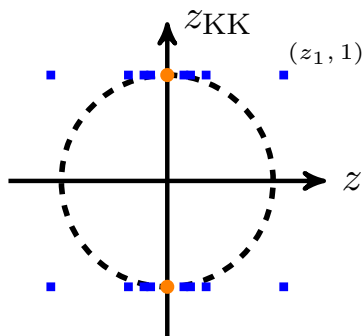
The reason we obtain a bound from the convex hull condition is that there is not a KK mode in every direction in charge space. We want to apply the convex hull condition to the charge-to-mass vectors

$$\vec{z} = (z, z_{\text{KK}}) = \frac{1}{m(n, \theta)} \left( \sqrt{\gamma}eqM_{\text{Pl}}, \frac{g_{\text{KK}}}{R} \left( n - q\frac{\theta}{2\pi} \right) \right), \quad (4.14)$$

$$m(n, \theta) = \sqrt{m_5^2 + \frac{1}{R^2} \left( n - q\frac{\theta}{2\pi} \right)^2}. \quad (4.15)$$

---

<sup>2</sup>The argument based on the magnetic WGC is still somewhat stronger, because we don't need to assume that the charged particle which generates the leading contribution to the axion potential has any other special role to play.



**Figure 1.** How the XWGC closes the small-action loophole when  $\theta = \pi$ . We depict the case  $g_{\text{KK}} = 1$  for convenience. The horizontal axis is the charge-to-mass ratio  $z$  for the U(1) gauge group giving rise to the axion. The vertical axis is the charge-to-mass ratio for Kaluza-Klein charge. The points on the vertical axis (orange circles) correspond to graviton KK modes. The points off the axis (blue squares) correspond to charged particle KK modes, which as  $n \rightarrow \infty$  accumulate near the orange points. We see that the convex hull condition demands that the horizontal coordinate  $z_1$  at  $n = 1$  be  $\geq 1$ , leading to the bound  $f < \frac{\sqrt{\gamma}}{\pi} M_{\text{Pl}}$ .

Without loss of generality, we specialize to the case  $q = 1$  for simplicity and set  $f = 1/(2\pi eR)$ . We choose  $\frac{\theta}{2\pi} = 1/2$  so that the KK charge of the particles is  $n - \frac{\theta}{2\pi}$ , an odd half-integer. We take  $m_5 = 0$  to study the small-action loophole. Any other value of  $m_5$  will, for fixed charges  $q$  and  $n$ , lead to a shorter vector  $\vec{z}$  and thus a tighter constraint. We have a set of charge-to-mass vectors

$$\vec{z}_n = \left( \frac{1}{n - \frac{1}{2}} \frac{\sqrt{\gamma} M_{\text{Pl}}}{2\pi f}, \text{sgn}\left(n - \frac{1}{2}\right) g_{\text{KK}} \right). \quad (4.16)$$

We also have a set of charge-to-mass vectors from KK gravitons or dilatons, which are uncharged under the U(1) but carry KK charge, and for unstabilized radion will always saturate the WGC for their direction in the charge lattice:

$$\vec{z}_{\text{grav};n} = (0, \text{sgn}(n) g_{\text{KK}}). \quad (4.17)$$

All of these vectors are outside the open unit ball, so in the direction of any  $\vec{z}_n$ , we satisfy WGC. But of course the striking thing about these vectors is that they all have a “ $\pm g_{\text{KK}}$ ” in the second entry. That is, the KK charge always satisfies the WGC bound (saturating it when the radion is unstabilized), and we’re at a point on the moduli space where *every* charged particle has KK charge due to the axion effect.

The convex hull requirement for these vectors is depicted in figure 1. We see that the requirement that the unit circle is contained in the convex hull imposes  $z_1 \geq 1$ , i.e.

$$f < \frac{\sqrt{\gamma}}{\pi} M_{\text{Pl}}. \quad (4.18)$$

If this requirement is not satisfied, a black hole with charge vector  $(Q, 0)$  under the two U(1) gauge groups will not be able to decay. Notice that we can easily construct such charge vectors from combinations of KK modes with opposite signs of the KK number. Thus, we

see that the XWGC (in its convex hull formulation) closes the small-action loophole for a single axion, despite the subtleties introduced by the KK U(1).

It is clear from figure 1 that, because the value of  $z_{\text{KK}}$  is the same for every state, we can simply project the problem down to the lower-dimensional problem of considering only  $z$ . This will continue to be true in a scenario where we consider multiple U(1) gauge bosons: a theory of  $n$  gauge bosons in 5d gives rise to an  $(n + 1)$ -dimensional convex hull problem in 4 dimensions, but so long as we go to a point on the moduli space where every charged particle in the theory has a common value (up to a sign) for  $z_{\text{KK}}$ , the added KK direction can be projected out of the argument. We will exploit this below in discussing a scenario with two U(1) groups in 5d.

Our main motivation for studying the SECC and XWGC is to apply them to a version of alignment inflation which the usual WGC is not strong enough to exclude. We will consider this scenario in the next section.

### 4.5 The Lattice Weak Gravity Conjecture

In [56], we introduced another form of the WGC:

*The Lattice Weak Gravity Conjecture (LWGC):* at every point in the charge lattice, there exists a state that satisfies the WGC.

This conjecture is obeyed, for instance, by string states of the SO(32) heterotic string as well as Kaluza-Klein reduction of pure gravity on a torus.

At first, this conjecture seems to highlight an apparent loophole in our arguments against  $N$ -flation in section 3. If there is an instanton for every point on the charge lattice lying outside the unit ball, then the leading instantons can lie on or just outside the unit ball, implying actions which are larger by a factor of  $\sqrt{N}$  relative to the case where only  $N$  instantons satisfy the convex hull condition.

However, the large number of subleading instantons required by the LWGC are sufficient to spoil the flatness of the potential. In particular, consider a theory with  $N$  axions that just marginally satisfies the LWGC in every possible direction. This implies the existence of infinite tower of instantons of every possible integral charge  $\vec{Q} = (Q_1, \dots, Q_N)$ . Further, setting the actions of the leading instantons to be  $\mathcal{O}(1)$ , and setting each of the decay constants to be Planckian,  $f_i \sim M_{\text{Pl}}$ , we have

$$f_{\text{eff}} \sim \sqrt{N} M_{\text{Pl}}. \tag{4.19}$$

The convex hull condition is satisfied because of the infinite number of subleading instantons, which densely fill the unit sphere. A necessary condition for this inflationary model is that the subleading instanton actions must scale with the instanton charges as

$$S_{\vec{Q}}^2 \sim \sum_i Q_i^2. \tag{4.20}$$

If, on the other hand, the actions were to grow linearly with the charges,  $S_{\vec{Q}} \sim \sum_i |Q_i|$ , then the subleading instantons would not densely fill in the unit sphere but instead a cube

of diagonal length  $M_{\text{Pl}}$  centered at the origin. This would not contain the unit ball, so the convex hull condition would be violated.

However, the particular growth of the instanton actions in (4.20) that allows this scenario to satisfy the LWGC is also what leads to its downfall. The instanton actions are smaller in this scenario than in a model in which the instanton actions grow linearly, which means their contributions to the potential are larger. We will now show that this enhancement yields large corrections to the inflationary potential, making it unsuitable for inflation.

