The Weak Gravity Conjecture and Axion Inflation

Anthony M. Charles

Department of Physics and Leinweber Center for Theoretical Physics
University of Michigan, 450 Church Street, Ann Arbor, MI 48109-1020, USA

Abstract

The inflationary approach is currently the most successful way to explain early-universe cosmology. However, there exist a very large number of candidate inflation models, many of which can be tuned to give a wide range of experimental predictions. Ruling out these models experimentally is hard, and so we need good ways to rule out inflationary models with theoretical constraints. In this note, we review large-field axion models and discuss their merits as inflationary candidates. We then discuss the weak gravity conjecture and show how its extension for axion fields is inconsistent with these large-field axion inflation models. Based on the author's journal club talk on January 16, 2018.

Contents

1 A Crash-Course on Inflation

In this section we review inflation and set up the conditions required for slow-roll inflation. No prior experience with cosmology is assumed, I promise. Feel free to skip this if you already feel comfortable with slow-roll inflation.

1.1 The FRW Metric and the Friedmann Equations

Experimental evidence from cosmological experiments tells us tjhat our universe seems to have three general large-scale features:

1. The universe is expanding over time.
2. The universe is (approximately) homogeneous.
3. The universe is (approximately) isotropic.

These three properties lead us to describe the our spacetime on large scales by the Friedmann-Robertson-Walker (FRW) metric

$$ds^2 = -dt^2 + a^2(t)\left(\frac{dr^2}{1-kr^2} + r^2 d\Omega_2^2\right) \ , \tag{1.1}$$

where $a(t)$ is a (dimensionless) function of time known as the scale factor, and k is a constant that depends on the curvature of the universe on a given time slice:

$$k = \begin{cases} +1 & \text{positive curvature} \\ 0 & \text{no curvature (flat)} \\ -1 & \text{negative curvature} \end{cases} . \tag{1.2}$$

The scale factor $a(t)$ characterizes the relative size of spacelike surfaces Σ at different times. For example, the distance an observer at the origin will measure to an object at coordinates (r, θ, ϕ) is $R = a(t)r$.

An important parameter when considering how the size of the universe changes over time is the Hubble parameter H, defined by

$$H = \frac{\dot{a}}{a} , \tag{1.3}$$

where dots denote derivatives with respect to time. $H > 0$ corresponds to a universe that is expanding over time, while $H < 0$ corresponds to a contracting universe.

The Einstein equation (in units where $c = G = 1$) tells us that

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi T_{\mu\nu} , \tag{1.4}$$

where $T_{\mu\nu}$ is the stress-energy tensor of the matter in the universe. If we assume that this matter behaves like an ideal fluid, then the stress-energy tensor is given by the diagonal matrix

$$T^{\mu}{}_{\nu} = \mathrm{diag}(\rho, -p, -p, -p) , \tag{1.5}$$

where ρ is the density and p the pressure of the fluid. If we use this stress-energy tensor in the Einstein equation, we find two distinct equations known as the Friedmann equations:

$$\begin{aligned} H^2 &= \frac{\rho}{3} - \frac{k}{a^2} , \\ \dot{H} + H^2 &= -\frac{1}{6}(\rho + 3p) . \end{aligned} \tag{1.6}$$

A useful relation that comes from combining these two equations with the time-derivative of the first equation. The result is known as the continuity equation:

$$\dot{\rho} + 3H(\rho + p) = 0 . \tag{1.7}$$

This relation is linear in time derivatives and can be integrated and solved in a very straightforward way. If we define the equation of state parameter w by

$$w = \frac{p}{\rho} , \tag{1.8}$$

then integrating the continuity equation tells us that $\rho \sim a^{-3(1+w)}$. We can plug this back into the first Friedmann equation to find that the scale factor depends on time as

$$a(t) \propto \begin{cases} t^{\frac{2}{3}(1+w)} & w \neq -1 , \\ e^{Ht} & w = -1 . \end{cases} \tag{1.9}$$

The expansion of the universe is therefore governed by the state parameter w. Some important values of w are $w = 0$ for non-relativistic matter and $w = \frac{1}{3}$ for radiation.

1.2 The Flatness Problem

In the work we've done so far, the curvature of the universe has come in only through the parameter k which takes only discrete values. It therefore does not serve as a very good parameter to measure just how flat the universe is. To quantify a notion of flatness, let's recall first Friedmann equation, which says that

$$H^2 = \frac{\rho(a)}{3} - \frac{k}{a^2} \ . \tag{1.10}$$

Note that we have implicitly labelled the density to be a function of a instead of t, which is allowed by the monotonic growth of the scale factor with time. For a completely flat universe $k = 0$, and so this tells us that

$$\text{flat universe} \implies \rho = 3H^2 \ . \tag{1.11}$$

We will therefore define the critical density ρ_{crit} to be the density for a flat universe, namely

$$\rho_{\text{crit}} = 3H^2 \ . \tag{1.12}$$

For a universe with curvature, we can now say that the close ρ is to ρ_{crit}, the flatter the universe is, with $\rho > \rho_{\text{crit}}$ corresponding to positive curvature and $\rho < \rho_{\text{crit}}$ corresponding to negative curvature. This quantification is typically done by introducing the curvature parameter Ω, defined to be

$$\Omega = \frac{\rho}{\rho_{\text{crit}}} \ . \tag{1.13}$$

Values of $\Omega > 1$ correspond to positive curvature, while $\Omega < 1$ indicates negative curvature. If we differentiate this expression with respect to the scale factor and using the continuity equation (1.7), we find that the curvature of the universe evolves with the scale factor as

$$\frac{d\Omega}{d\log a} = (1+3w)\Omega(\Omega - 1) \ . \tag{1.14}$$

Note that experimental evidence tells us that the scale factor $a(t)$ increases monotonically over time in our cosmological history, and so we can treat the scale factor as a time coordinate.

