

This is the accepted manuscript made available via CHORUS. The article has been published as:

Inflation that runs naturally: Gravitational waves and suppression of power at large and small scales

Quinn E. Minor and Manoj Kaplinghat

Phys. Rev. D **91**, 063504 — Published 3 March 2015

DOI: [10.1103/PhysRevD.91.063504](https://doi.org/10.1103/PhysRevD.91.063504)

Inflation that runs naturally: gravitational waves and suppression of power at large and small scales

Quinn E. Minor

*Department of Science, Borough of Manhattan Community College,
City University of New York, New York, NY 10007, USA and*

Department of Astrophysics, American Museum of Natural History, New York, NY 10024, USA

Manoj Kaplinghat

Department of Physics and Astronomy, University of California, Irvine CA 92697, USA

We point out three correlated predictions of the axion monodromy inflation model: large amplitude of gravitational waves, suppression of power on horizon scales and on scales relevant for the formation of dwarf galaxies. While these predictions are likely generic to models with oscillations in the inflaton potential, the axion monodromy model naturally accommodates the required running spectral index through Planck-scale corrections to the inflaton potential. Applying this model to a combined data set of Planck, ACT, SPT, and WMAP low- ℓ polarization cosmic microwave background (CMB) data, we find a best-fit tensor-to-scalar ratio $r_{0.05} = 0.07^{+0.05}_{-0.04}$ due to gravitational waves, which may have been observed by the BICEP2 experiment. Despite the contribution of gravitational waves, the total power on large scales (CMB power spectrum at low multipoles) is lower than the standard Λ CDM cosmology with a power-law spectrum of initial perturbations and no gravitational waves, thus mitigating some of the tension on large scales. There is also a reduction in the matter power spectrum of 20-30% at scales corresponding to $k = 10 \text{ Mpc}^{-1}$, which are relevant for dwarf galaxy formation. This will alleviate some of the unsolved small-scale structure problems in the standard Λ CDM cosmology. The inferred matter power spectrum is also found to be consistent with recent Lyman- α forest data, which is in tension with the Planck-favored Λ CDM model with power-law primordial power spectrum.

Keywords: cosmology: cosmic background radiation, cosmology: inflation, galaxies: dwarf

I. INTRODUCTION

The Planck mission has established that the cosmic microwave background anisotropies on small angular scales are well described by the standard Λ CDM cosmology with a nearly scale-invariant power-law spectrum [1]. At large scales ($\ell < 40$), however, there appears an overall deficit of power compared to what is expected in the benchmark Λ CDM model [2]. The unusual shape and amplitude of the power spectrum at low multipoles was first observed by the WMAP mission [3, 4] and remains unexplained, although hypotheses include a running spectral index [5–9], a breakdown of slow-roll inflation or pre-slow roll phase [10–14], a contracting pre-inflation phase [15], open inflation [16], a non-Bunch-Davies initial vacuum state [17], or the presence of an extra neutrino species [18, 19].

The tension at low ℓ is exacerbated significantly if there exists a stochastic gravitational wave background with tensor-to-scalar ratio $r \gtrsim 0.1$, since tensor perturbations add to the expected CMB temperature anisotropies at low multipoles [20]. Such a large r has been suggested by the BICEP2 experiment in its detection of B-mode polarization in the sky at degree-angular scales [21]. Under the assumption that the observed B-mode anisotropies are sourced by primordial gravitational waves, they infer $r_{0.05} = 0.2^{+0.07}_{-0.05}$. By comparison, from the TT spectrum alone, [1] inferred $r_{0.05} < 0.135$ at the 95% confidence level, in tension with the BICEP2 result. At

present, the BICEP2 result is highly uncertain due to the likely presence of contamination by foreground dust polarization in the observed field-of-view [22–24]. Nevertheless, large-field inflation models—including the simplest chaotic models—predict at minimum $r \gtrsim 0.01$ [25] and thus would increase the apparent tension with Λ CDM.

The simplest way to accommodate the deficit of power at low ℓ is to allow for a running spectral index, thus departing from a power-law spectrum. When a constant running of the spectral index $\alpha = dn_s/d \ln k$ is allowed, [26] finds $\alpha = -0.011 \pm 0.008$ in the case $r = 0$, where the preference for negative running is driven largely by the temperature likelihood at low multipoles. More significantly, running also allows for a higher tensor contribution, leading to $r_{0.05} < 0.275$ when both running and tensors are allowed.

Such a large constant running is difficult to implement in the underlying inflation model, since it yields an insufficient number of e-foldings to solve the horizon problem [27]. Within the context of single-field slow-roll inflation, the only way to achieve the 50-60 remaining e-foldings necessary after the mode $k = 0.05 \text{ Mpc}^{-1}$ leaves the horizon, is if the running diminishes or turns positive at larger k . Plausible mechanisms exist for the running to diminish to zero at larger k , for example through radiative corrections [28, 29] or GUT symmetry breaking [30]; this would, however, imply a special scale at which the running “turns over” and becomes small, a scale comparable to or just smaller than scales observable in the CMB,

which would seem a remarkable coincidence. Another, perhaps more natural possibility is that the power spectrum oscillates, implying that the inflaton potential may contain an oscillatory component. Since it would seem unnatural for only one such oscillation to occur during the course of inflation, the intriguing possibility arises that the inflaton potential contains a gentle oscillation which may occur all the way to the end of inflation. In fact, many large-field models that include oscillations in a natural way have been investigated [31–36].

In this article we will test axion monodromy models against the CMB temperature anisotropy spectrum, but our results are broadly applicable to inflationary potentials with gentle oscillations. We will assume that the oscillation scale (in $\log k$) is “long”, i.e. comparable to the range of multipoles observed in the CMB, naturally leading to a running spectral index as described above. We will show that the best-fit model gives three correlated predictions: 1) a significant gravitational wave amplitude of order $r \sim 0.1$; 2) a reduction of power at large scales (low ℓ) despite the tensor contribution, thus mitigating the existing tension at large scales; 3) there is a corresponding significant suppression of power at small scales, particularly at the scales relevant to dwarf galaxy formation, which will alleviate some of the small-scale structure problems of Λ CDM. Finally, although axion monodromy allows for a gravitational wave background, we will find that the e-folding requirement surprisingly places an upper bound on the tensor-to-scalar ratio ($r \lesssim 0.2$), which is in tension with the BICEP2 measurement unless a significant portion of the observed B-mode signal is due to foreground contamination rather than primordial gravitational waves.

The paper is outlined as follows. In Section II we discuss the theoretical motivation for axion monodromy models and derive formulae for the power spectra for scalar and tensor perturbations. In Section III we discuss our choice of parameters and the priors for each. The resulting constraints are shown in Section IV; the constraints on the oscillation parameters are discussed in Section IV A, the best-fit model is discussed in Section IV B, and the constraint on the tensor-to-scalar ratio r is discussed in Section IV C. In Sections IV D and IV E we compare our model to the usual constant-running model at low and high multipoles, respectively. In Section IV F we discuss whether our model can be amended to allow for a higher tensor-to-scalar ratio. Section V investigates the implications of an oscillating power spectrum for the small-scale problems of Λ CDM: the effect on dwarf galaxy formation is discussed in Section V A, while in Section V B our results are compared to recent Lyman- α forest data and the prospects for other small-scale probes of the matter power spectrum are discussed. We conclude with a summary of our main points in Section VI.