For this, we need only consider the special class of instantons with charges  $(\pm 1, \dots, \pm 1)$ . Of course, there are many more instantons that will contribute to the inflationary potential, but for our purposes it suffices to show that just the instantons in this special class combine to give large contributions. There are  $2^N$  such instantons, which have action  $S \approx \sqrt{N} S_{\text{leading}}$  relative to the actions of the leading instantons,  $S_{\text{leading}}$ . The  $N$ -flation potential generated by the leading instantons is of the form

$$V(\phi_i) \supset \mathcal{A} e^{-S_{\text{leading}}} \cos(\phi_i/f) \tag{4.21}$$

The subleading instantons under consideration, on the other hand, give a potential of the form

$$V(\phi_i) \supset \mathcal{A} e^{-\sqrt{N} S_{\text{leading}}} \cos\left(\sum_i \eta_i \phi_i/f\right). \tag{4.22}$$

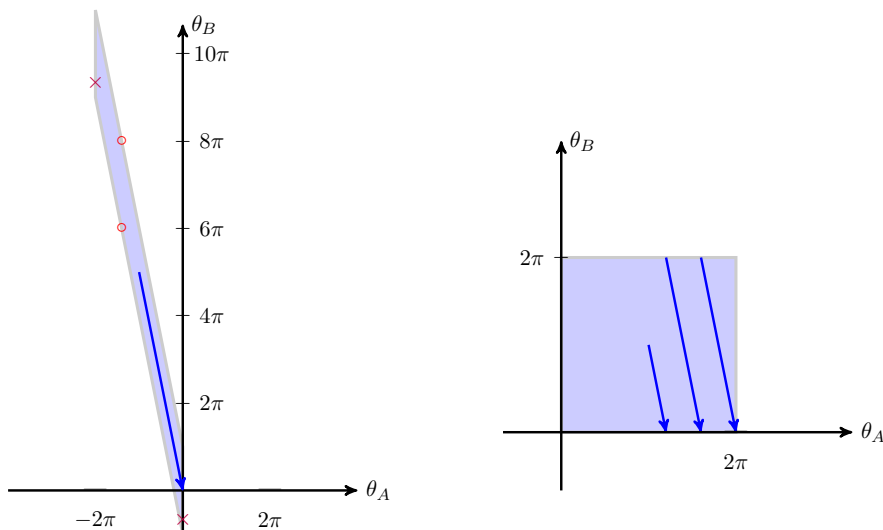
Here  $\eta_i = \pm 1$  depending on which instanton is being considered. We consider inflation along the diagonal direction  $\phi_i = \phi$ . A necessary condition for inflation is that the potential contributions from these instantons must in fact be negligible compared to those from the leading instantons. However, there are  $2^N$  instantons of the form in question, and each one introduces a potential contribution of magnitude  $e^{-\sqrt{N} S_{\text{leading}}}$ . Thus, the total potential contribution from these instantons grows roughly as  $2^N e^{-\sqrt{N} S_{\text{leading}}}$ . To protect the  $N$ -flation potential of (4.21) from these subleading effects at large  $N$ , we must therefore take  $S_{\text{leading}} \gtrsim \sqrt{N}$ . However, the LWGC implies  $f \lesssim M_{\text{Pl}}/S_{\text{leading}}$ , so  $f_{\text{eff}} \sim \sqrt{N} f \lesssim M_{\text{Pl}}$ . Hence, parametric enhancement of the effective decay constant via isotropic  $N$ -flation is inconsistent with the LWGC. The LWGC also restricts models of decay constant alignment, which we discuss in the following section.

## 5 Decay constant alignment

The arguments we have given in section 3 break down when the charge assignments of the dominant instanton effects are not small integers in the same basis where the magnetic monopole charges satisfying the magnetic WGC are. This case is especially subtle. For concreteness, we will focus the two-axion model of [48]. We assume a basis in which two compact axion fields,  $\theta_A$  and  $\theta_B$ , have diagonal kinetic terms with decay constants  $f_A$  and  $f_B$ :

$$\mathcal{L}_{\text{kin}} = \frac{1}{2} f_A^2 \partial_\mu \theta_A \partial^\mu \theta_A + \frac{1}{2} f_B^2 \partial_\mu \theta_B \partial^\mu \theta_B. \tag{5.1}$$





**Figure 2.** Two views of the fundamental domain (shaded) of the two axions for the case  $N = 5$ , together with a trajectory (thick blue arrow) beginning at a maximum of the potential and ending at the origin. For clarity, the view on the left and right are not drawn to the same scale. The right hand view is a “natural” parametrization with  $0 \leq \theta_{A,B} \leq 2\pi$  but requires that we discontinuously change the value of  $\theta$  and execute a corresponding monodromy on the Kaluza-Klein spectrum when wrapping around the torus. The left-hand view chooses a parametrization in which the values of  $\theta_{A,B}$  and the Kaluza-Klein masses change smoothly during all of inflation. To illustrate the periodic identifications imposed on the boundaries, we show two points labeled with a red  $\circ$  that are identified and two points labeled with a purple  $\times$  that are identified.

We further assume that in this basis the magnetic monopoles satisfying WGC have charge assignments  $(1, 0)$  and  $(0, 1)$ , leading to the constraints

$$f_A, f_B \lesssim M_{\text{Pl}}. \tag{5.2}$$

However, we assume that the dominant instanton effects arise from electric charges  $(1, 0)$  and  $(N, 1)$  that are highly aligned in this basis. As a result, the potential behave as

$$V(\theta_A, \theta_B) \approx V_0 (1 - \cos(\theta_A)) + \tilde{V}_0 (1 - \cos(N\theta_A + \theta_B)). \tag{5.3}$$

As emphasized in [48], this provides a UV setting for the alignment mechanism of Kim, Nilles, and Peloso [14] which appears to satisfy the WGC constraint. Inflation occurs on a trajectory for which  $\theta_A$  winds once around the circle while  $\theta_B$  winds  $N$  times, leading to an effective decay constant

$$f_{\text{eff}} \approx N f_B. \tag{5.4}$$

This inflationary trajectory is illustrated in figure 2. The instantons can be generated from worldlines of charged particles, but [48] also discusses a scenario in which the factor of  $N$  can be the level of a Chern-Simons coupling  $A \wedge G^a \wedge G^a$  in the 5d theory, potentially arising from a quantized flux in an even higher-dimensional theory.

### 5.1 Parametrizing the fundamental domain

If we consider the axion fields to range over  $0 \leq \theta_A, \theta_B \leq 2\pi$ , as depicted in the right-hand side of figure 2, then the periodic identification  $\theta_A \rightarrow \theta_A + 2\pi$  that wraps the left-hand side of the square onto the right-hand side shifts the mass of a state with charge  $(N, 1)$  by  $N/R$ , suggesting the possibility that a cutoff near  $1/R$  is too low for consistency. On the other hand, the inflaton trajectory winds multiple times, and we can think of this large mass shift as a consequence of the monodromies induced every time we wrap around the torus and shift our coordinate  $\theta$  discontinuously.

A more useful parametrization of the moduli space is to “unwind” it so that it is aligned with the inflaton trajectory, as in the left panel of figure 2. In this case, moving off the right edge of the space wraps back to the left edge at a lower point, corresponding to  $\theta_B \rightarrow \theta_B - 2\pi$  with  $\theta_A$  unchanged. The identification of the upper left edge with the bottom right edge corresponds to  $(-2\pi, \theta_B) \rightarrow (0, \theta_B - 2N\pi)$ . This is a large change in  $\theta_B$ , but it leaves  $N\theta_A + \theta_B$  fixed, so for a particle of charge  $(N, 1)$  there is no change in the mass spectrum for this transformation. Thus, if the only particles in our effective theory have charges  $(1, 0)$  and  $(N, 1)$ , the SECC imposes no obstacle to taking  $\Lambda \sim 1/R$ . The masses of the Kaluza-Klein modes of these fields are at most of order  $1/R$  throughout the moduli space.

This illustrates a general fact: the physical criterion that we would like to impose is that there is a single effective theory that is valid everywhere on the moduli space. Some parametrizations of the moduli space might obscure the existence of this theory, while others make it manifest.

### 5.2 Case 1: dominant instantons satisfy electric WGC

The first case we consider is that the same charges  $(1, 0)$  and  $(N, 1)$  that control the axion potential also are responsible for satisfying the electric WGC. The argument of [48], emphasized to us by the authors [65], is that once the magnetic WGC constraint that  $f_A, f_B \lesssim M_{\text{Pl}}$  is imposed, there is no further WGC constraint. The charged particles can be arbitrarily light, and direct calculation confirms that higher instanton contributions are numerically small. However, this changes if we impose our stronger XWGC conjecture from section 4.4. The potential has a local maximum where the arguments of both cosines are  $\pi$ , i.e.  $N\theta_A + \theta_B = \pi$  and  $\theta_A = \pi$ . At this point, the lightest KK modes for both charged particles have mass  $1/(2R)$  (assuming that  $m_5 = 0$ ). As in the XWGC discussion above, there is also a third, Kaluza-Klein,  $U(1)$  gauge group to consider, but at this point in moduli space every charged particle in the theory has  $z_{\text{KK}} = \pm g_{\text{KK}}$ , so the third dimension can be projected out of the argument. Thus the XWGC tells us that, assuming we started with the best-case scenario where the charged particles have negligible 5d masses, the charge-to-mass vectors at the maximum of the potential (and indeed at any generic point in moduli space) are of order

$$\begin{aligned} \vec{z}_1 &= \left( \frac{\sqrt{2}e_A M_{\text{Pl}}}{m_1}, 0 \right) \sim (e_A R M_{\text{Pl}}, 0) \sim \left( \frac{M_{\text{Pl}}}{f_A}, 0 \right), \\ \vec{z}_2 &= \left( \frac{\sqrt{2}N e_A M_{\text{Pl}}}{m_2}, \frac{\sqrt{2}e_B M_{\text{Pl}}}{m_2} \right) \sim \left( \frac{N M_{\text{Pl}}}{f_A}, \frac{M_{\text{Pl}}}{f_B} \right). \end{aligned} \tag{5.5}$$