Now, what is the flatness relation (1.14) telling us? First off, it says that $\Omega = 1$ (e.g. perfect flatness) is a fixed point of time evolution. This fixed point is unstable for $1 + 3w > 0$, (e.g. time evolution will drives the universe away from flatness) and stable (attractive) for $1 + 3w < 0$ (e.g. time evolution drives the universe towards flatness). Most of the matter present in the evolution of the universe can be approximated as non-relativistic (with $w = 0$) or radiation (with $w = \frac{1}{3}$). Importantly, both of these satisfy $1 + 3w > 0$, and so a universe dominated by these forms of matter will tend to evolve away from flatness.

This leads us to the *flatness problem*. The universe today appears to be very close to flat, despite the fact that most ordinary matter tends to drive us away from flatness. Naïvely, this tells us that at the time of the big bang, the curvature of the universe would have to be fine-tuned to be even flatter than it appears today. To avoid these kinds of fine-tuned initial conditions on the universe, we need there to be some period of time in the early universe where $1 + 3w < 0$. This corresponds to the universe undergoing accelerated expansion, referred to as *inflation*.

1.3 (Slow-Roll) Inflation to the Rescue!

To see how to actually achieve inflation, we will consider a low-energy effective field theory approach in which our universe is described by Einstein-Hilbert gravity coupled to some matter content. For simplicity, we will take this matter to be a scalar field ϕ with minimal coupling to gravity. This scalar field will be what drives inflation, so we will often refer to it as the *inflaton*.

The action for our toy model of inflation is

$$S = S_{\mathrm{EH}} + S_\phi = \int d^4x\, \sqrt{-g}\left[\frac{1}{2}R + \frac{1}{2}(\partial\phi)^2 - V(\phi)\right] . \tag{1.15}$$

Extremizing this action with respect to the metric results in the Einstein equation

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi T_{\mu\nu} , \tag{1.16}$$

where the stress-energy tensor $T_{\mu\nu}$ for the scalar matter is

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S_\phi}{\delta g^{\mu\nu}} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}(\partial\phi)^2 + V(\phi)\right) . \tag{1.17}$$

If we assume the scalar field to be completely homogeneous, e.g. $\phi(\vec{x}, t) = \phi(t)$, then the stress-tensor for the scalar matter is that of a perfect fluid (1.5) with

$$\rho = \frac{1}{2}\dot{\phi}^2 - V(\phi) , \quad p = \frac{1}{2}\dot{\phi}^2 + V(\phi) . \tag{1.18}$$

The state parameter is thus given by

$$w = \frac{p}{\rho} = \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)} . \tag{1.19}$$

As discussed in section 1.2, the flatness problem requires that at some point in the cosmological history of our universe, the universe undergoes inflation, e.g. accelerated expansion with $w < -\frac{1}{3}$. We therefore need there to be some period of time in which the potential energy dominates over the kinetic energy of the inflaton:

$$V(\phi) \gg \frac{1}{2}\dot{\phi}^2 . \tag{1.20}$$

Moreover, this period of accelerated expansion can't be arbitrarily short. Current cosmological experimental data on correlations in the CMB indicate that inflation must have lasted around 60 e-folds, meaning that the ratio of the scale factor at the beginning of inflation (e.g. when the CMB was created) to the scale factor at the end of inflation must satisfy

$$\log \frac{a_{\rm end}}{a_{\rm CMB}} \sim 60 \ . \tag{1.21}$$

In order to sustain the accelerated expansion for this period of time, we need the second derivative of ϕ to satisfy

$$|\ddot{\phi}| \ll |3H\dot{\phi}| \ . \tag{1.22}$$

We can therefore define the slow-roll parameters

$$\begin{aligned} \varepsilon &= -\frac{\dot{H}}{H^2}\left(= \frac{3}{2}(w+1) \text{ for a homogeneous scalar}\right) , \\ \eta &= -\frac{\ddot{\phi}}{H\dot{\phi}} \ . \end{aligned} \tag{1.23}$$

Then, in order to have an inflationary epoch consistent with cosmological data, we require that

$$\boxed{\varepsilon, |\eta| \ll 1} \ . \tag{1.24}$$

These are known as the *slow-roll conditions*, because they correspond to the inflaton "slowly" rolling down its potential. A sample slow-roll potential is shown in figure 1.

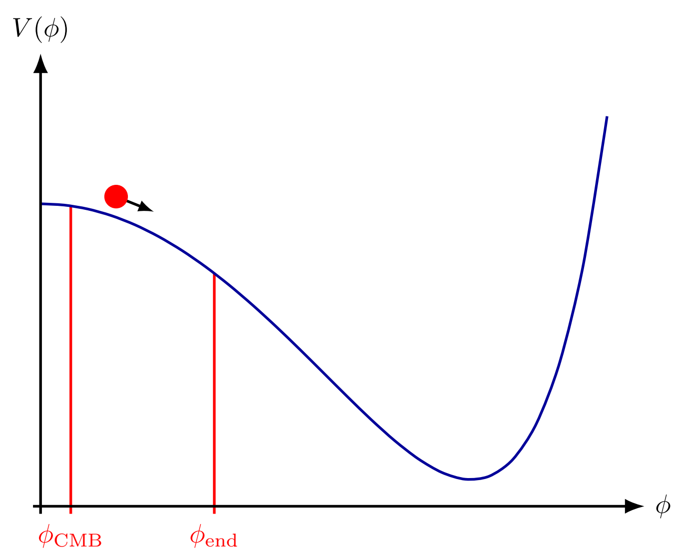

Figure 1. A sample slow-roll inflation potential. As the inflaton rolls down the potential, inflation begins when the CMB is created at $\phi_{\rm CMB}$ and ends when $\varepsilon = 1$ at $\phi_{\rm end}$. The potential energy of the inflaton dominates over its kinetic energy in the region between, where $\varepsilon, |\eta| \ll 1$.