II. THEORETICAL BACKGROUND

A. Motivation for large-field inflation models

During slow roll inflation, it is easily shown that a relation exists between the tensor-to-scalar ratio r and the overall shift $\Delta\phi$ in the scalar field from CMB scales to the end of inflation:

$$\frac{\Delta\phi}{M_p} = \mathcal{O}(1) \times \left(\frac{r}{0.01}\right)^{1/2} \quad (1)$$

where $M_p = (8\pi G)^{-1/2}$ is the reduced Planck mass. The importance of this relation, known as the Lyth bound [25], is that a significant gravitational wave contribution $r \gtrsim 0.01$ implies that the scalar field value changes by more than the Planck energy M_p . These so-called large-field models imply that any fields coupled to the inflaton with at least gravitational strength will receive corrections to the coupling strengths, resulting in an infinite series of Planck-suppressed terms contributing to the scalar field potential. Such terms spoil the flatness of the potential, inhibiting inflation, unless there is an exquisite degree of fine-tuning in the coefficients. The only way to avoid fine-tuning is if a symmetry “protects” the potential from large contributions—in particular, an approximate shift symmetry $\phi \rightarrow \phi + a$ in the corresponding Lagrangian ensures that the correction to the potential is *at worst* periodic, and thus its magnitude can in principle be controlled.

While many different potentials can approximately satisfy the above mentioned shift symmetry, there is no guarantee that such a potential admits a UV completion in quantum gravity; indeed, it is far more likely that the class of potentials which can be derived from a corresponding UV-complete theory obeying the same approximate shift symmetry is relatively restricted. For this reason, merely considering generic renormalizable effective field theories is not sufficient for large-field models. Instead, one should consider inflation models that can be derived from quantum gravity, with string theory being the best developed to date. Although a few alternatives exist [37–40], one of the best-motivated of these models is axion monodromy [32, 34, 41, 42]. Axion fields a in string theory naturally obey a discrete symmetry $a \rightarrow a + 2\pi$. The phenomenon of monodromy appears when axions are coupled to fluxes in a compactified higher-dimensional space (in the string theoretic description, this occurs in the presence of a wrapped brane in a Calabi-Yau manifold). The shift symmetry is slightly broken, allowing the field potential energy to change by a large amount as one traverses many cycles in the compactified space, while all the remaining microphysics is periodic in field space; this is analogous to a spiral staircase where the overall height can change without bound through many cycles.

The resulting potential for axion monodromy models consists of a monomial term plus a sinusoidal term:

$$V(\phi) = \lambda\phi^p + A \sin\left(\frac{\phi}{f} + \psi\right) \quad (2)$$

where we are now working in Planckian units where $M_p = (8\pi G)^{-1/2} = 1$ for the remainder of this paper. The parameter f corresponds to the period of oscillation and is known as the *axion decay constant*. Although in its original incarnation the monodromy potential has $p = 1$, depending on the flux coupling and brane configuration one can also achieve other discrete values such as $p = 2/3$, $p = 2$, $p = 3$ and so on [43]. Since we are interested here in what p -value(s) are preferred by CMB data, we will vary p as a free parameter, with the ansatz that p can be any positive real number.

B. Axion monodromy power spectrum

For fitting purposes, we find it useful to make the transformation $\phi \rightarrow \phi_{min} - \phi$, where ϕ_{min} will be chosen so that $\phi = 0$ corresponds to the CMB pivot scale $k_* = 0.05 \text{ Mpc}^{-1}$. With a suitable relabeling of parameters, we can recast the potential in the following form,

$$\frac{V}{V_*} = (1 - a \sin \delta) (1 - \phi/\phi_{min})^p + a \sin\left(\frac{\phi}{f} + \delta\right) \quad (3)$$

which is written so that $V(\phi = 0) = V_*$. We choose ϕ_{min} to be positive so the field rolls in the positive direction, i.e. ϕ increases as it rolls.

To find the primordial power spectra for scalar and tensor perturbations, one typically uses the slow roll approximation, which assumes the Hubble parameter varies slowly enough that the quasi-de Sitter solution of the equation for quantum perturbations can be used ([44]; for a review see [45]). Careful consideration is required here, however, because slow roll can break down if oscillations in the potential are sufficiently rapid, in which case the relevant equation must be solved directly. This occurs if the amplitude is sufficiently large, or if the axion decay constant f is small [41]. Since in this paper we focus only on long-wavelength oscillations in the power spectrum, the slow roll approximation holds very well in the parameter region of interest. The proof of this is given in appendix I for the interested reader.

The power spectrum for curvature perturbations in the slow-roll approximation is given by $\Delta_{\mathcal{R}}^2 \sim \frac{H^2}{\epsilon} \sim \frac{V^3}{V_{,\phi}^2}$. Normalizing the power spectrum by the scalar amplitude at the pivot scale A_s , we find

$$\Delta_{\mathcal{R}}^2 = A_s \mathcal{N}^2 \times \frac{\left[(1 - a \sin \delta) (1 - \phi_k/\phi_{min})^p + a \sin\left(\frac{\phi_k}{f} + \delta\right) \right]^3}{\left[(1 - a \sin \delta) (1 - \phi_k/\phi_{min})^{p-1} - \left(\frac{\phi_{min}}{pf}\right) a \cos\left(\frac{\phi_k}{f} + \delta\right) \right]^2}, \quad (4)$$

where ϕ_k corresponds to the scalar field value at the time when the mode with wavenumber k left the horizon (to be determined shortly), and

$$\mathcal{N} = 1 - a \sin \delta - \left(\frac{\phi_{min}}{nf}\right) a \cos \delta. \quad (5)$$

Note that at the pivot scale ($\phi_k = 0$), this reduces to $\Delta_{\mathcal{R}}^2 = A_s$ as it should.

Meanwhile, the primordial tensor power spectrum is given by $\Delta_t^2 \sim H^2 \sim V$, and since it is normalized by $A_t = r A_s$ where r is the tensor-to-scalar ratio at the pivot scale k_* , we have

$$\Delta_t^2 = r A_s \frac{V(\phi_k)}{V_*} \quad (6)$$

where V/V_* is given in equation 3.[46]

Finally, we still need the mapping $\ln\left(\frac{k}{k_*}\right) \rightarrow \phi_k$. In the slow roll approximation, this is given by

$$\frac{k}{k_*} = \sqrt{\frac{V}{V_*}} \exp\left\{ \int_0^\phi \left| \frac{V}{V_{,\phi}} \right| d\phi \right\}. \quad (7)$$

For a given set of parameter values, this integral can be calculated and inverted numerically; however, this would be computationally expensive to perform while the parameters are being varied during the MCMC procedure described in Section III. Thus, an analytic approximation is desirable here, which can be obtained as follows. Taking the log of equation 7, the formula is easily integrated and inverted if we first consider the no-oscillation case where $a = 0$. This yields

$$\phi_{k,0} \equiv \frac{p}{\phi_{min}} \ln\left(\frac{k}{k_*}\right). \quad (8)$$

For the case where $a \neq 0$, an approximate solution is found by expanding in $a \ll 1$ and keeping terms to first order in $\frac{a\phi_{min}}{nf}$. Upon integrating, the formula can be inverted approximately by substituting equation 8 into the sine term, producing the formula

$$\phi_k \approx \phi_{k,0} - \frac{a\phi_{min}}{p} \left\{ \sin\left(\frac{\phi_{k,0}}{f} + \delta\right) - \sin \delta \right\}. \quad (9)$$

Using this approximation in eqs. 4 and 6, we find the power spectrum differs from the exact numerical solution by less than 1% over the range $2 < \ell < 2500$ for all the test cases considered (using the approximate formula $\ell \approx k x_c$ where $x_c \approx 14100 \text{ Mpc}$ is the comoving angular diameter distance to last scattering).

III. SAMPLING AND PRIORS

We sample the model parameter space with a Markov Chain Monte Carlo (MCMC) method using the CosmoMC software package [47], which has been modified

to incorporate the power spectra in eqs. 4 and 6 along with the model parameters. For the sampling we use the Metropolis-Hastings algorithm extended by a “fast-slow” algorithm for efficiently sampling nuisance parameters [48]. In addition to Planck data [49], the likelihood includes WMAP 9-year polarization data [50] as well as small-scale CMB data from the ACT [51] and SPT surveys [52].