Now we require that the convex hull of the vectors  $\pm\vec{z}_1, \pm\vec{z}_2$  contains the unit sphere. The line passing through the points  $(-\alpha, 0)$  and  $(N\alpha, \beta)$  in the  $(x, y)$  plane is  $(N+1)\alpha y = \beta(\alpha+x)$ , so if we read off where this crosses the  $y$ -axis, we obtain the constraint  $\beta \geq N+1$ . Substituting the vectors we're interested in,

$$\frac{M_{\text{Pl}}}{f_B} \gtrsim N+1 \Rightarrow f_{\text{eff}} = Nf_B \lesssim M_{\text{Pl}}. \tag{5.6}$$

This shows that our conjecture that the electric WGC should be satisfied at all stationary points in the axion moduli space is strong enough to exclude decay constant alignment in the scenario where the same particles are responsible for both satisfying the WGC and supplying the dominant instanton effects.

The LWGC postulates the existence of an instanton satisfying the WGC at every point on the charge lattice. If this version of the WGC is true, we must therefore consider models with additional instantons. We now turn our attention to these models.

### 5.3 Case 2: additional particles satisfy electric WGC

On the other hand, we could consider a different scenario. The reason we obtained such a large bound on  $f_B$  from the convex hull condition is that its charge-to-mass vector was nearly aligned with that of  $f_A$ . This was necessary to obtain a large field range in inflation, but what if we satisfy the convex hull condition (and hence the electric WGC) with *different* vectors than the ones that dominate the instanton contributions? This possibility was, again, suggested to us by the authors of [48]. Suppose that we have three relevant charges,  $(1, 0)$ ,  $(N, 1)$ , and  $(0, 1)$ . The first two supply the dominant instanton contributions, while the first and third satisfy the WGC. We can take the 5d mass of the  $(0, 1)$  particle to be somewhat large compared to  $1/R$ , so that its instanton contribution is suppressed, but not parametrically large by a factor of  $N$ , so that we do not shrink its mass-to-charge vector by enough to obtain the desired bound on  $f_B$ .

This is the point at which the SECC becomes crucial. We can consider which modes are present in our effective theory as we move around the moduli space. The particles with charge  $(q_A, q_B)$  have their masses shifted by  $\frac{1}{2\pi R}(q_1\theta_A + q_2\theta_B)$ . As discussed in section 5.1, in order to adiabatically track the mass of a particular mode, we should work in the “unwound” moduli space where the inflaton trajectory is continuous, as in the left-hand panel of figure 2. In this fundamental domain neither  $\theta_A$  nor  $N\theta_A + \theta_B$  is parametrically large, but  $\theta_B$  itself is. Precisely because the inflaton direction winds around the  $\theta_B$  circle multiple times, when we unwind the moduli space we find that  $0 \leq \theta_B \leq 2\pi N$ . Thus, the new particle of charge  $(0, 1)$  that we invoked to satisfy the electric WGC without running into difficulty with the convex hull condition is not part of a single consistent effective field theory defined over the entire axion moduli space unless we satisfy the bound

$$2\pi R\Lambda \gtrsim 2\pi N. \tag{5.7}$$

Again, this is the precise parametric bound that we need to obtain  $f_{\text{eff}} \lesssim M_{\text{Pl}}$ . In fact, in this particular case we do not even have to invoke the full SECC. We only need to require a

consistent set of modes *along the inflaton trajectory*. This leaves open the possibility that this weaker requirement and the XWGC are sufficient assumptions, without the full SECC.

Notice that this argument, unlike many of our previous arguments, relies on the EWGC: if the state of charge  $(1, 0)$  which satisfies the convex hull condition is never in the effective theory to begin with, then the SECC does not further constrain the model. However, so long as there is *some* effective field theory description of this state, the SECC will rule out the model, whereas theories which violate the EWGC may present other control problems, as explained in the introduction.

#### 5.4 Status of decay constant alignment

To summarize, we have excluded the particular example of decay constant alignment with charges  $(1, 0)$  and  $(N, 1)$  using the two conjectured constraints from section 4. The XWGC is crucial for avoiding the small-action loophole in which we use the same light fields to generate the dominant instanton effects and to satisfy the electric WGC. The SECC is needed for the case when the particles generating the instanton effects and satisfying the electric WGC are not the same. In both cases, the parametric constraint that we extract is *precisely* what is necessary to obtain a maximum field range of  $M_{\text{Pl}}$ . This is, at the least, very suggestive. It calls for further effort to understand how strongly motivated our conjectured constraints are. Again, the possible counterargument to the XWGC is that stationary points in the moduli space that are not minima have a finite lifetime before tunneling to a lower point on the potential, so arguments based on concerns about exactly stable remnants are not directly relevant. For the SECC, the concern is that despite the inconsistency of a single effective field theory, an ultraviolet completion may somehow allow a patchwork theory to be constructed. We do not have definitive rebuttals to these possible counterarguments, but our constraints appear plausible and well-motivated to us, and it is very interesting that the XWGC and the SECC precisely exclude a model that otherwise appears to sail through the WGC's tests with flying colors.

Throughout this section we have referred to charged particles of charge  $(N, 1)$ . As explained in [48], it may be possible to generate instanton effects through other means, such as Chern-Simons couplings to nonabelian gauge theories that confine. This will not affect our arguments. Some set of charged particles must exist to satisfy the electric WGC, and even if they do not contribute significant instanton effects, their charge assignments still lead to uncontrolled large gauge transformations at the boundaries of moduli space.

We have not attempted to derive a bound in detail for arbitrary charge assignments or arbitrary numbers of axions, but by putting together the ideas of this section with those we employed in section 3, we expect that any model based on compact axion fields with no crucially new physics idea as input can be excluded.

## 6 Assessing strong conjectures

The original weak gravity paper [39] posited a “strong form” of the WGC, which stipulates that in theories with a single  $U(1)$ , the *lightest* charged particle in the spectrum, and not just any charged particle, should have a charge-to-mass ratio larger than that of an

extremal black hole. As pointed out in [48, 65] and later discussed in [49, 51–54], there is a loophole in the electric WGC which leaves open the possibility of axion inflation, but which would be closed by this strong form of the WGC. In particular, consider a theory with a single axion  $a$  and two instantons of action  $S_1 \ll S_2$  and associated decay constants  $f_1, f_2$ , respectively. Each instanton will introduce a term in the axion potential of the form,

$$V(a) \supset \mathcal{A}_i e^{-S_i} \cos a/f_i, \tag{6.1}$$

with  $\mathcal{A}_i$  some coefficients. Now, as long as  $f_2 S_2 < M_{\text{Pl}}$ , the ordinary, mild form of the WGC will be satisfied.  $f_1$  is left unbounded, and since  $S_1 \ll S_2$ , the potential contributions from the first instanton will dominate those from the second. On the other hand, the strong form of the WGC requires that the instanton of smaller action, which is  $S_1$  in our scenario, must also satisfy the bound  $f_1 S_1 < M_{\text{Pl}}$ . Thus, if we assume  $S_1 > 1$ , we find that  $f_1$  is constrained to be sub-Planckian, and the axion is unsuitable for inflation.