Up until now, we have been working in units where $M_{\rm pl} = 1$. We will restore the Planck mass in what follows. In the slow-roll regime, we can approximate the slow-roll parameters purely in terms of the potential by

$$\begin{aligned} \varepsilon &\approx \frac{M_{\rm pl}^2}{2}\left(\frac{V'}{V}\right)^2 , \\ \eta &\approx M_{\rm pl}^2\left[\frac{V''}{V} - \frac{1}{2}\left(\frac{V'}{V}\right)^2\right] . \end{aligned} \tag{1.25}$$

These expressions are easier to use, because they do not require solving the Friedmann equations. These are only approximate expressions, though, and are valid only in areas where both V' and V'' are relatively small.

2 Large-Field Inflation and Axions

At this point, we've studies what conditions are necessary in order for a model of slow-roll inflation to be consistent with current cosmological data. However, these conditions are not very constraining; there are plenty of inflationary models out there, with not a whole lot of ways to rule them out experimentally with current data sets.

So, how do we choose what inflationary models to look for? That is, what makes a theory of inflation *good*, in some sense? There are plenty of differing opinions on this subject, but as someone whose background is in string theory, I claim that a good theory of inflation should have two reasonable features (as well as one feature that is maybe unreasonable):

1. **Experimental signals**. Good models make new predictions beyond the data they are built to match, allowing them to be confirmed or refuted by future experimental efforts.

2. **Lack of fine-tuning**. If a model's parameters have to be very precisely tuned to match data, the model is unsatisfying, as it doesn't provide a physics reason to expect the parameters to take these fine-tuned values.

3. **Sensitivity to Planck-scale physics**. The Standard Model is an effective theory that only describes physics below the Planck scale, where quantum gravity effects become important. Inflationary models with Planck-sensitivity offer us a window into these quantum gravity effects and could give us a better understanding of how to deal with gravity at high energies.

In this section, we will detail how large-field axion inflation models satisfy all three of these criteria, making them an interesting class of models to study in-depth.

2.1 Small-Field versus Large-Field

The CMB is created at some inflaton field value $\phi_{\rm CMB}$ that occurs a few e-folds after the Big Bang. The inflaton then undergoes a slow-roll period, where its potential dominates over its kinetic energy. This lasts for ~ 60 e-folds until the inflaton field value hits ϕ_{end}, signaling the end of inflation. We can define the field excursion range of the inflaton associated with inflation by

$$\Delta\phi = |\phi_{\text{end}} - \phi_{\text{CMB}}| \ . \tag{2.1}$$

We can heuristically break down models of inflation into two classes based on how the inflaton field range $\Delta\phi$ compares to the Planck scale M_{pl}:

- **Small-field inflation**, where $\Delta\phi < M_{\text{pl}}$. These are sometimes called "sub-Planckian" models.
- **Large-field inflation**, where $\Delta\phi > M_{\text{pl}}$. These are often referred to as "trans-Planckian" models.

Large-field inflation models satisfy two of the conditions presented above: they give experimentally-testable predictions for gravitational wave signals, and they are more sensitive to UV physics. Additionally, small-field inflation models tend to be fairly fine-tuned to whatever the most up-to-date constraints on non-Gaussianity in the CMB are. The bulk of these models would be ruled out (or, at the very least, constrained) through measuring such a non-Gaussianity. Large-field models tend to be more robust in that sense.

Experimental Predictions

We have so far talked about FRW universes that are perfectly homogeneous. However, quantum fluctuations of matter density in the early universe (referred to as *primordial fluctuations*) will break this homogeneity. These fluctuations lead to a departure of the metric from its FRW form.

The perturbations of the spacetime metric away from its FRW form are characterized by two types of perturbations: scalar perturbations and tensor perturbations. As the perturbations evolve in time, distances measured by the metric will change in time as well, which generically would lead us to expect gravitational waves being emitted. Importantly, only the tensor perturbations actually lead to gravitational waves, and not the scalar modes. The relative strength of these two types of metric perturbations are characterized by the *tensor-to-scalar ratio* r. Higher values of r correspond to larger gravitational wave signals.

We won't go very in-depth into this, but you can compute r by studying correlation functions of the primordial fluctuations. Doing so leads to the relation

$$r = \frac{8}{M_{\text{pl}}^2}\left(\frac{d\phi}{dN}\right)^2 \ , \tag{2.2}$$

where $dN \equiv d\log a$ is the differential of the number of e-folds. By combining this with current experimental constraints on the number of e-folds that inflation lasts, we end up with the *Lyth bound* [1]:

$$\frac{\Delta\phi}{M_{\text{pl}}} \approx \mathcal{O}(1)\left(\frac{r}{0.01}\right)^{1/2} \ . \tag{2.3}$$

For large-field inflation models with $\Delta\phi > M_{\rm pl}$, we therefore predict to measure a tensor-to-scalar ratio of $r > 0.01$. Current experimental bounds set an upper limit of $r < 0.11$ (95% CL) [2], and the next decade of experiments should probe values closer to $r = 0.01$. Large-field inflation models therefore couple more strongly to gravitational waves in the early universe, allowing experimental probes of these models in the near future.