A. Choice of model parameters

To sample the parameter space, at first it would seem most straightforward to vary the five model parameters in equation 3 (ϕ_{min}, p, f, a , and δ). However, if this is done then the tensor-to-scalar ratio r *cannot* be varied freely, but must be set according to the relation $r = 16\epsilon_V$, where $\epsilon_V = \frac{1}{2} \left(\frac{V_{,\phi}}{V} \right)^2$ is the first potential slow-roll parameter. Since r is the more interesting observable, we prefer to vary r freely while using this constraint to fix the value of ϕ_{min} as a function of the other parameters. Substituting the potential from equation 3, we find our constraint equation,

$$\frac{p}{\phi_{min}}(1 - a \sin \delta) - \frac{a}{f} \cos \delta = \sqrt{\frac{r}{8}}. \quad (10)$$

While a gives the amplitude of the oscillation in the potential, the power spectrum amplitude is more directly observable in the CMB. By expanding equation 4 in the amplitude a , it can be seen that the scalar power spectrum amplitude is $\approx 2 \frac{a \phi_{min}}{nf}$ (this dominates over a since ϕ_{min} is typically of order 10 while f will be of order 0.1). Now, if a is exactly zero, then we see from equation 10 that $\frac{p}{\phi_{min}} = \sqrt{\frac{r}{8}}$. In practice, we will find that a must be small, so this will still hold to a reasonably good approximation for realistic values of a . With this in mind, we will define the (approximate) power spectrum oscillation amplitude

$$b \equiv \frac{2a}{f} \sqrt{\frac{8}{r}}. \quad (11)$$

where, again, typically $b \gg a$. It should be emphasized that our analysis will not assume that $b \ll 1$. While it is true that the power spectrum amplitude may differ somewhat from b unless $b \ll 1$, this does not preclude our using it as a parameter since the amplitude is still proportional to b . In practice, our inferred b values will satisfy $b \lesssim 1$, but b will not necessarily be very small.

Our power spectrum parameters to vary, then, are b , f , δ and p , in addition to r and the scalar amplitude A_s . Expressing equation 10 in terms of b , we find

$$\frac{p}{\phi_{min}} = \sqrt{\frac{r}{8}} \left(\frac{1 + \frac{1}{2}b \cos \delta}{1 - \frac{f}{2}\sqrt{\frac{r}{8}}b \sin \delta} \right). \quad (12)$$

From the above formula it is obvious that as $b \rightarrow 0$, we recover $\frac{p}{\phi_{min}} \approx \sqrt{\frac{r}{8}}$. Equation 12 will be used to eliminate ϕ_{min} in the power spectrum formulae (eqs. 4, 6).

B. e-folding prior

The number of e-foldings from the time the mode k_* exits the horizon to the end of inflation is constrained theoretically to lie within the approximate range 50-60. We will therefore impose a corresponding prior on the number of e-foldings, which is given by the integral

$$N = \int_0^{\phi_e} \left| \frac{V}{V_{,\phi}} \right|, \quad (13)$$

where ϕ_e denotes the scalar field value at the end of inflation determined by the solution to the equation $\epsilon_V(\phi) \approx 1$. In the case with zero amplitude ($a \approx 0$), using the expression for the potential in equation 3 one finds an expression for the number of e-foldings N_0 ,

$$\phi_{min}^2 = \frac{p}{2} (4N_0 + p). \quad (14)$$

Even for nonzero amplitude, equation 14 holds approximately true, so long as ϕ_{min} is defined in terms of the amplitude according to equation 12. Combining eqs. 12 and 14, we find an expression for the approximate number of e-foldings N_0 :

$$N_0 = p \left\{ \frac{4}{r} \left(\frac{1 - \frac{f}{2}\sqrt{\frac{r}{8}}b \sin \delta}{1 + \frac{1}{2}b \cos \delta} \right)^2 - \frac{1}{4} \right\}. \quad (15)$$

To determine the exact number of e-foldings N , the value for ϕ_e and the integral in equation 13 must be calculated numerically. In practice, the presence of the oscillations cause N to differ from that determined by equation 14 only by a small amount (typically $N > N_0$ by less than 3), although it depends on the oscillation amplitude and period. Since the integral in equation 13 is computationally expensive to calculate while varying parameters, we will instead use equation 15 for the approximate number of e-foldings to enforce a prior in N_0 during the MCMC routine.

While equation 15 typically gives a reasonable approximation to the number of e-foldings, a catastrophe can occur near the end of inflation if the oscillations dominate over the monomial term in the potential—in this case the potential may cease to become monotonic and a local minimum (false vacuum) can occur, rendering the number of e-foldings effectively infinite.[53] This occurs if either the amplitude b or the exponent p are too large; in the latter case, the slope of the monomial term becomes

very shallow before inflation ends, allowing the oscillations to dominate. To deal with this issue, we will refine the e-folding prior during post-processing by performing the numerical integral (equation 13) to find N for each point in the MCMC chain. This allows us to eliminate regions of parameter space where the number of e-foldings N becomes large or infinite.

We therefore obtain a flat prior in the number of e-foldings as follows. During the MCMC routine, we sample the parameter space with a flat prior in the approximate number of e-foldings N_0 over a liberal range from 40 to 70. During post-processing, the exact number of e-foldings N for each point in parameter space is calculated by first finding the field value at the end of inflation ϕ_e via a grid search. If a local stationary point is encountered before inflation ends, then the number of e-foldings is effectively infinite and thus the point is discarded. For the remaining points, we compute N by performing the integral in equation 13 numerically; points with N outside the canonical range from 50 to 60 are then discarded. By this method, we obtain a flat prior in the number of e-foldings N to good approximation (this will be verified in Section IV).

One last subtlety remains in implementing the e-folding prior: the parameter N_0 is entirely determined by the model parameters discussed in Section III A and thus cannot be included as a separate parameter. We therefore impose the N_0 prior by making a transformation of variables. In Section IV we will show that δ prefers to be small, and since $f\sqrt{\frac{p}{8}} \ll 1$, we can approximate equation 15 as

$$N_0 \approx \frac{4p}{r} \left(1 + \frac{1}{2}b\right)^{-2}. \quad (16)$$

Thus, to a good approximation the number of e-foldings depends only on the parameters p , r , and b , with very little dependence on f and δ . Since N_0 will not be one of our primary model parameters, we enforce the e-folding prior by starting with N_0 as a parameter (with a given prior), then making the transformation from N_0 to b . We then derive our prior in b from the prior in N_0 and the resulting Jacobian $|\partial N_0 / \partial b|$ using equation 15. As we will see in Section IV, b is fairly well-constrained (apart from a small non-Gaussian tail) and the Jacobian has only a minor effect on the inferred parameters.

C. Priors in the model parameters

Apart from the amplitude b whose prior is defined by the e-folding constraint (equation 15), we choose a flat prior in the remaining model parameters (p , r , f , δ). Here we discuss the preferred range of each parameter.

In order to sample the full range of possible phases and amplitudes, the most straightforward approach would be to vary δ over the range $(-\pi, \pi)$ and allow b to vary from 0 to some large amplitude b_{max} . However, the same can be

achieved by varying δ in the range $(-\frac{\pi}{2}, \frac{\pi}{2})$ and allowing b to have a *negative* amplitude, since the point $(b, \delta = \pi)$ is equivalent to $(-b, \delta = 0)$. We favor the latter approach, since it avoids having a bimodal posterior distribution in the phase shift δ . A negative amplitude corresponds to having a *positive* running of the spectral index near the pivot scale, and the posterior will run continuously from positive to negative b -values. We choose a liberal $b_{max} = 2$ so our allowed range in b will therefore be $(-2, 2)$.