However, there are a few problems with invoking the “strong form” of the WGC to close such loopholes. First off, straightforwardly generalizing the strong form to theories with multiple U(1)s proves problematic, as it implies constraints on the spectrum that are clearly far too strong. To see why, consider a very simple theory with two U(1)s and two particles of mass  $m_1, m_2$  with charges  $(q_1, 0)$  and  $(0, q_2)$ . If one considers either of the U(1)’s in this basis, the naïve “strong form of the WGC” holds that lightest particle charged under each U(1) should have  $q/m > 1/M_{\text{Pl}}$ . However, suppose we now make a very small basis rotation of our U(1)s, so that particles 1 and 2 now have charges

$$(q_1 + \mathcal{O}(\varepsilon^2), q_1 \varepsilon + \mathcal{O}(\varepsilon^2)), \quad (-q_2 \varepsilon + \mathcal{O}(\varepsilon^2), q_2 + \mathcal{O}(\varepsilon^2)),$$

respectively. In this new basis, the statement that the lightest particle charged under each U(1) should have sufficiently large charge-to-mass ratio is problematic: if  $m_1 < m_2$ , then by taking  $\varepsilon$  small enough, we can ensure that  $q_1 \varepsilon/m_1$  is too small to satisfy the bound. If  $m_2 < m_1$ , then we can do the same with  $q_2 \varepsilon/m_2$ . The only way this conjecture could hold is if  $m_1 = m_2$  i.e. if every particle in the spectrum has precisely the same mass. This is clearly unacceptable.

This problem may be remedied. Namely, we may define the “strong form” of the WGC for theories with  $N$  U(1)s to be the statement that the lightest particles whose charge-to-mass vectors span the full  $\mathbb{R}^N$  should satisfy the convex hull condition. It is easy to check that in a theory with a single U(1), this definition of the strong WGC reduces to the usual one. Furthermore, the 0-form generalization of this strong form would indeed place strong restrictions on axion moduli space diameters and close the aforementioned loophole.

However, even if we use this updated  $N$ -species strong form of the conjecture, there are other problems with invoking the strong WGC to rule out axion inflation. To begin with, it does not rule out a closely related loophole achieved by taking  $\mathcal{A}_1 < \mathcal{A}_2$  in (6.1). In this case, one could take  $f_2$  arbitrarily large and  $S_1 \lesssim S_2$  and still satisfy the strong WGC. As long as  $\mathcal{A}_1 e^{-S_1} \cos a/f_1$  is sufficiently smaller than  $\mathcal{A}_2 e^{-S_2} \cos a/f_2$ , the potential will be dominated by the latter term. Secondly, it does not close the small action loophole discussed previously, in which the instanton actions are taken smaller than 1. This limit

is difficult to arrange in a controlled string compactification, but as we have seen, it is not such a problem in simpler extranatural scenarios. Most important, however, is the fact that the strong form of the WGC does not derive from arguments based on either effective field theory or black hole thermodynamics. Though further developments could change the situation, we currently see no compelling reason to believe that the WGC holds in its stronger form.<sup>3</sup>

## 7 Conclusions

We have argued that the original magnetic form of the Weak Gravity Conjecture and the UV cutoff that it implies, appropriately generalized to multiple U(1) gauge fields, exclude a variety of  $N$ -flation models including models of kinetic alignment. We summarize the claimed constraints, and the assumptions on which they rely, in table 1. The theories that are excluded in this way have in common the feature that the magnetic charges satisfying the magnetic WGC and the electric charges leading to the dominant instanton effects are simple (not parametrically large and aligned) in the same basis. We believe that these arguments are robust. They can possibly be evaded by considering a theory with a cancellation or tuning in the monopole masses (so that the monopoles are much lighter than the semiclassical self-energy estimate). The only other potential way out is if the compactification radius could somehow be consistently taken to be much smaller than the smallest distance  $\Lambda^{-1}$  for which we trust the monopole solution. Because we expect the description of a local U(1) gauge theory to break down at  $\Lambda$ , this would require going beyond the abelian effective field theory to a more complete ultraviolet description.

A very interesting model in which the electric charge vectors for the instantons are highly aligned in the basis where magnetic charges are simple has been previously claimed to evade the Weak Gravity Conjecture [48]. We agree that it cannot be straightforwardly ruled out by the original WGC. However, this model implies surprising features that arise from the nontrivial dependence of the masses of charged particles on the values of the axion fields. We have proposed additional conjectures that would exclude such surprises. The two assumptions are that the mass spectrum at all extrema in the moduli space should satisfy the Weak Gravity Conjecture and that a light mode that is present in some region of moduli space should be part of the effective theory throughout the entire moduli space (rather than moving above the cutoff). We have no definitive proof of these statements, but they appear to be plausible, and we find it very compelling that they precisely parametrically exclude the one scenario that otherwise evades our arguments. Further study of these conjectures, as well as possible application of them to other models like axion monodromy inflation [16–18, 20, 21] (with non-periodic potentials, unlike all cases considered in this paper), seem to us to be the most likely avenue for progress. We expect that the general argument sketched in section 4.3 can exclude many such models, although they must be considered on a case-by-case basis to see if loopholes exist. Ultraviolet completions of other theories that apply

---

<sup>3</sup>See however [56] for a discussion of the “lattice WGC” (LWGC), a candidate strong form which avoids some of these pitfalls and can be motivated by consistency considerations and string theory examples.

Assumptions \ Reference:	ref. [39]	ref. [48]	ref. [54]	sec. 3.1	sec. 3.2	sec. 5.2	sec. 5.3
Single axion	✓	✓					
$K$ diagonal	✓	✓	WLOG	✓		✓	✓
$S_{\text{inst}} > 1$	✓		✓			✓	✗
Instanton charges obey EIWGC	✓		✓	✓	✓	✓	✓
MagWGC obeyed by simple charges							
Instantons simple, $\sum a_i \cos(\theta_i)$				✓	✓		
Constrained by:	EIWGC	MagWGC	EIWGC	MagWGC	MagWGC	EIXWGC + MagWGC	Single EFT + MagWGC

**Table 1.** WGC constraints on (compact) axion inflation models with various assumptions. Each column is a scenario for which a constraint has been claimed; a ✓ indicates that an assumption is made, and an ✗ indicates that the *opposite* assumption is made. The entry “WLOG” for “without loss of generality” indicates that this assumption was made but, due to the lack of other related assumptions, it is a completely general basis choice. The single-axion assumption implies the  $K$  diagonal assumption. The assumption that “instanton charges obey the electric WGC” misses the loophole where the states that satisfy the electric WGC make negligible contributions to the potential (for instance, they may have mass near the Planck scale, outside the effective theory). In theories with multiple axions the Convex Hull Condition is always assumed to be part of the definition of the appropriate WGC. Abbreviations: EIWGC = Electric Weak Gravity Conjecture, MagWGC = Magnetic Weak Gravity Conjecture, EIXWGC = Electric Extended Weak Gravity Conjecture.

large field ranges to the hierarchy problem [73] or to generating a light dilaton [74–76] may be susceptible to similar constraints.

More generally, noncompact fields (like the scalar moduli which are supersymmetric counterparts of string axions) could give rise to large-field inflation, and the prospects for constraining them with the WGC are not clear. Still, it is thought that noncompact fields in string theory are highly constrained, and that effective field theory always breaks down in the presence of super-Planckian field ranges [38, 40, 41]. Super-Planckian field excursions in *space*, rather than in time, tend to collapse into black holes [77], which may point in the direction of general arguments against the consistency of effective field theories of super-Planckian fields coupled to gravity [78].

The Weak Gravity Conjecture has been established as a powerful tool to cull the space of theories of inflation. The possibility of future measurements of nonzero  $r$  offers the hope that we can confront our understanding of general properties of quantum gravity against real empirical knowledge of our universe. We hope that the study of the WGC can offer tentative steps in the direction of a phenomenology of quantum gravity.