Sensitivity to UV Physics

From an effective field theory perspective, we cannot simply write down some fine-tuned potential for the inflaton that is chosen specifically to make inflation work. The action must instead consist of all possible interaction terms, even higher-derivative ones, that are consistent with the symmetries of the theory. For example, let's suppose that our theory has a $\mathbb{Z}_2$ symmetry $\phi \to -\phi$. In the absence of any cut-off on our theory up to the Planck scale, the most general effective field theory we can write down for the inflaton is

$$\mathcal{L} = -\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4 + \sum_{p=1}(\lambda_p\phi^4 + c_p(\partial\phi)^2)\left(\frac{\phi}{M_{\rm pl}}\right)^{2p} + \dots\ , \qquad (2.4)$$

where λ_p, c_p are some dimensionless coupling constants and we have suppressed lots of other possible higher-dimension operators. Note that the only scale we have to suppress the higher-dimension operators is the Planck scale. Additionally, without any extra symmetries or constraints to impose on the theory, we generically expect the coefficients λ_p, c_p to be $\mathcal{O}(1)$ to prevent fine-tuning.

This leads to a problem for our effective field theory of inflation: for large-field inflation models with Planck-scale field excursions $\Delta\phi \gtrsim M_{\rm pl}$, all higher-dimension operators will contribute at the same order as the relevant dimension-four operators! We can no longer use the standard effective field theory approach and truncate our theory up to some level of sub-leading terms. These large field models will thus generically require summing up an infinite number of higher-dimension terms. The inflationary dynamics are therefore incredibly sensitive to the UV physics.

What about fine-tuning?

We now have a problem: without knowing everything about the UV dynamics of the theory, it's almost impossible to say whether our model will be consistent with constraints coming from inflationary data. Even if you start with a nice slow-roll potential, this infinite tower of corrections could destroy this potential and leave you with no accelerated expansion. We therefore need some way to set all of the coupling constants λ_p, c_p to very tiny. However, this would make our theory fine-tuned, which is a trait we specifically wanted to avoid!

The way around this is to find a physical mechanism that forces these higher-order coupling constants to vanish, without having to fine-tune the theory by hand. In particular, we will look for new symmetries that can be used to constrain the set of possible operators appearing in the effective action. A particularly nice symmetry that accomplishes this is having a shift leading us to consider axion models of inflation in the next section.

2.2 Axions and the Peccei-Quinn Mechanism

The main idea behind the Peccei-Quinn mechanism [3] is to start with a theory with a continuous global symmetry (typically some $U(1)$) that is *spontaneously* broken due to fields acquiring vacuum expectation values. The breaking of the global symmetry introduces Nambu-Goldstone bosons at this breaking scale. These Nambu-Goldstone bosons will have a discrete shift symmetry that ensures that their potential receives very little in the way of corrections that could destroy its slow-roll potential.

To see an explicit example, consider a theory of a fermion ψ and a complex scalar field Φ coupled to gravity:

$$S = \int d^4x \sqrt{-g}\left[|\partial_\mu \Phi|^2 - V(|\Phi|) + \xi|\Phi|^2 R + i\bar{\psi}\partial\!\!\!/\psi - (y\bar{\psi}_L\psi_R\Phi + \text{h.c.})\right] . \tag{2.5}$$

Here, ξ is a real coupling constant and y a complex Yukawa coupling constant. This action is invariant under a global chiral $U(1)$ symmetry:

$$\begin{aligned} \psi_L &\to e^{i\alpha}\Psi_L \ , \\ \psi_R &\to e^{-i\alpha}\psi_R \ , \\ \Phi &\to e^{i\alpha}\Phi \ , \end{aligned} \tag{2.6}$$

for some phase angle α. Loop effects in the theory will typically generate a potential for Φ of the form

$$V(|\Phi|) = \lambda\left(|\Phi|^2 - \frac{f^2}{2}\right)^2 , \tag{2.7}$$

where f is some mass scale. The resulting scalar field VEV for this potential is

$$\langle\Phi\rangle = \frac{fe^{i\phi/f}}{\sqrt{2}} , \tag{2.8}$$

where ϕ is some real phase.

When considering fluctuations of Φ around its vacuum expectation value, we can think of two modes: the radial mode, oscillating around the minimum of the potential, and the axial modes ϕ. The axial modes are massless, while the radial modes have a mass

$$m_{\text{radial}} \sim \sqrt{\lambda}f \ . \tag{2.9}$$

Therefore, for energy scales well below the breaking scale ($E \ll f$), the radial modes become very hard to excite. In the low energy limit, then, our theory has a massless field ϕ. Moreover, this field has a discrete shift symmetry $\phi \to \phi + 2\pi f$, making it an axion. This shift symmetry protects the axion from receiving radiative corrections, and so its potential remains flat at all orders in perturbation theory.

Now, what can we say about the symmetry breaking scale f? The input here comes from the fact that, at low energies $E \ll f$, our theory must look like the Einstein-Hilbert action coupled to the axion ϕ and the fermion ψ. The Ricci scalar term in the action must therefore, at low energies, take the usual Einstein-Hilbert term such that

$$S = \int d^4x \sqrt{-g} \left(\frac{M_{\rm pl}^2}{16\pi} R + \dots \right) . \tag{2.10}$$

The actual Ricci scalar term in the action, when evaluated at the vacuum expectation value for Φ, takes the form

$$\xi |\Phi^2| R = \frac{\xi f^2}{2} R . \tag{2.11}$$

Upon identification of (2.10) and (2.11), we find that the coupling constant ξ is given by

$$\xi = \frac{M_{\rm pl}^2}{8\pi f} . \tag{2.12}$$

A natural (e.g. not finely-tuned) value for ξ will be $\mathcal{O}(1)$, which sets f to be roughly around the Planck scale. We therefore conclude that the global $U(1)$ is spontaneously broken at $f \sim M_{\rm pl}$, and the resulting theory at energies $E \ll M_{\rm pl}$ contains a shift-symmetric axion with a perturbatively flat potential.