We can get a sense of the desired range in f by considering the oscillation period in $\log \ell \sim \log k + \text{const.}$ Combining eqs. 4 and 8, we find that the axion decay constant f is related to the period in $\log k$ by

$$f = \frac{\ln 10}{2\pi} \left(\frac{p}{\phi_{min}} \right) P_{\log k} \quad (17)$$

Using equation 14, this becomes

$$f \approx 0.26 P_{\log k} \sqrt{\frac{p}{N_0}}. \quad (18)$$

From the largest scales down to the Silk damping tail probed by ACT/SPT, observable primary anisotropies in the CMB span a range of $\ell \approx 2 - 3500$, corresponding to $\Delta \log \ell \approx 3$. At higher multipoles, there is no strong preference for negative running in either the Planck or ACT data (although SPT does prefer negative running at high ℓ to some extent—see [54]). We therefore expect at worst a mild suppression of power at high ℓ . In order to have suppression at low ℓ without an equally large suppression at higher ℓ , the entire range of multipoles $\Delta \log \ell \sim 3$ should fit less than roughly three quarters of a full period; in other words, the period $P_{\log \ell} = P_{\log k}$ should be roughly greater than ~ 4 . For p running from 0.5 to 4 and 50-60 e-foldings, we find from equation 18 that f should be larger than ≈ 0.1 .

On the other hand, a large period of oscillation would require a large amplitude to achieve the desired suppression of power at low ℓ , and this would approach the constant running limit (over CMB scales) that has been considered before. For $f \approx 1$, the oscillation period is at least several times larger than the relevant ℓ -range (unless $p \ll 1$) such that the running is effectively constant in this regime. As we will see in Section IV however, regions of parameter space with a large period $P_{\log k}$ are forbidden by the e-folding constraint unless the amplitude $b \lesssim 1$, and hence the running of the spectral index is small (see also appendix III). Thus, allowed solutions with $f \gtrsim 1$ tend to have small running and cannot fit the low- ℓ power spectrum significantly better than the usual power-law power spectrum model. In Section IV we will show that this expectation is correct and thus the constraints do not change significantly when f is extended up to 2 (see Figure 4). We therefore choose the fiducial range to be $f \in (0.1, 1)$.

Given that the period $P_{\log k}$ is more directly observable in the power spectrum than f is, it might be tempting

to use $P_{\log k}$ as our model parameter instead of f . We choose not to do this, for two reasons. On the theory side, a super-Planckian axion decay constant $f \gtrsim 1$ is difficult to implement in the underlying string theory; in fact $f \ll 1$ seems necessary to embed the corresponding model, although it may be possible for multiple axions to combine to produce a larger effective axion decay constant [55, 56]. Given these difficulties, we prefer to impose an upper limit $f \leq f_{\max}$, with $f_{\max} = 1$ being the fiducial value. This does *not* translate to a clean upper bound in $P_{\log k}$, however, since from equation 18 it is possible to have $f \gtrsim 1$ even if $P_{\log k}$ is relatively small provided p is large enough.

The second reason for choosing f as our parameter instead of $P_{\log k}$ is more subtle. As we will see in Section IV, the posterior distribution is multi-modal and there exist regions of parameter space which fit the high- ℓ likelihood at the expense of lower ℓ . For example, if p is made small, by equation 18 the period becomes large unless f is also small. However in the latter case, we find another mode emerges which has negligible running and only fits the high- ℓ likelihood (for discussion see appendix III). Since we are primarily interested in improving the fit at low as well as high multipoles, by imposing a lower bound $f_{\min} = 0.1$ we avoid this spurious mode entirely. Thus, our fiducial range in f is (0.1, 1). In Section IV C we will consider the effect of varying the prior of f , including the allowed range (see Figure 4).

IV. RESULTS

A. Constraints on the oscillation parameters b, δ, f

After sampling the parameter space with the data and priors discussed in Section III, we display marginal posterior probability distributions in the model parameters in a “triangle plot” in Figure 1. Starting with the one-dimensional posteriors, it is clear that the oscillation amplitude b and phase shift δ have distributions that are well peaked, albeit with significant non-Gaussian tails. By contrast, the axion decay constant f is very poorly constrained, with a mild preference for $f \approx 0.5$ but largely prior-dominated.

Before proceeding, is important to verify that the constraints on the spectral index n_s and running α_* at the pivot scale $k_* = 0.05 \text{ Mpc}^{-1}$ are consistent with those obtained from the usual power-law spectrum and constant running models. To this end, analytic expressions for n_s and α_* are derived in Appendix II and given in equations 31 and 33, respectively. Using these formulas we plot the corresponding derived posteriors in Figures 2(a) and 2(b). The probability is peaked around $n_s \approx 0.96$, entirely consistent with the base Planck model; likewise, the allowed running α_* lies in the approximate range $(-0.03, 0)$, consistent with the usual constant running model. In Figure 2(c) we plot a posterior in the number of e-foldings N (calculated from eq. 13), which shows N

to be entirely dominated by the chosen flat prior in the fiducial range 50-60. Finally, in Figure 2(d) we plot a posterior in the oscillation period $P_{\log k}$ (using eq. 17), which is well-peaked in contrast to the constraint on f .

Returning to the model parameter constraints, the structure of the posterior distribution can be better understood from the joint 2-dimensional posteriors in Figure 1. In the f vs. b plot, one sees a mild correlation in f and b , particularly for $f \lesssim 0.6$. This can be understood in terms of the running of the spectral index: if the period (corresponding to f) is increased, one must also increase the amplitude b to keep the running α_* constant (this can also be seen in the formula for α_* in equation 33). The correlation is not very tight, however, because the required running depends on r : the greater the tensor contribution at low ℓ , the greater the (negative) running must be to adequately suppress a corresponding amount of low- ℓ scalar power.

In Figure 1, we can see in the b vs. δ plot the distribution is obviously multi-modal, where outside the maximum probability region there are three separate “wings”. Two of the wings run off to large or small δ , while the amplitude b is very small or even negative. The third wing runs off to high amplitudes b . The multi-modal structure is discussed in further detail in appendix III; here we will simply note that in each of these wings, f takes on either very small or large values (as expected since b and f are correlated). If one stays within the high probability region, f is somewhat better constrained than the f -posterior in Figure 1 would suggest.

B. Best-fit model

Next, to find the best-fit point by minimizing the likelihood requires some care, because the exact number of e-foldings N is computationally too expensive to perform during the minimization procedure. We therefore obtain the best-fit point by factoring in a prior in N_0 (given by eq. 15), where in this case our prior is chosen to be strongly peaked (a Gaussian with dispersion $\sigma_{N_0}=0.5$) about a particular value $N_{0,i}$. This procedure is performed over a grid of 16 values $N_{0,i}$ regularly spaced from 45 to 65. After finding each best-fit point, the exact number of e-foldings N is calculated for each; the best-fit point with the highest likelihood value that remains within the range $N \in [50, 60]$ is chosen as the global best-fit point.

The resulting best-fit parameter values are given in the first row of Table I. Errors are given for the inflation parameters r, p, b and δ , determined by the 16% and 84% percentiles of the posterior probability distribution in each parameter.