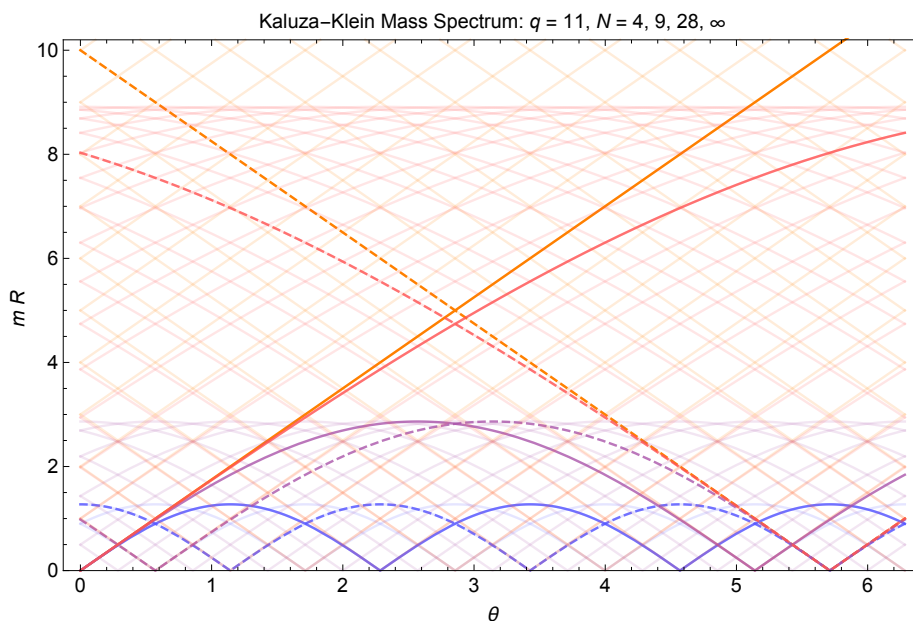
## Acknowledgments

We thank Thomas Bachlechner, Cliff Cheung, Thomas Dumitrescu, Peng Gao, Cody Long, Liam McAllister, Grant Remmen, Prashant Saraswat, Matt Schwartz, Gary Shiu, Eva Silverstein, Raman Sundrum, and Cumrun Vafa for discussions or correspondence. MR thanks Thomas Bachlechner, Liam McAllister, and Alexander Westphal for useful discussions after the first version of the paper, and in particular for suggesting the addition of table 1. BH is supported by the Fundamental Laws Initiative of the Harvard Center for the Fundamental Laws of Nature. The work of MR is supported in part by the NSF Grant PHY-1415548. TR is supported in part by the National Science Foundation under Grant No. DGE-1144152. While preparing the revised version of this paper, MR was supported in part by the National Science Foundation under Grant No. PHYS-1066293 and the hospitality of the Aspen Center for Physics.

## A The SECC and gauge-invariant regulators

Our argument about the SECC relies, in part, on a discussion of short-wavelength modes of charged fields. In a gauge theory we expect that arbitrarily high-frequency gauge transformations should not be considered: gauge theories can be emergent long-distance descriptions of other physics, and the short distance degrees of freedom may be entirely different. However, because  $\partial_\mu$  is not gauge invariant, while  $D_\mu$  is, one may be concerned that the problem we are discussing is not actually a physical one. We believe that tracking physical (gauge invariant) masses of individual Kaluza-Klein modes adiabatically as the physical Wilson loop variable  $\theta$  varies and seeing that they pass through the cutoff already demonstrates that the problem is physical. For completeness, in this appendix we provide a different perspective by looking at how two different gauge-invariant regulators treat the spectrum of modes across the moduli space.





**Figure 3.** Kaluza-Klein mode spectrum for a particle of charge  $q = 11$  with an  $N$ -site lattice regulator for multiple choices of  $N$ . Blue curves are  $N = 4$ ; purple,  $N = 9$ ; red,  $N = 28$ ; and orange, the continuum result  $N \rightarrow \infty$ . Observe that the spectrum of low-lying modes at  $mR \lesssim 1$  is the same for any number of lattice sites. However, the behavior of an individual mode tracked as  $\theta$  is continuously varied is dramatically different at small  $N$  and large  $N$ . The solid lines are modes with zero mass at the origin of moduli space, and the dashed lines are modes with zero mass at  $\frac{\theta}{2\pi} = \frac{10}{11}$ . Faint curves in the background show the other modes.

Our first regulator is the lattice, with spacing  $a = \Lambda^{-1}$ . This regulator gives a good example in which 5d locality breaks down at short distances, but gauge invariance is preserved and we have a convenient setting to discuss only physical quantities. We will only discretize the fifth dimension, for simplicity. In other words, we will apply the idea of dimensional deconstruction [79–82]. We replace the continuous fifth dimension with a periodic lattice with  $N$  sites. The SECC criterion of section 4.2 tells us that, in order to work within the context of a single effective field theory valid everywhere in moduli space, we should take  $N \gg q$ . Let us explore what happens at smaller  $N$  and how the spectrum alters as we take  $N$  large. For a scalar field, we find a finite set of  $N$  modes with masses

$$m_n^2 = \frac{N^2}{\pi^2 R^2} \sin^2 \left( \frac{n\pi - q\theta/2}{N} \right), \tag{A.1}$$

with integer  $n$  in the range  $-N/2 < n \leq N/2$ . In the large  $N$  limit, for  $n \ll N$ , this reduces to the continuum Kaluza-Klein mode result  $m_n^2 = (n - \frac{q\theta}{2\pi})^2/R^2$ . Notice that the heaviest modes will have mass  $\approx N/(\pi R)$ , so to recover the continuum result for mode numbers up to  $q$  we will need to take  $N$  large compared to  $\pi q$ .

From the expression (A.1) we see that there is a zero mode whenever  $\theta = 2\pi n/q$  (for integer  $n$ ). This means that the lattice regulator brings down a light mode whenever the 5d continuum theory with a high cutoff would tell us to expect one, *independent of how*

small we take  $N$ . One could take this as a hint that gauge invariance is subtle and any UV regulator will guarantee the consistency of the effective field theory. On the other hand, the origin of the light mode varies dramatically as we change  $N$ . In the continuum theory with a cutoff satisfying the SECC, the zero mode at  $\theta = 2\pi n/q$  is different for each  $n$ . In the lattice theory at small  $N$ , there simply aren't enough modes for this to be true: a mode can be a zero mode, increase in mass as  $\theta$  is varied, then turn around and decrease again to become a new zero mode at a different value of  $\theta$ . We illustrate this in figure 3. If the cutoff is sufficiently high, i.e.  $N \gg q$ , then a mode that becomes light at order-one values of  $\theta$  was heavy at small  $\theta$ . When  $N \ll q$ , the modes that become light at any point in the moduli space were already light at the origin.

On the other hand, we could consider a different gauge invariant regulator. Suppose that we use a Pauli-Villars regulator, adding a mode with wrong-sign kinetic term and mass  $\Lambda$  in the bulk. When  $n - \frac{q\theta}{2\pi} \gg \Lambda R$ , the contribution of a mode and its Pauli-Villars partner nearly cancel. In this case, in some sense all the modes are present for all  $n$ , but the regulator suppresses their contributions if they are not low-lying states. If we adiabatically track a mode that becomes light at large  $\theta$ , it will have been heavy at the origin, and vice versa. This is very different from the case of a lattice regulator, where individual modes periodically become light multiple times.

What we have found is that, for different gauge-invariant regulators with the same UV cutoff, the spectrum of light states will be similar but the behavior of individual modes as we move across the moduli space can be radically different. At the risk of belaboring the point, consider a physical state containing particles that are light at  $\theta = 0$ , and adiabatically vary  $\theta$  to order-one values. The Pauli-Villars regulator would tell us that our particles keep becoming heavier. The lattice regulator would suggest that they would again become light. This is a sign of UV sensitivity and a breakdown of effective field theory. Only by taking a UV cutoff large enough to encompass at least  $q$  modes at one time can we obtain a consistent answer that is insensitive to the regulator. Cutoffs that do not satisfy  $q\theta \lesssim \Lambda R$  require additional ultraviolet knowledge beyond the EFT. An interesting question to pursue further is just how mild this additional ultraviolet knowledge is: for instance, the lattice regulator discussed here is incompatible with knowing that the ultraviolet completion obeys Lorentz invariance in the extra dimension.

## B Do entropy bounds exclude large-field inflation?

One constraint widely believed to apply in any theory of quantum gravity is an entropy bound, loosely speaking that the logarithm of the number of microstates accessible to a system bounded by a surface of area  $A$  is at most of order  $AM_{\text{Pl}}^2$ . Such conjectures originated with Bekenstein [83] and were given a sharp covariant form by Bousso [84]. Entanglement entropy in quantum field theory [85, 86] has been shown to satisfy such a bound [87–89], suggesting that it will be difficult to place theories in the swampland simply by arguing that they violate an entropy bound. On the other hand, several authors have argued for precisely such statements regarding large-field inflation models [66–68], including recently in the context of the Weak Gravity Conjecture [51]. Our goal in this section is to

critically review these arguments. We find that they rely on unjustified assumptions. Our assessment is that models of large-field inflation are consistent with entropy bounds.