This toy model gives a good idea of how the Peccei-Quinn mechanism works, but it is missing one ingredient: the model lacks any gauge field. This is important, because gauge fields naturally give rise to topological terms in the action that are total derivatives. The axion can couple to these topological terms, as they are consistent with their shift symmetry. In particular, if we have a theory with an axion and a non-Abelian gauge field A_μ^a, we expect an interaction of the form

$$\mathcal{L} = \theta \, \mathrm{tr} \, F_{\mu\nu}^a F^{a\mu\nu} \ , \quad \theta = \frac{\phi}{f} . \tag{2.13}$$

These kinds of couplings give rise to non-perturbative effects that break the original global symmetry, giving rise to a potential for the axion. This potential was explicitly derived in [4] in the context of QCD, but their arguments hold for more general gauge groups.

The essence of their argument is as follows. By studying the vacuum structure of the theory, you find a bunch of topologically distinct vacua labelled by θ. Moreover, the vacuum energy actually depends on θ. You can explicitly construct instantons (e.g. non-perturbative tunneling solutions that solve the classical equations of motion) that tunnel between these vacua. These non-perturbative effects break the global symmetry, and they generate a potential for the axion field of the form

$$V(\phi) = \Lambda^4 \left[1 + \cos\left(\frac{\phi}{f} \right) \right] , \tag{2.14}$$

where Λ is some dynamically-generated scale that comes from the details of the gauge group that the axion couples to. (For example, $\Lambda = \Lambda_{\rm QCD}$ when using the Peccei-Quinn mechanism to solve the strong CP problem in QCD.)

Non-perturbative contributions hence generate a potential for the axion that is stable and won't be destroyed by an infinite tower of UV effects. A plot of this potential is shown in figure 2.

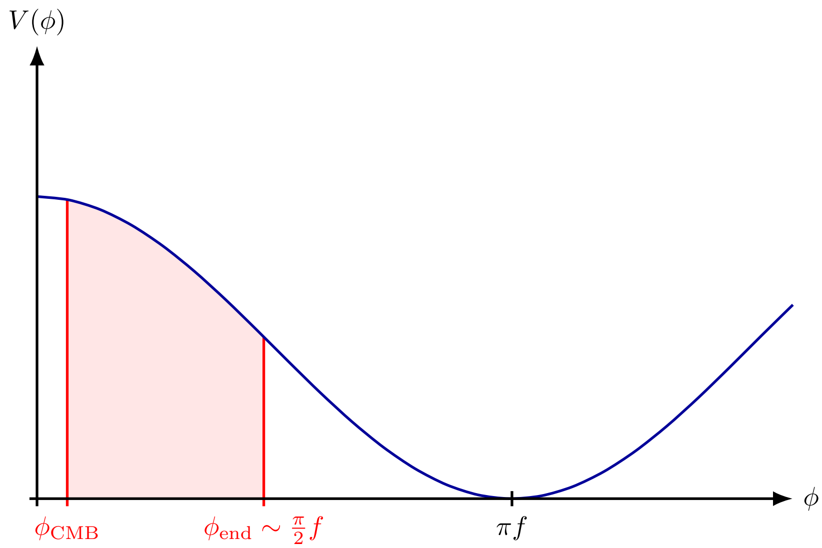

Figure 2. The axion potential (2.14) generated by non-perturbative effects in the Peccei-Quinn mechanism. The area shaded is where we expect slow-roll conditions to be satisfied.

Slow-roll inflation can be achieved when the potential energy of the axion dominates over its kinetic energy, which occurs approximately in the shaded region of figure 2. This region has a width of roughly $\sim f$, and so achieving a trans-Planckian field excursion range requires

$$f \gtrsim M_{\text{pl}} \ . \tag{2.15}$$

We can make this more precise by computing the the slow-roll parameters for the potential:

$$\varepsilon = \frac{M_{\text{pl}}^2}{2f^2} \tan^2\left(\frac{\phi}{f}\right) \ , \quad \eta = -\frac{M_{\text{pl}}^2}{f^2} \frac{\cos(\phi/f)}{1+\cos(\phi/f)} \ . \tag{2.16}$$

Slow roll ends when $\varepsilon = 1$, which occurs at

$$\phi_{\text{end}} = f \tan^{-1}\left(\frac{\sqrt{2}f}{M_{\text{pl}}}\right) \ . \tag{2.17}$$

For $\phi < \phi_{\text{end}}$, the slow-roll parameters are both small, so inflation occurs. The maximum size for the field excursion range will occur when $\phi_{\text{CMB}} = 0$, which bounds the excursion range such that

$$\Delta\phi \leq f \tan^{-1}\left(\frac{\sqrt{2}f}{M_{\text{pl}}}\right) \leq \frac{\pi}{2} f \ . \tag{2.18}$$

The axion decay width f therefore determines the size of the axion field excursion range. In particular, we require $f \gtrsim M_{\rm pl}$ in order to have trans-Planckian field ranges, in accordance with our earlier graphical estimate (2.15).