In Figure 3 we plot the CMB angular power spectrum over the multipole range $2 \leq \ell \leq 2500$ for the best-fit axion monodromy model (dark line) compared to the base Planck model (red line; defined as the Λ CDM model with a power-law spectrum, i.e. zero running). Note that

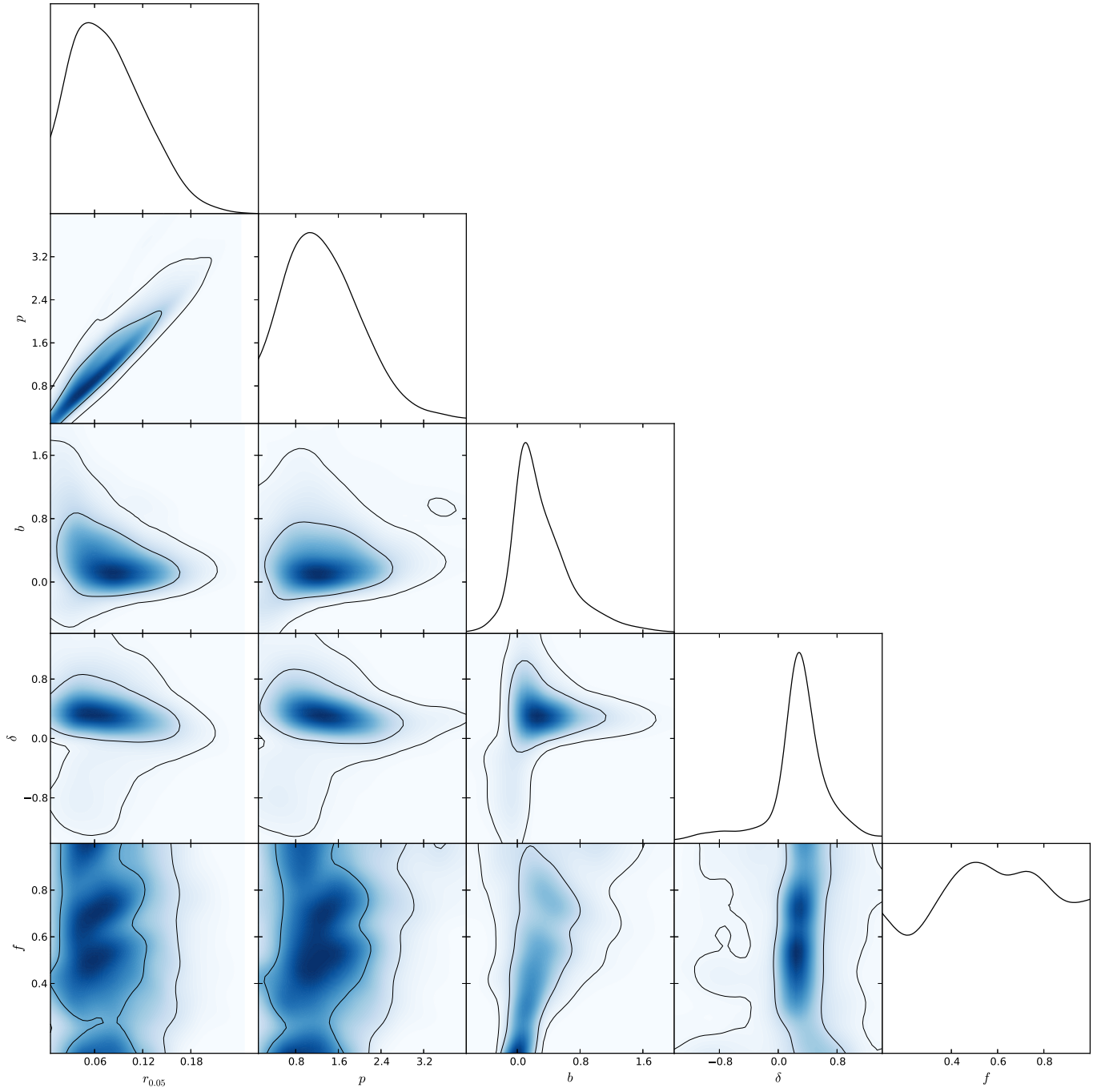


FIG. 1. Posteriors in the inflation model parameters (p, b, δ, f) discussed in Section III A and the tensor-to-scalar ratio r at the pivot scale $k_* = 0.05 \text{ Mpc}^{-1}$. Contours enclose 68% and 95% of the total probability in each joint posterior. Note the small probability for $r = 0.2$, which lies just outside the 99% confidence region. The correlation between the exponent p and r is a consequence of the e-folding constraint (eq. 15). In contrast to the amplitude b and phase shift δ which have peaked distributions, the axion decay constant f is poorly constrained although it exhibits a mild correlation with b . Finally, note that the distributions in r and p exhibit very little dependence on f , and are thus fairly insensitive to the assumed prior in f .

for $\ell \gtrsim 30$, the two models are indistinguishable, while at lower multipoles the axion monodromy power is suppressed by up to $\approx 20\%$ compared to the base Planck model. This suppression is a consequence of the running spectral index coming from the sinusoidal term in the

potential (eq. 3).

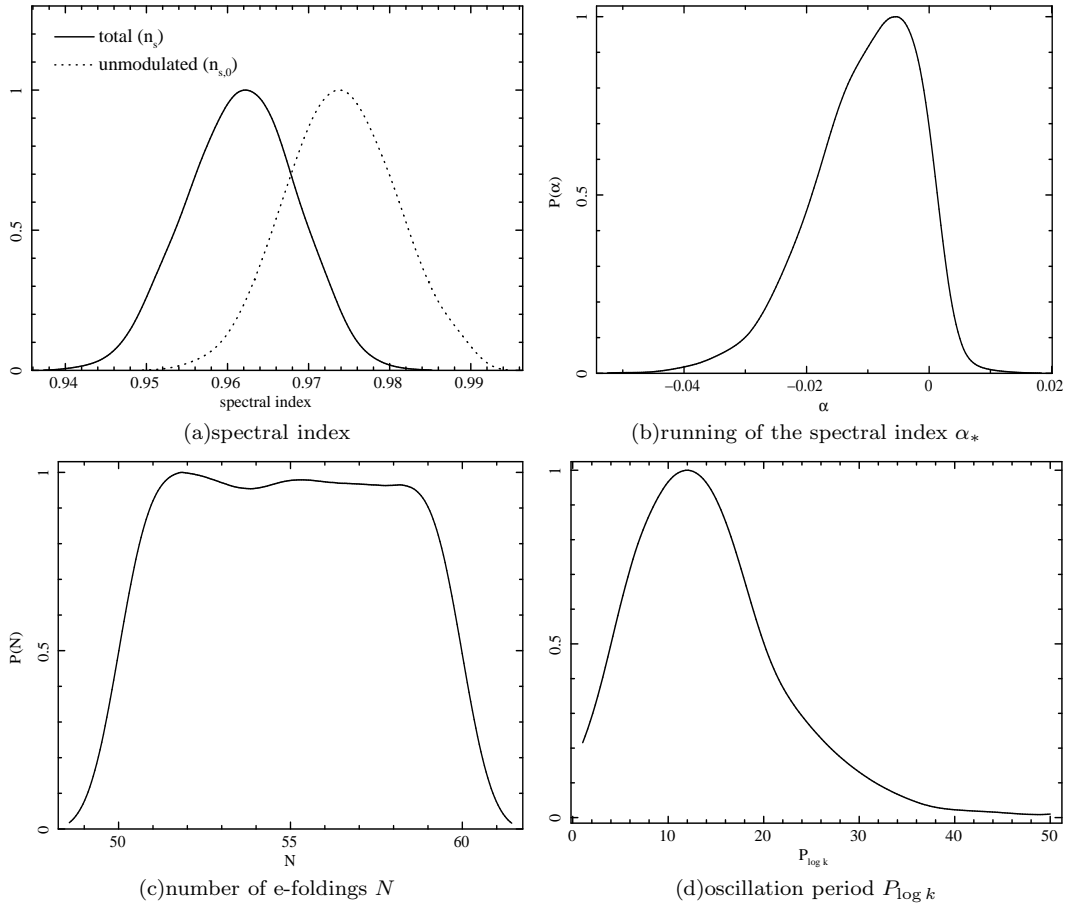


FIG. 2. Marginal posterior probability in four derived parameters: (a) The spectral index at the pivot scale $k_* = 10 \text{ Mpc}^{-1}$. The total spectral index n_s (solid line) is plotted together with the unmodulated spectral index $n_{s,0}$ (dashed line) which excludes the oscillatory part. (b) Running of the spectral index α_* at the pivot scale. (c) The number of e-foldings N from the time that mode k_* left the horizon, to the end of inflation. (d) Period of oscillation in the power spectrum in terms of $\log_{10} k$.