### B.1 The reheating objection

An argument against theories with a large number of  $e$ -folds has been given based on entropy production during reheating [51, 66]. Essentially, the objection is that to the extent that we can trust the semiclassical picture of the post-reheating universe as a radiation-dominated phase, it will contain a hot plasma with entropy per unit volume  $s \propto T^3$ . Because the entropy scales with volume and the entropy bound scales with area, it seems there is a potential conflict. But if we consider a radiation-dominated universe,

$$\rho \sim H^2 M_{\text{Pl}}^2 \sim T^4, \tag{B.1}$$

so in a volume of radius  $R \sim H^{-1}$ , the entropy associated with the radiation is

$$S \sim T^3 R^3 \sim \frac{1}{T} H^2 M_{\text{Pl}}^2 H^{-3} \sim \frac{M_{\text{Pl}}^2 H}{H^2 T} \sim S_{\text{Bek}} \frac{T}{M_{\text{Pl}}}. \tag{B.2}$$

Thus, the Bekenstein bound is safely satisfied in a Hubble-size volume provided the temperature is much less than the Planck energy, which is surely true whenever effective field theory is valid.

The potential to derive a contradiction between inflation with many  $e$ -folds and entropy bounds arises from considering volumes of radius  $R \gg H^{-1}$ . [66] framed the problem in terms of the particle horizon, which grows exponentially during inflation, so the ratio of volume to area becomes much larger than for a Hubble-sized volume. The uncertainty about what region to apply entropy bounds to was a major motivating reason for Bousso's covariant entropy conjecture [84], which had a significant impact because it gave a precise criterion for which such problems do not arise. Given that substantial evidence has accumulated that Bousso's variation of the entropy bound is the correct criterion, it does not appear that reheating after inflation is problematic even in theories with many  $e$ -folds.

### B.2 Bit-counting arguments

Another argument that axions with super-Planckian decay constants can violate entropy bounds arises from a bit-counting argument. We impose an ultraviolet cutoff on the theory by defining a minimum distance scale  $\ell$  and count pixels of area  $\ell^2$  on the horizon. There are  $\sim 1/(\ell^2 H^2)$  such pixels. This exceeds the bound if  $\ell < M_{\text{Pl}}^{-1}$ . If we keep the UV cutoff of our theory below the Planck scale, there should be no problem.

The UV cutoff in refs. [67, 68] was taken to be  $\ell \sim f_a^{-1}$ . The argument was that the free-particle two-point function of our axion field,  $\langle a(x)a(0) \rangle \sim \frac{1}{x^2}$ , can no longer be a valid estimate at  $x \lesssim f_a^{-1}$ , because it grows without bound even though the field itself is compact and satisfies  $a(x) < 2\pi f_a$ . This is a reasonable argument. We could refine it a bit, by noting that the two-point function is not physical because  $a(x)$  is not gauge invariant, but  $e^{ia(x)/f_a}$  is, so we can build a Lagrangian out of the latter and expand to find non-renormalizable interactions suppressed by the scale  $f_a$ , which fixes the cutoff. Although

the argument that compactness of the field space implies a UV cutoff is correct, it only imposes an *inequality*,  $\ell > f_a^{-1}$ . A contradiction with entropy bounds arises only if we take this to be an equation. But of course, in any theory of quantum gravity we expect it will not be sensible to talk about distances below the Planck length, so we should never take  $\ell < M_{\text{Pl}}^{-1}$ . Nothing in this argument precludes the possibility of theories with a range of scales  $f_a \gg M_{\text{Pl}} \gg \ell^{-1}$ .

### B.3 The classical entropy current argument

A further argument considered the inflaton as a perfect fluid and constructed the entropy density from its stress energy tensor [68]. This leads to an equation

$$\dot{S}_\phi = \frac{8\pi^2}{H^3} \dot{\phi}^2. \tag{B.3}$$

[68] then gives an argument with the following logical structure:  $\Delta S_\phi$  is computed by integrating the above derivative. The expression for  $\Delta S_\phi$  is broken into two pieces and bounded by considering the absolute value of these two pieces. This establishes that  $\Delta S_\phi < S_{\text{max}} \sim f_a^2/H^2$ . It is then observed that  $S_{\text{max}} > S_{\text{Bek}}$  for large-field inflation. But, without some estimate of *how close*  $\Delta S_\phi$  actually comes to  $S_{\text{max}}$ , this proves nothing.

In fact, we can obtain a much better estimate of  $\Delta S_\phi$ . For slow-roll inflation we have  $\dot{\phi} \approx -\frac{1}{3H} \frac{\partial V}{\partial \phi}$ , so we can rewrite the integral as:

$$\Delta S_\phi = \int dt \frac{8\pi^2}{H^3} \dot{\phi} \left( -\frac{1}{3H} \frac{\partial V}{\partial \phi} \right) \approx -\frac{8\pi^2}{3H^4} \int d\phi \frac{\partial V}{\partial \phi} \approx \frac{8\pi^2}{3H^4} |\Delta V|, \tag{B.4}$$

with the approximation that  $H$  is relatively constant during the time period considered. Given that during inflation  $V \approx 3H^2 M_{\text{Pl}}^2$ , this is exactly consistent with the Gibbons-Hawking entropy of a Hubble patch of de Sitter space [90], so no bound is violated. In fact, a detailed study of de Sitter thermodynamics in the context of inflation, including perturbations, was undertaken some time ago by Frolov and Kofman, who found no inconsistencies [91].

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

### References

- [1] D.H. Lyth, *What would we learn by detecting a gravitational wave signal in the cosmic microwave background anisotropy?*, *Phys. Rev. Lett.* **78** (1997) 1861 [[hep-ph/9606387](https://arxiv.org/abs/hep-ph/9606387)] [[INSPIRE](https://arxiv.org/abs/hep-ph/9606387)].
- [2] R. Easther, W.H. Kinney and B.A. Powell, *The Lyth bound and the end of inflation*, *JCAP* **08** (2006) 004 [[astro-ph/0601276](https://arxiv.org/abs/astro-ph/0601276)] [[INSPIRE](https://arxiv.org/abs/astro-ph/0601276)].
- [3] D. Baumann and D. Green, *A field range bound for general single-field inflation*, *JCAP* **05** (2012) 017 [[arXiv:1111.3040](https://arxiv.org/abs/1111.3040)] [[INSPIRE](https://arxiv.org/abs/1111.3040)].