Our axion model therefore naturally gives rise to an inflationary potential with a trans-Planckian field range. This model is referred to as *natural inflation*, and it currently fits very well into experimental data constraints coming from the 2015 Planck data [5].

2.3 Stringy Axions

As it turns out, another interesting aspect of axions is how often they come up in string theory. In particular, they are everywhere when compactifying gauge fields in string theory. As a toy example, consider a $d = 5$ theory with a $U(1)$ gauge field $\mathcal{A}_M$, where $M = (0, \ldots, 4)$ indexes the one time direction and four spatial directions.

Let's suppose now that the $d = 5$ space has an $\mathcal{M}_4 \times S^1$ topology with a radius R for the compact fourth spatial direction. If we compactify along x^4, $\mathcal{A}_M$ decomposes into a vector A_μ and a scalar ϕ:

$$\begin{aligned} A_\mu(x) &= \oint_0^R dx^4 \mathcal{A}_\mu(x, x^4) \ , \\ \phi(x) &= \oint_0^R dx^4 \mathcal{A}_4(x, x^4) \ . \end{aligned} \tag{2.19}$$

Importantly, the original gauge field had a $U(1)$ gauge symmetry that these dimensionally-reduced fields inherit. In particular, if we consider large gauge transformations along the circle, ϕ has a discrete $\phi \to \phi + 2\pi R$ symmetry.

More generally, dimensional reduction of a theory in $d > 4$ with (potentially non-Abelian) gauge fields leads to a $d = 4$ theory with (non-Abelian) gauge fields and axions. Since the axions and gauge fields descend from the same field in the $d > 4$ theory, the axions will naturally couple to gauge field strengths. The Peccei-Quinn mechanism is in full force again, and the axion will end up getting a potential from non-perturbative instanton effects of the form [6]

$$V(\phi) \sim \Lambda^4 \sum_{n=1}^{\infty} e^{-(n-1)S_{\rm inst}} \left[1 + \cos\left(\frac{n\phi}{f}\right) \right] \ . \tag{2.20}$$

Note that we have included higher harmonics in this potential. Roughly speaking, these higher-harmonics come from higher-order instantons that require more energy to tunnel. These higher harmonics are correspondingly weighted by higher powers of the tunneling probability $e^{-S_{\rm inst}}$.

3 Weak Gravity Conjecture

We've so far spent a lot of time motivating how a large class of trans-Planckian axion inflation models are interesting both for experimental and theoretical reasons. However, we are ultimately interested in theories that have a good embedding into a theory of quantum gravity in the UV. We therefore need to determine if these trans-Planckian axion inflation models do have a sensible UV embedding. If not, then we should rule them out and move on to a new theory. These theories that do not come from a UV theory of gravity are referred to as being in "*the swampland*" [7].

Determining if a given theory is in the swampland is easier said than done. Since the only UV-complete theory of quantum gravity that we know of is string theory (which remains a somewhat formidable foe to deal with), a top-down method where you run a quantum gravity theory down to low energies is out of the question. Instead, we need conditions that can be posed in the low-energy effective field theory itself.

3.1 Gravity as the Weakest Force

One potential criteria for determining if a theory is in the swampland or not is the *weak gravity conjecture* (abbreviated WGC), originally posited in [8]. Heuristically, this conjecture is the notion that gravity is the weakest force in the universe around us.

Quantitatively, the conjecture is as follows. Consider a $d = 4$ low-energy effective field theory with a $U(1)$ gauge field in it, and then couple this theory to gravity. In order for this theory to have a UV-completion, there must be a particle in the theory with mass m and charge q that is *superextremal*:

$$\frac{q}{m} \geq \frac{1}{M_{\text{pl}}} \ . \tag{3.1}$$

The motivation for this conjecture comes from black hole considerations. If our low-energy effective field theory has a $U(1)$ gauge field and a graviton, it follows that charged black holes should be allowed to form. However, not any mass M and charge Q are allowed; cosmic censorship requires that the ratio of these charges satisfy

$$\frac{Q}{M} \leq \frac{1}{M_{\text{pl}}} \ . \tag{3.2}$$

An extremal (non-rotating) black hole (abbreviated EBH) is one with the maximum amount of charge Q allowed for a given mass M, the ratio of which is equal to M_{pl}^{-1}.

A more intuitive way to see this bound is to note that, in the world around us, the electromagnetic force is stronger than the gravitational force. So, if we were to throw a bunch of test charges into a black hole, eventually the charge of the black hole would be large enough that the electromagnetic force would win over the gravitational force, and so the black hole would start to repel the charges. The point where all charges start to repel is precisely the extremality bound.

A feature of these extremal black holes is that they have no Hawking temperature,

$$T\big|_{\text{EBH}} = 0 \ , \tag{3.3}$$

and are thus stable under Hawking radiation. Additionally, conservation of charge and mass prevent an extremal black hole from decaying into other black holes. This can't be the full story, though; we would like a mechanism for them to decay. Otherwise, our universe could potentially fill up with an incredibly large number of stable objects that are not protected by any symmetry. This is known as the *remnant problem* [9]. Solving this was the motivation for the original WGC, as the existence of a superextremal particle in the spectrum is precisely what is needed to allow an extremal black hole to decay into a (smaller) black hole and some particles.