	r	p	b	δ	f	$10^9 A_s$	H_0	Ω_m	Ω_b	τ	N	α_*	n_s	$n_{s,0}$
Best-fit	$0.07^{+0.05}_{-0.04}$	$1.55^{+0.56}_{-0.92}$	$0.44^{+0.24}_{-0.45}$	$0.30^{+0.32}_{-0.31}$	0.53	2.209	67.9	0.307	0.048	0.093	58.5	-0.014	0.959	0.979
$r = 0.13$	0.13	2.37	0.41	0.20	0.58	2.216	68.2	0.304	0.048	0.094	59.9	-0.020	0.960	0.970
$r = 0.19$	0.19	3.08	0.31	0.12	0.54	2.210	68.3	0.303	0.048	0.093	51.9	-0.025	0.960	0.961

TABLE I. Best-fit axion monodromy models

C. Constraints on the tensor-to-scalar ratio r and potential parameter p

Marginal posteriors in p and r are shown in Figure 1. Strikingly, the BICEP2 result $r = 0.2$ lies just outside the 99% confidence region, while the highest probability r -value is ≈ 0.06 , similar to the best-fit value $r \approx 0.07$ (Table I). Likewise, $p > 3$ lies outside the 95% confidence region.

From the e-folding constraint given by equation 15 it is apparent that the exponent p and the tensor-to-scalar ratio r should be strongly correlated, provided the amplitude b does not get too large. The joint posterior in

p and r shown in Figure 1 shows that this is indeed the case; a high r -value is correlated with a large exponent p . The width of the posterior around this correlation is determined by the allowed number of e-foldings, and also the range of b -values.

Since the axion decay constant f is poorly constrained, it is important to check whether these results are sensitive to the chosen prior on f . From the r vs. f and p vs. f plot in Figure 1, one can anticipate that the inferred probability of r and p are weakly dependent on f , since essentially no correlation with f is seen for either parameter. To verify this, in Figure 4 we plot posterior distributions in r for four different assumed priors in f :

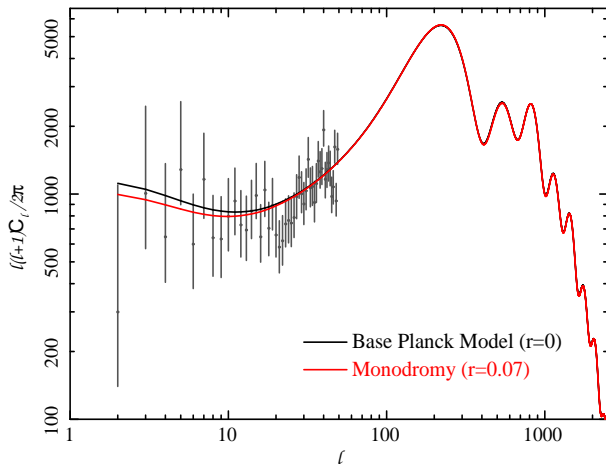


FIG. 3. Best-fit TT angular power spectrum for the base Planck model with zero running (dark line), and axion monodromy model (red line). Error bars are shown for $2 \leq \ell < 50$. These best-fit spectra are determined using a combination of Planck+lensing+WP+high ℓ data.

a flat prior in the fiducial range $f \in (0.1, 1)$ (solid line), a flat prior in the ranges $f \in (0.05, 0.25)$ (dashed line) and $f \in (0.1, 2)$ (dot-dashed line), and finally a log prior in f over the fiducial range. As can be seen in the figure, the inferred probability in r (and likewise p) is quite robust to changes in the assumed f prior. Since the other model parameters (b, δ) are fairly well-constrained, our result for r is thus quite insensitive to the priors in the axion model parameters.

Why does the data prefer a small (but nonzero) r value? Generally, higher r means a larger tensor contribution to TT anisotropies at low ℓ . This exacerbates the deficit in power at low ℓ , and thus requires a higher negative running to make up for it. However, the likelihood at high ℓ shows no strong preference for negative running [1], and the fit worsens at high ℓ as the running increases. Thus, the fit generally worsens as r is increased toward large values. This is not the whole story however, because r is significantly more constrained in the axion monodromy model compared to the usual constant running model, for which $r < 0.26$ at the 95% confidence level [1]. This can be seen in Figure 5, where we plot a joint posterior in r vs. n_s for the constant running model (red) and the axion monodromy model (blue). While both models prefer a similar n_s , r is more constrained in axion monodromy: the point $r = 0.2$, $n_s \approx 0.96$ lies outside the 95% CL contour for axion monodromy, while it is well within the same contour for the constant running model.

Additionally, note from Figure 5 that r prefers to be zero in the constant running model, whereas the axion monodromy model has its peak probability near the best-fit $r \approx 0.07$. The reason why axion monodromy does not prefer $r = 0$ is simple: one can see from equations 31 and 33 that n_s and α_* are dependent on r , whereas in the constant running model, these parameters can be cho-

sen independently of r . The best-fit constant running model has a zero tensor contribution ($r = 0$) and running $\alpha \approx -0.012$. This shows that having a nonzero tensor contribution cannot be entirely made up for by negative running—even if running is allowed, the likelihood itself still prefers $r = 0$. In the axion monodromy model however, this solution is impossible, since the running is proportional to r according to equation 32. Thus, $r \approx 0$ would necessarily imply negligible running, and likewise the spectral index would be a poor fit. Instead, the best-fit axion monodromy model settles for an r -value (≈ 0.07) which is high enough to give the necessary running and spectral index, but low enough that the tensor contribution doesn't spoil the fit too much. This compromise necessarily results in a slightly worse fit at low ℓ compared to the constant running model.

We still need to understand why axion monodromy is more restrictive at the large r end compared to constant running. To investigate this, Table II shows the change in χ^2 for four models compared to the base Planck model with zero running ($\alpha = 0$). The four models are: the best-fit axion monodromy model (for which $r \approx 0.07$), the best-fit constant running model (for which $r = 0$), and the same two best-fit models when r is fixed to 0.19 rather than varied. Note that while the total χ^2 is decreased by a similar amount in both best-fit models, χ^2 is increased in the $r = 0.19$ models, with the axion monodromy $r = 0.19$ giving the worst fit. To understand why this is the case, we must break this down into the various likelihoods representing different multipole ranges.

Starting with the high- ℓ likelihoods, note that while the fit to the CAMSpec likelihood (comprising the majority of multipoles in the Planck data) is barely affected for the general best-fit models, the fit is dramatically worsened in the models for which $r = 0.19$. This is a consequence of the aforementioned high running required to fit the low- ℓ likelihood well for large r . The same is true (though not as dramatically) for the ACT/SPT likelihoods. For the Commander likelihood (low ℓ), on the other hand, all models show an improved fit, but the constant running model gives a significantly better fit; this disparity is greater for the $r = 0.19$ models. In the following section we investigate the nature of this disparity more closely and why it worsens with large r .