- [4] S. Antusch and D. Nolde, *BICEP2 implications for single-field slow-roll inflation revisited*, *JCAP* **05** (2014) 035 [[arXiv:1404.1821](#)] [[INSPIRE](#)].
- [5] J. Bramante, S. Downes, L. Lehman and A. Martin, *Last stand of single small field inflation*, *Phys. Rev. D* **90** (2014) 023530 [[arXiv:1405.7563](#)] [[INSPIRE](#)].
- [6] M. Kamionkowski and J. March-Russell, *Planck scale physics and the Peccei-Quinn mechanism*, *Phys. Lett. B* **282** (1992) 137 [[hep-th/9202003](#)] [[INSPIRE](#)].
- [7] R. Holman, S.D.H. Hsu, T.W. Kephart, E.W. Kolb, R. Watkins and L.M. Widrow, *Solutions to the strong CP problem in a world with gravity*, *Phys. Lett. B* **282** (1992) 132 [[hep-ph/9203206](#)] [[INSPIRE](#)].
- [8] R. Kallosh, A.D. Linde, D.A. Linde and L. Susskind, *Gravity and global symmetries*, *Phys. Rev. D* **52** (1995) 912 [[hep-th/9502069](#)] [[INSPIRE](#)].
- [9] J.P. Conlon, *The QCD axion and moduli stabilisation*, *JHEP* **05** (2006) 078 [[hep-th/0602233](#)] [[INSPIRE](#)].
- [10] P. Svrček and E. Witten, *Axions in string theory*, *JHEP* **06** (2006) 051 [[hep-th/0605206](#)] [[INSPIRE](#)].
- [11] K. Freese, J.A. Frieman and A.V. Olinto, *Natural inflation with pseudo-Nambu-Goldstone bosons*, *Phys. Rev. Lett.* **65** (1990) 3233 [[INSPIRE](#)].
- [12] T. Banks, M. Berkooz, S.H. Shenker, G.W. Moore and P.J. Steinhardt, *Modular cosmology*, *Phys. Rev. D* **52** (1995) 3548 [[hep-th/9503114](#)] [[INSPIRE](#)].
- [13] N. Arkani-Hamed, H.-C. Cheng, P. Creminelli and L. Randall, *Extra natural inflation*, *Phys. Rev. Lett.* **90** (2003) 221302 [[hep-th/0301218](#)] [[INSPIRE](#)].
- [14] J.E. Kim, H.P. Nilles and M. Peloso, *Completing natural inflation*, *JCAP* **01** (2005) 005 [[hep-ph/0409138](#)] [[INSPIRE](#)].
- [15] S. Dimopoulos, S. Kachru, J. McGreevy and J.G. Wacker, *N-flation*, *JCAP* **08** (2008) 003 [[hep-th/0507205](#)] [[INSPIRE](#)].
- [16] E. Silverstein and A. Westphal, *Monodromy in the CMB: gravity waves and string inflation*, *Phys. Rev. D* **78** (2008) 106003 [[arXiv:0803.3085](#)] [[INSPIRE](#)].
- [17] L. McAllister, E. Silverstein and A. Westphal, *Gravity waves and linear inflation from axion monodromy*, *Phys. Rev. D* **82** (2010) 046003 [[arXiv:0808.0706](#)] [[INSPIRE](#)].
- [18] M. Berg, E. Pajer and S. Sjörs, *Dante's inferno*, *Phys. Rev. D* **81** (2010) 103535 [[arXiv:0912.1341](#)] [[INSPIRE](#)].
- [19] C. Germani and A. Kehagias, *UV-protected inflation*, *Phys. Rev. Lett.* **106** (2011) 161302 [[arXiv:1012.0853](#)] [[INSPIRE](#)].
- [20] N. Kaloper, A. Lawrence and L. Sorbo, *An ignoble approach to large field inflation*, *JCAP* **03** (2011) 023 [[arXiv:1101.0026](#)] [[INSPIRE](#)].
- [21] N. Kaloper and A. Lawrence, *Natural chaotic inflation and ultraviolet sensitivity*, *Phys. Rev. D* **90** (2014) 023506 [[arXiv:1404.2912](#)] [[INSPIRE](#)].
- [22] C.P. Burgess, M. Cicoli, F. Quevedo and M. Williams, *Inflating with large effective fields*, *JCAP* **11** (2014) 045 [[arXiv:1404.6236](#)] [[INSPIRE](#)].
- [23] T.C. Bachlechner, M. Dias, J. Frazer and L. McAllister, *Chaotic inflation with kinetic alignment of axion fields*, *Phys. Rev. D* **91** (2015) 023520 [[arXiv:1404.7496](#)] [[INSPIRE](#)].

- [24] C. Csáki, N. Kaloper, J. Serra and J. Terning, *Inflation from broken scale invariance*, *Phys. Rev. Lett.* **113** (2014) 161302 [[arXiv:1406.5192](#)] [[INSPIRE](#)].
- [25] K. Furuuchi and Y. Koyama, *Large field inflation models from higher-dimensional gauge theories*, *JCAP* **02** (2015) 031 [[arXiv:1407.1951](#)] [[INSPIRE](#)].
- [26] K. Harigaya and M. Ibe, *Phase locked inflation — Effectively trans-planckian natural inflation*, *JHEP* **11** (2014) 147 [[arXiv:1407.4893](#)] [[INSPIRE](#)].
- [27] T. Higaki and F. Takahashi, *Axion landscape and natural inflation*, *Phys. Lett. B* **744** (2015) 153 [[arXiv:1409.8409](#)] [[INSPIRE](#)].
- [28] G. Shiu, W. Staessens and F. Ye, *Large field inflation from axion mixing*, *JHEP* **06** (2015) 026 [[arXiv:1503.02965](#)] [[INSPIRE](#)].
- [29] A.R. Liddle, A. Mazumdar and F.E. Schunck, *Assisted inflation*, *Phys. Rev. D* **58** (1998) 061301 [[astro-ph/9804177](#)] [[INSPIRE](#)].
- [30] E.J. Copeland, A. Mazumdar and N.J. Nunes, *Generalized assisted inflation*, *Phys. Rev. D* **60** (1999) 083506 [[astro-ph/9904309](#)] [[INSPIRE](#)].
- [31] A. Mazumdar, S. Panda and A. Perez-Lorenzana, *Assisted inflation via tachyon condensation*, *Nucl. Phys. B* **614** (2001) 101 [[hep-ph/0107058](#)] [[INSPIRE](#)].
- [32] A. Jokinen and A. Mazumdar, *Inflation in large- $N$  limit of supersymmetric gauge theories*, *Phys. Lett. B* **597** (2004) 222 [[hep-th/0406074](#)] [[INSPIRE](#)].
- [33] A. Ashoorioon, H. Firouzjahi and M.M. Sheikh-Jabbari, *M-flation: inflation from matrix valued scalar fields*, *JCAP* **06** (2009) 018 [[arXiv:0903.1481](#)] [[INSPIRE](#)].
- [34] A. Ashoorioon and M.M. Sheikh-Jabbari, *Gauged M-flation, its UV sensitivity and spectator species*, *JCAP* **06** (2011) 014 [[arXiv:1101.0048](#)] [[INSPIRE](#)].
- [35] A. Ashoorioon and M.M. Sheikh-Jabbari, *Gauged M-flation after BICEP2*, *Phys. Lett. B* **739** (2014) 391 [[arXiv:1405.1685](#)] [[INSPIRE](#)].
- [36] T. Banks, M. Dine, P.J. Fox and E. Gorbatov, *On the possibility of large axion decay constants*, *JCAP* **06** (2003) 001 [[hep-th/0303252](#)] [[INSPIRE](#)].
- [37] T. Rudelius, *On the possibility of large axion moduli spaces*, *JCAP* **04** (2015) 049 [[arXiv:1409.5793](#)] [[INSPIRE](#)].
- [38] C. Vafa, *The string landscape and the swampland*, [hep-th/0509212](#) [[INSPIRE](#)].
- [39] N. Arkani-Hamed, L. Motl, A. Nicolis and C. Vafa, *The string landscape, black holes and gravity as the weakest force*, *JHEP* **06** (2007) 060 [[hep-th/0601001](#)] [[INSPIRE](#)].
- [40] H. Ooguri and C. Vafa, *On the geometry of the string landscape and the swampland*, *Nucl. Phys. B* **766** (2007) 21 [[hep-th/0605264](#)] [[INSPIRE](#)].
- [41] M.R. Douglas and Z. Lu, *Finiteness of volume of moduli spaces*, [hep-th/0509224](#) [[INSPIRE](#)].
- [42] Y. Kats, L. Motl and M. Padi, *Higher-order corrections to mass-charge relation of extremal black holes*, *JHEP* **12** (2007) 068 [[hep-th/0606100](#)] [[INSPIRE](#)].
- [43] T. Banks, M. Johnson and A. Shomer, *A note on gauge theories coupled to gravity*, *JHEP* **09** (2006) 049 [[hep-th/0606277](#)] [[INSPIRE](#)].
- [44] C. Cheung and G.N. Remmen, *Naturalness and the weak gravity conjecture*, *Phys. Rev. Lett.* **113** (2014) 051601 [[arXiv:1402.2287](#)] [[INSPIRE](#)].