We will now show that superextremal particles are necessary for extremal black holes to decay. Suppose that we have an EBH with mass M and charge Q at rest. This EBH then decays into a smaller black hole with mass M' and charge Q' as well as a particle of mass m and charge q. Since the decay products can be ejected with some velocity, conservation of energy tells us that

$$m + M' \leq M \ , \tag{3.4}$$

while conservation of charge dictates that

$$q + Q' = Q \ . \tag{3.5}$$

Combining these two equations tells us that

$$\frac{q + Q'}{m + M'} \geq \frac{Q}{M}\bigg|_{\text{EBH}} = \frac{1}{M_{\text{pl}}} \ . \tag{3.6}$$

Then, since the smaller black hole must also respect the extremality bound, we know that $M' \geq Q' M_{\text{pl}}$. We therefore find that

$$\frac{q + Q'}{m + Q' M_{\text{pl}}} \geq \frac{q + Q'}{m + M'} \geq \frac{1}{M_{\text{pl}}} \ . \tag{3.7}$$

Moving over the denominator on the left-hand side of this inequality yields

$$q + Q' \geq \frac{m}{M_{\text{Pl}}} + Q' \ . \tag{3.8}$$

Cancelling common factors of Q', we therefore recover the WGC:

$$\boxed{\frac{q}{m} \geq \frac{1}{M_{\text{pl}}}} \ . \tag{3.9}$$

Note that this seems like a plausible argument, but it is not a full proof. The best evidence we have for the WGC actually comes from string theory. Namely, every known compactification in string theory satisfies the WGC!

3.2 Generalization to p-forms in d-dimensions

More generally, there are extended objects in string theory known as *branes* that can be charged under higher-form gauge fields. In particular, any $(p-1)$-brane will naturally be charged under a p-form gauge field; we denote this charge by q_p. Instead of a mass, these branes will have some tension T_{p-1}. Additionally, these objects can exist in string compactifications for a wide range of dimensions. Moreover, these objects can also stack up and eventually undergo gravitational collapse into extended objects with horizons that (due to cosmic censorship) satisfy some sort of extremality relation between the tension and charge of the object. Our logic from before tells us that these objects must be able to decay, which requires the existence of superextremal branes in the theory.

All of this leads to the following generalization of the WGC. Suppose you have a d-dimensional theory with a p-form gauge field. Then, there must exist a $(p-1)$-dimensional brane (with tension T_{p-1}) that couples to this p-form gauge field. Moreover, the corresponding charge q_p of the brane must satisfy

$$\left(\frac{d-2}{2p(d-p-2)}\right)\frac{q_p}{T_{p-1}} \geq \frac{1}{M_d^{d-2}} \ , \tag{3.10}$$

where M_d is the d-dimensional Planck mass [10]. For $p=1$, $d=4$, we recover the form of the $U(1)$ WGC presented above. This works more generally, though, with the numerical factors chosen to be consistent with the extremality bounds for extended gravitational objects with horizons. This extension of the WGC has also been shown to hold in all known string compactifications for $p \geq 1$.

4 Weak Gravity Constraints on Axion Inflation

In this section, we bring everything together and discuss how the weak gravity conjecture puts constraints on the large-field axion inflation models discussed earlier. We end by mentioning areas in which ongoing research is being done to understand these constraints better.

4.1 Generalizing the Weak Gravity Conjecture to Axions

It is conjectured that the WGC for p-forms can be extended to the case when $p=0$ as well, where a 0-form is an axion with a shift symmetry and the (-1)-dimensional object is an instanton that couples to the axion.[1] The coupling of the axion to the gauge group is $1/f$, which takes the place of the "charge", while the "mass" is S_{inst}, the on-shell action for the instanton, since this is a measure of how much energy it takes to excite a given instanton tunneling transition. The axion version of the WGC posits that

$$\frac{\text{“charge”}}{\text{“mass”}} \geq \frac{1}{M_{\text{pl}}} \ , \tag{4.1}$$

which translates into [11]

$$S_{\text{inst}} \leq \frac{M_{\text{pl}}}{f} \ . \tag{4.2}$$

[1] A particle is a 0-dimensional object that traces out a 1-dimensional wordline in spacetime, while a string is a 1-dimensional object that traces out a 2-dimensional wordline. Since instantons are solutions to the Euclidean equations of motion at a single point in spacetime, they can be thought to have a 0-dimensional worldine. We therefore identify them as (-1)-dimensional objects.

Morally, what this tells us is that the amount of energy it takes to excite any given instanton transition can't be unboundedly high. Moreover, having very large axion decay widths f makes the bound even tighter, making it even easier to excite all instanton transitions.

4.2 Ruling out Trans-Planckian Field Ranges

With this in mind, let's reconsider the axion potential

$$V(\phi) \sim \Lambda^4 \sum_{n=1}^{\infty} e^{-(n-1)S_{\text{inst}}} \left[1 + \cos\left(\frac{n\phi}{f}\right)\right] . \tag{4.3}$$

Suppose that we have an axion inflation model for which $f \sim NM_{\text{pl}}$ for some $N > 1$. (Remember, having $f > M_{\text{pl}}$ is required to get a trans-Planckian axion field range, as discussed in section 2.2). Each harmonic in the potential is weighted by the tunneling prefactor. If the axion version of the WGC holds, then this tunneling prefactor is bounded by

$$e^{-(n-1)S_{\text{inst}}} \geq e^{-(n-1)\frac{M_{\text{pl}}}{f}} \sim e^{-\frac{(n-1)}{N}} . \tag{4.4}$$

This indicates that the tunneling prefactor is $\mathcal{O}(1)$ for each harmonic up to $n \sim N$. This is bad, though, because the higher harmonic terms in the potential oscillate faster and shorten the field excursion range of the axion over inflation. The n-th harmonic term is periodic under

$$\phi \to \phi + \frac{2\pi N}{n} M_{\text{pl}} . \tag{4.5}$$

The potential energy therefore becomes comparable to the kinetic energy of the field rolling down the slope at $\phi \sim \frac{\pi N}{2n} M_{\text{pl}}$, and so the field excursion range becomes bounded by

$$\Delta\phi \lesssim \frac{\pi N}{2n} M_{\text{pl}} . \tag{4.6}$$

Since harmonics up to $n = N$ contribute we therefore find that it is *hard* to get the field excursion range to be much larger than M_{pl}, because the $n = N$ term limits the field range such that $\Delta\phi \lesssim \frac{\pi}{2} M_{\text{pl}}$.