D. Comparison to constant running model at low multipoles

In Figure 6 we plot the best-fit angular TT power spectrum for $2 \lesssim \ell \lesssim 50$ for the base Planck model with zero running (red solid line), axion monodromy (blue dashed line), and constant running (magenta dotted line) models. The black error bars give the errors in each individual multipole. Note that while both models achieve similar reduction in power at very low ℓ , the axion monodromy model achieves very little reduction in power for $\ell \gtrsim 25$. The deficit in power in the data (points with error bars)

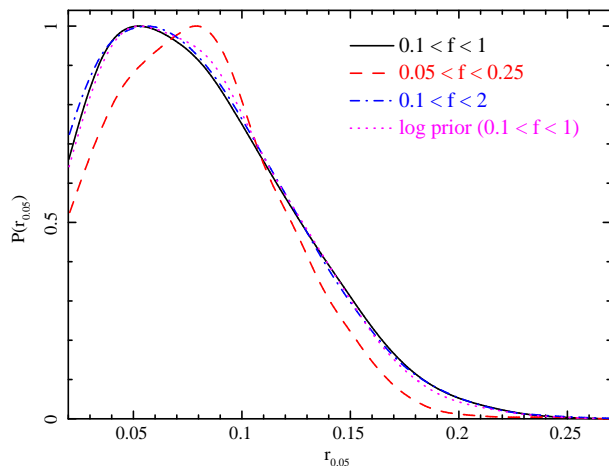


FIG. 4. Posterior distribution in the tensor-to-scalar ratio r for four different priors in f : three uniform priors covering different ranges in f (solid, dashed, dot-dashed lines) and a log prior in f (dotted line). The inferred probability in r (and likewise p) is quite robust since it is insensitive to the assumed prior in f , while the other model parameters (b, δ) are fairly well-constrained.

is most noticeable over the range $20 \lesssim \ell \lesssim 30$, and in this range the constant running model is a significantly better fit.

Why does the constant running model fit the low- ℓ likelihood better? There are a few reasons. As discussed in the previous section, axion monodromy does not have the freedom to choose r and α , n_s independently. In particular, $r = 0$ is disfavored because it produces a negligibly small running and incorrect spectral index. Unfortunately however, a nonzero r slightly worsens the fit at low ℓ by including a tensor contribution.

Even if r is fixed to the same value in both models, however, the constant running model achieves a better fit at low ℓ . In Figure 7 we plot the primordial scalar power spectrum for axion monodromy models (solid lines) and constant running models (dashed lines), with r fixed to 0.07 (dark lines) and 0.19 (red lines) in each case. For either r -value, we find that the constant running model achieves greater suppression of power at low k (and hence, low ℓ). For the $r = 0.19$ case we need a much greater suppression compared to $r = 0.07$ to make up for the additional tensor power. However, it is evident that the axion monodromy $r = 0.19$ model achieves much less suppression at low k compared to the corresponding constant running model. This occurs because the $r = 0.19$ model has a relatively high exponent $p \approx 3$, as required by the e-folding constraint (eq. 15). As a result, the relatively large monomial term “softens” the magnitude of the running at low k . The higher the p (and hence, r), the more the running is mitigated at low k . This is the principle reason why axion monodromy provides a worse fit compared to constant running when r is large.

Even for relatively small amplitudes, as r is increased to ever higher values, p must also be high (eq. 15). How-

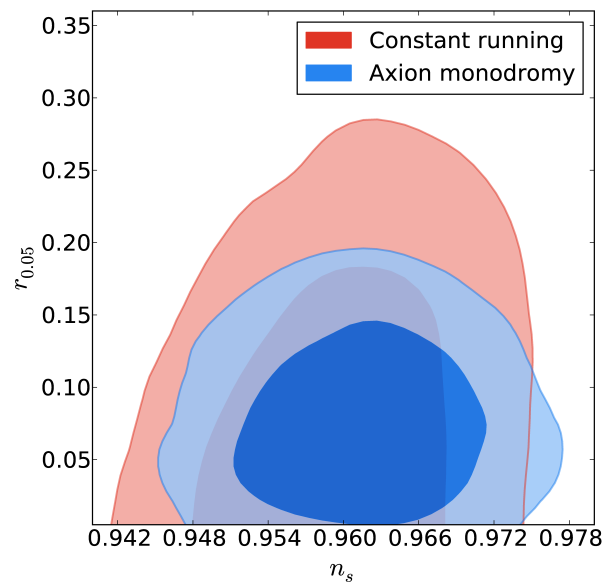


FIG. 5. Joint posteriors in the spectral index n_s and the tensor-to-scalar ratio r , both evaluated at the pivot scale $k = 0.05 \text{ Mpc}^{-1}$. Constraints (68% and 95%) are shown for the constant running model (red) and axion monodromy model (blue). While the constant running model prefers $r = 0$, in axion monodromy the highest probability occurs for $r \approx 0.07$ because a nonzero r -value is required to fit the running and spectral index well. Note also the constant running model allows for a higher r , primarily because it is not subject to the e-folding constraint.

ever, sufficiently high p solutions lead to a very large or infinite number of e-foldings due to the appearance of a local minimum in the potential from the oscillation (unless the amplitude b is very close to zero), as discussed in Section III. This effect excludes nearly all regions of parameter space where $p > 4$, and a large fraction of parameter space where $p > 3$. Thus, the direct effect of the oscillations on the total number of e-foldings N restricts the allowed parameter space for large r even further.

E. Comparison to constant running model at high multipoles

As Table II shows, axion monodromy actually improves the fit for the ACT/SPT likelihood (high ℓ) compared to the constant running model. We can understand why this is the case from the analytic expression for n_s at the pivot scale $k = k_*$ (derived in appendix II). The spectral index n_s has two contributions,

$$n_s \approx n_{s,0} + \Delta n_s, \quad (19)$$

where $n_{s,0}$ is the contribution from the monomial term in the potential (i.e. the unmodulated spectral index), while Δn_s is the contribution from the sinusoidal term. By

	Best-fit model	Best-fit constant α	Best-fit ($r = 0.19$)	Best-fit constant α ($r = 0.19$)
Lowlike (WMAP Pol.)	-0.5	0.1	-1.1	-1.0
Lensing	0.8	0.9	1.0	1.2
Commander ($2 \leq \ell < 50$)	-1.8	-2.5	-0.6	-1.7
CAMspec ($50 \lesssim \ell \lesssim 2500$)	0.2	0.1	1.5	2.2
ACT/SPT ($600 \lesssim \ell \lesssim 3500$)	-0.5	-0.2	0.2	-0.1
Total	-1.7	-1.6	1.0	0.6

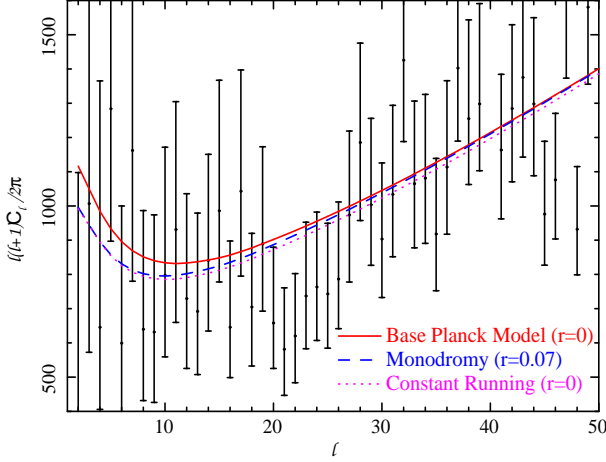
TABLE II. $\Delta\chi^2$ compared to best-fit base Planck model (with $\alpha = 0$)

FIG. 6. Best-fit CMB TT angular power spectrum at low multipoles for the base Planck model with zero running (red solid line), the axion monodromy model (blue dashed line), and the constant running model (magenta dotted line). Although axion monodromy and the constant running model exhibit a similar suppression of power at the lowest multipoles, the latter has a greater suppression of power in the range $10 < \ell < 50$, partly because there is no tensor contribution ($r = 0$) and because it is not subject to the e-folding constraint.

comparing to equation 31 and substituting the approximate e-folding constraint (equation 16), we find that

$$n_{s,0} = 1 - \frac{r}{8} \left(1 + \frac{2}{p} \right) \quad (20)$$

$$\approx 1 - \frac{r}{8} - \frac{1}{N_0} \quad (21)$$

while for the oscillating term we have

$$\Delta n_s \approx -\frac{b}{f} \sqrt{\frac{r}{8}} \sin \delta + \frac{r}{4} \left(1 - \frac{1}{p} \right) b \cos \delta. \quad (22)$$

Generally, the first term in equation 22 dominates over the second term provided that the phase shift δ is not very small or negative, since b and f are of a similar order. Given that the data prefers a positive δ , we may therefore expect $\Delta n_s < 0$, and therefore $n_{s,0}$ should be at least slightly higher than n_s .