- [45] C. Cheung and G.N. Remmen, *Infrared consistency and the weak gravity conjecture*, *JHEP* **12** (2014) 087 [[arXiv:1407.7865](#)] [[INSPIRE](#)].
- [46] L. Susskind, *Trouble for remnants*, [hep-th/9501106](#) [[INSPIRE](#)].
- [47] T.C. Bachlechner, C. Long and L. McAllister, *Planckian axions in string theory*, [arXiv:1412.1093](#) [[INSPIRE](#)].
- [48] A. de la Fuente, P. Saraswat and R. Sundrum, *Natural inflation and quantum gravity*, *Phys. Rev. Lett.* **114** (2015) 151303 [[arXiv:1412.3457](#)] [[INSPIRE](#)].
- [49] T. Rudelius, *Constraints on axion inflation from the weak gravity conjecture*, *JCAP* **09** (2015) 020 [[arXiv:1503.00795](#)] [[INSPIRE](#)].
- [50] M. Montero, A.M. Uranga and I. Valenzuela, *Transplanckian axions!?*, *JHEP* **08** (2015) 032 [[arXiv:1503.03886](#)] [[INSPIRE](#)].
- [51] J. Brown, W. Cottrell, G. Shiu and P. Soler, *Fencing in the swampland: quantum gravity constraints on large field inflation*, *JHEP* **10** (2015) 023 [[arXiv:1503.04783](#)] [[INSPIRE](#)].
- [52] T.C. Bachlechner, C. Long and L. McAllister, *Planckian axions and the weak gravity conjecture*, [arXiv:1503.07853](#) [[INSPIRE](#)].
- [53] A. Hebecker, P. Mangat, F. Rompineve and L.T. Witkowski, *Winding out of the swamp: evading the weak gravity conjecture with F-term winding inflation?*, *Phys. Lett. B* **748** (2015) 455 [[arXiv:1503.07912](#)] [[INSPIRE](#)].
- [54] J. Brown, W. Cottrell, G. Shiu and P. Soler, *On axionic field ranges, loopholes and the weak gravity conjecture*, [arXiv:1504.00659](#) [[INSPIRE](#)].
- [55] D. Junghans, *Large-field inflation with multiple axions and the weak gravity conjecture*, [arXiv:1504.03566](#) [[INSPIRE](#)].
- [56] B. Heidenreich, M. Reece and T. Rudelius, *Sharpening the weak gravity conjecture with dimensional reduction*, [arXiv:1509.06374](#) [[INSPIRE](#)].
- [57] W.A. Hiscock and L.D. Weems, *Evolution of charged evaporating black holes*, *Phys. Rev. D* **41** (1990) 1142 [[INSPIRE](#)].
- [58] G. Gibbons, *Vacuum polarization and the spontaneous loss of charge by black holes*, *Commun. Math. Phys.* **44** (1975) 245.
- [59] B.W. Schumacher, *Is Bekenstein's conjecture true for charged black holes?*, *Phys. Rev. Lett.* **54** (1985) 2643 [[INSPIRE](#)].
- [60] I.B. Khriplovich, *Particle creation by charged black holes*, *Phys. Rept.* **320** (1999) 37 [[INSPIRE](#)].
- [61] F. Denef and S.A. Hartnoll, *Landscape of superconducting membranes*, *Phys. Rev. D* **79** (2009) 126008 [[arXiv:0901.1160](#)] [[INSPIRE](#)].
- [62] C.-M. Chen, S.P. Kim, I.-C. Lin, J.-R. Sun and M.-F. Wu, *Spontaneous pair production in Reissner-Nordstrom black holes*, *Phys. Rev. D* **85** (2012) 124041 [[arXiv:1202.3224](#)] [[INSPIRE](#)].
- [63] C.-M. Chen, J.-R. Sun, F.-Y. Tang and P.-Y. Tsai, *Spinor particle creation in near extremal Reissner-Nordström black holes*, *Class. Quant. Grav.* **32** (2015) 195003 [[arXiv:1412.6876](#)] [[INSPIRE](#)].
- [64] T.C. Bachlechner, C. Long, L. McAllister and J. Stout, to appear.

- [65] A. de la Fuente, P. Saraswat and R. Sundrum, private communication (2014).
- [66] N. Kaloper and A.D. Linde, *Cosmology versus holography*, *Phys. Rev. D* **60** (1999) 103509 [[hep-th/9904120](#)] [[INSPIRE](#)].
- [67] J.P. Conlon, *Quantum gravity constraints on inflation*, *JCAP* **09** (2012) 019 [[arXiv:1203.5476](#)] [[INSPIRE](#)].
- [68] L. Boubekeur, *On the scale of new physics in inflation*, [arXiv:1312.4768](#) [[INSPIRE](#)].
- [69] Y. Hosotani, *Dynamical mass generation by compact extra dimensions*, *Phys. Lett. B* **126** (1983) 309 [[INSPIRE](#)].
- [70] H.-C. Cheng, K.T. Matchev and M. Schmaltz, *Radiative corrections to Kaluza-Klein masses*, *Phys. Rev. D* **66** (2002) 036005 [[hep-ph/0204342](#)] [[INSPIRE](#)].
- [71] N. Arkani-Hamed, S. Dubovsky, A. Nicolis and G. Villadoro, *Quantum horizons of the standard model landscape*, *JHEP* **06** (2007) 078 [[hep-th/0703067](#)] [[INSPIRE](#)].
- [72] N. Seiberg and E. Witten, *Electric-magnetic duality, monopole condensation and confinement in  $N = 2$  supersymmetric Yang-Mills theory*, *Nucl. Phys. B* **426** (1994) 19 [*Erratum ibid.* **B 430** (1994) 485] [[hep-th/9407087](#)] [[INSPIRE](#)].
- [73] P.W. Graham, D.E. Kaplan and S. Rajendran, *Cosmological relaxation of the electroweak scale*, *Phys. Rev. Lett.* **115** (2015) 221801 [[arXiv:1504.07551](#)] [[INSPIRE](#)].
- [74] R. Contino, A. Pomarol and R. Rattazzi, unpublished work discussed in talks at Xmas10 and Planck 2010 (2010)
- [75] B. Bellazzini, C. Csáki, J. Hubisz, J. Serra and J. Terning, *A naturally light dilaton and a small cosmological constant*, *Eur. Phys. J. C* **74** (2014) 2790 [[arXiv:1305.3919](#)] [[INSPIRE](#)].
- [76] F. Coradeschi, P. Lodone, D. Pappadopulo, R. Rattazzi and L. Vitale, *A naturally light dilaton*, *JHEP* **11** (2013) 057 [[arXiv:1306.4601](#)] [[INSPIRE](#)].
- [77] A. Nicolis, *On super-planckian fields at sub-planckian energies*, *JHEP* **07** (2008) 023 [[arXiv:0802.3923](#)] [[INSPIRE](#)].
- [78] Z. Komargodski, private communication (2014).
- [79] N. Arkani-Hamed, A.G. Cohen and H. Georgi, *(De)constructing dimensions*, *Phys. Rev. Lett.* **86** (2001) 4757 [[hep-th/0104005](#)] [[INSPIRE](#)].
- [80] C.T. Hill, S. Pokorski and J. Wang, *Gauge invariant effective Lagrangian for Kaluza-Klein modes*, *Phys. Rev. D* **64** (2001) 105005 [[hep-th/0104035](#)] [[INSPIRE](#)].
- [81] N. Arkani-Hamed, A.G. Cohen and H. Georgi, *Electroweak symmetry breaking from dimensional deconstruction*, *Phys. Lett. B* **513** (2001) 232 [[hep-ph/0105239](#)] [[INSPIRE](#)].
- [82] C.T. Hill and A.K. Leibovich, *Deconstructing 5D QED*, *Phys. Rev. D* **66** (2002) 016006 [[hep-ph/0205057](#)] [[INSPIRE](#)].
- [83] J.D. Bekenstein, *A universal upper bound on the entropy to energy ratio for bounded systems*, *Phys. Rev. D* **23** (1981) 287 [[INSPIRE](#)].
- [84] R. Bousso, *A covariant entropy conjecture*, *JHEP* **07** (1999) 004 [[hep-th/9905177](#)] [[INSPIRE](#)].
- [85] M. Srednicki, *Entropy and area*, *Phys. Rev. Lett.* **71** (1993) 666 [[hep-th/9303048](#)] [[INSPIRE](#)].



- [86] H. Casini and M. Huerta, *Entanglement entropy in free quantum field theory*, *J. Phys. A* **42** (2009) 504007 [[arXiv:0905.2562](#)] [[INSPIRE](#)].
- [87] H. Casini, *Relative entropy and the Bekenstein bound*, *Class. Quant. Grav.* **25** (2008) 205021 [[arXiv:0804.2182](#)] [[INSPIRE](#)].
- [88] R. Bousso, H. Casini, Z. Fisher and J. Maldacena, *Proof of a quantum bousso bound*, *Phys. Rev. D* **90** (2014) 044002 [[arXiv:1404.5635](#)] [[INSPIRE](#)].
- [89] R. Bousso, H. Casini, Z. Fisher and J. Maldacena, *Entropy on a null surface for interacting quantum field theories and the Bousso bound*, *Phys. Rev. D* **91** (2015) 084030 [[arXiv:1406.4545](#)] [[INSPIRE](#)].
- [90] G.W. Gibbons and S.W. Hawking, *Cosmological event horizons, thermodynamics and particle creation*, *Phys. Rev. D* **15** (1977) 2738 [[INSPIRE](#)].
- [91] A.V. Frolov and L. Kofman, *Inflation and de Sitter thermodynamics*, *JCAP* **05** (2003) 009 [[hep-th/0212327](#)] [[INSPIRE](#)].