We can do a more quantitative analysis by computing the slow-roll parameters. For the n-th harmonic alone, these are given by

$$\begin{aligned} \varepsilon^{(n)} &= \frac{n^2}{2N^2} \tan^2\left(\frac{n\phi}{NM_{\text{pl}}}\right) , \\ \eta^{(n)} &= -\frac{\cos\left(\frac{n\phi}{NM_{\text{pl}}}\right)}{1 + \cos\left(\frac{n\phi}{NM_{\text{pl}}}\right)} . \end{aligned} \tag{4.7}$$

Solving for $\varepsilon^{(n)} = 1$ tells us that inflation ends when

$$\phi_{\text{end}}^{(n)} = \frac{2NM_{\text{pl}}}{n} \tan^{-1}\left(\frac{\sqrt{2}N}{n}\right) . \tag{4.8}$$

For values of ϕ well below this, the slow-roll parameters remain small and thus inflation occurs. Since we only care about terms for which $N \geq n$, and the arctan function asymptotes to $\frac{\pi}{2}$ this sets a bound on when inflation ends:

$$\phi_{\text{end}} \leq \frac{\pi N M_{\text{pl}}}{2n} . \tag{4.9}$$

The CMB must be created at some field value smaller than this, and so we immediately recover the constraint on the field range estimated earlier in (4.6).

Admittedly, this analysis treated each harmonic separately, but adding all harmonics from $n = 1$ to $n = N$ together and repeating the analysis is fairly straightforward. The result is the same; namely, that the field excursion range is bounded by

$$\Delta\phi \lesssim \frac{\pi}{2} M_{\text{pl}} , \tag{4.10}$$

This result was obtained by assuming that only terms up to $n = N$ contribute meaningfully, which is true for a theory that exactly saturates the WGC bound. If we had a theory that was below this bound, though, we'd find that even higher harmonics contribute, which would limit the field excursion range even more. The WGC conjecture therefore puts a very hard constraint on trans-Planckian axion inflation, such that saturating the WGC bound only allows for field excursion ranges of the form

$$\boxed{\Delta\phi \lesssim \mathcal{O}(1) M_{\text{pl}}} . \tag{4.11}$$

Therefore, according to the axion version of the weak gravity conjecture, large-field axion inflation models are solidly in the swampland.

4.3 Loopholes?

There are attempts to evade these constraints within the string phenomenology community by cooking up theories with multiple axions. Even if each individual axion has a sub-Planckian field range, they can be aligned to give an effective net field excursion range that is trans-Planckian [12]. However, such theories could be in danger of violating the convex hull conjecture, a generalization of the weak gravity conjecture for theories with multiple types of p-forms [13, 14]. Whether or not axion alignment can evade the weak gravity conjecture is a current subject of debate.

A possibly more damning piece of evidence against these models is that no one has found any explicit realization of trans-Planckian axion inflation in string theory. It's hard to say if that's because string theory doesn't allow for them, or if this is just selection bias within the class of compactifications that we actually understand. This is still a very hotly-contested area of research, and it will be exciting to see where things end up in the next several years.

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) 1861–1863, [hep-ph/9606387].

[2] **Planck** Collaboration, P. A. R. Ade et al., *Planck 2015 results. XX. Constraints on inflation*, *Astron. Astrophys.* **594** (2016) A20, [arXiv:1502.02114].

[3] R. D. Peccei and H. R. Quinn, CP, *Phys. Rev. Lett.* **38** (Jun, 1977) 1440–1443.

[4] C. G. Callan, R. Dashen, and D. J. Gross, *Toward a theory of the strong interactions*, *Phys. Rev. D* **17** (May, 1978) 2717–2763.

[5] E. Pajer and M. Peloso, *A review of Axion Inflation in the era of Planck*, *Class. Quant. Grav.* **30** (2013) 214002, [arXiv:1305.3557].

[6] T. Rudelius, *Constraints on Axion Inflation from the Weak Gravity Conjecture*, *JCAP* **1509** (2015), no. 09 020, [arXiv:1503.00795].

[7] C. Vafa, *The String landscape and the swampland*, hep-th/0509212.

[8] 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].

[9] L. Susskind, *Trouble for remnants*, hep-th/9501106.

[10] B. Heidenreich, M. Reece, and T. Rudelius, *Evidence for a sublattice weak gravity conjecture*, *JHEP* **08** (2017) 025, [arXiv:1606.08437].

[11] B. Heidenreich, M. Reece, and T. Rudelius, *Weak Gravity Strongly Constrains Large-Field Axion Inflation*, *JHEP* **12** (2015) 108, [arXiv:1506.03447].

[12] T. C. Bachlechner, C. Long, and L. McAllister, *Planckian Axions and the Weak Gravity Conjecture*, *JHEP* **01** (2016) 091, [arXiv:1503.07853].

[13] 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].

[14] J. Brown, W. Cottrell, G. Shiu, and P. Soler, *On Axionic Field Ranges, Loopholes and the Weak Gravity Conjecture*, *JHEP* **04** (2016) 017, [arXiv:1504.00659].