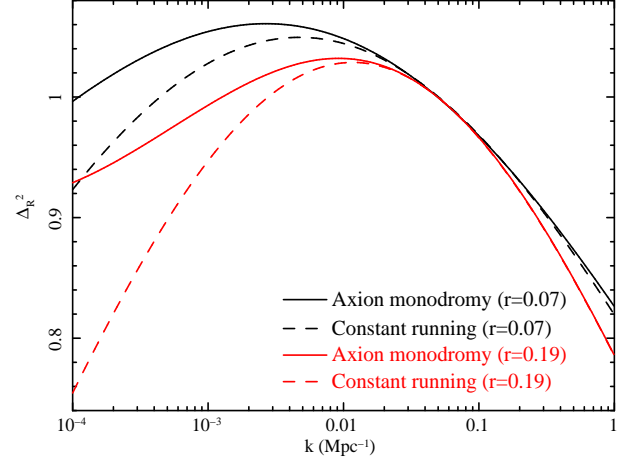


FIG. 7. Best-fit scalar power spectrum for axion monodromy (solid lines) vs. constant running models (dashed lines), all normalized to 1 at $k = 0.05 \text{ Mpc}^{-1}$ for the sake of comparison. For $r = 0.07$ (dark lines), axion monodromy achieves only slightly less suppression at low k (corresponding to low ℓ) compared to the constant running model. For $r = 0.19$ (red lines) however, axion monodromy has milder negative running at low k and thus cannot achieve enough suppression to make up for the tensor contribution at low ℓ . This is a consequence of the e-folding constraint, which enforces a large exponent p and limits the oscillation amplitude b of the power spectrum.

In Figure 2 we plot derived posteriors in n_s and $n_{s,0}$ using equations 19, 20, and 22. The constraint in n_s ($n_s \approx 0.96$) is quite similar to that obtained in the base Planck model [1], which is expected since it is well constrained by the data. By contrast, the best-fit value for $n_{s,0}$ is ≈ 0.98 , which is strikingly high. This follows from equation 21: since the number of e-foldings is restricted to the approximate range 50-60, the lower the r -value (and the corresponding p -value), the higher $n_{s,0}$ must be. Since the best-fit model has $r \approx 0.07$, this relatively low r accounts for why $n_{s,0}$ is so high.

The question remains, why does the data at high ℓ prefer such a high $n_{s,0}$? According to Table I, the best-fit axion monodromy model shows the greatest improvement in χ^2 for the ACT/SPT likelihood. We have seen that at high multipoles, the data shows no strong preference for a running spectral index, with $n_s \approx 0.96$ preferred even at higher ℓ . This is also largely true when the high- ℓ

data from ACT/SPT is factored in, although it should be noted that ACT and SPT are in tension here; SPT prefers negative running while ACT does not [54]. Thus, while negative running improves the fit at low ℓ , it worsens the fit at the high- ℓ end unless the running is relatively small. In the best-fit axion monodromy model however, the running diminishes in magnitude as one proceeds to smaller scales (higher ℓ), and hence the spectral index Δn_s is slightly higher at high ℓ . This is why the relatively high value of $n_{s,0}$ is preferred, since it allows a smaller running at high ℓ .

A similar conclusion regarding a high $n_{s,0}$ was reached by [57], in which their best-fit solution has $n_{s,0} \approx 1$ for the case where they incorporate the BICEP2 constraint $r \sim 0.2$. However, in their work the full inflation model is not used; the spectral index $n_{s,0}$ is chosen completely independently of r , and no e-folding constraint is applied. Equation 21 shows that for $r = 0.2$, $n_{s,0}$ cannot be larger than 0.975, and this upper limit is only possible in the limit $N_0 \rightarrow \infty$. From this it can be concluded that while the high- ℓ data prefers a high $n_{s,0}$, this is forbidden by the e-folding constraint unless r is relatively small. Indeed, the $r = 0.19$ axion monodromy model in Table I has $n_s \approx n_{s,0}$ and the fit is actually slightly worsened at the high- ℓ end.

F. Can the number of e-foldings be amended to allow a higher tensor-to-scalar ratio?

While the e-folding constraint disfavors high r -values, one might question whether our model might be amended so that the number of e-foldings satisfies the constraints even for high r . In particular, our model assumes that the oscillation amplitude of the potential a remains constant all the way to the end of inflation. For this reason, solutions with sufficiently large r (and hence p) tend to dramatically inflate the number of e-foldings near the end of inflation, excluding these solutions. However, if the oscillation amplitude were to diminish before the end of inflation, these high- r solutions could in principle still satisfy the e-folding constraint.

Furthermore, from the point of view of the underlying microphysics in the axion monodromy model, there is no necessary reason to expect the modulation parameters to remain constant during inflation. On the contrary, these parameters are determined by the values of dynamical moduli fields (e.g. related to the size and shape of the compactified extra-dimensional manifold) which might well evolve with time as inflation progresses. Of course, if the oscillation amplitude were to increase with time, the resulting constraint on r would be even stricter. However it is reasonable to consider the case where the oscillation amplitude dies out well before the end of inflation. To consider this, we note that the approximate e-folding parameter N_0 (eq. 15) gives the number of e-foldings without the direct contribution of the sinusoidal oscillation; further, by far the largest contribution of the

oscillation to the number of e-foldings N occurs at scales smaller than that probed by the CMB. This is particularly the case for solutions with high p (and hence, high r) values. Therefore, we can use N_0 as a conservative estimate of the number of e-foldings in the case where the oscillations die out at scales smaller than that of the CMB.

With this in mind, we apply a flat prior in the number of e-foldings $N_0 \in (50, 60)$ to approximate the case where the amplitude dies out at smaller scales. The resulting constraint on r is shown in Figure 8, plotted next to the fiducial result which uses a flat prior in N . As can be seen, there is greater probability for large r , but even in this case $r > 0.2$ is disfavored at the 95% CL, for the same reason outlined in the previous section: the large amplitude (and low p) required to achieve enough suppression of power at low ℓ is forbidden by the e-folding constraint (eq. 15).

In spite of the above difficulties, certainly there are ways to decrease the number of e-foldings to allow for higher r . For example, if the exponent p switches to a lower value at small scales, inflation would end more quickly and the number of e-foldings would be decreased. From the standpoint of single-field inflation this seems highly unnatural, although such a transition might naturally occur if multiple fields contribute to the effective inflaton potential (for an example of this scenario see [58]). From the Occam's razor point of view however, it seems more likely that the tensor-to-scalar ratio is simply smaller than the BICEP2 result suggests, particularly given the concerns about contamination by dust polarization.

V. CAN AXION MONODROMY ALLEVIATE THE SMALL-SCALE PROBLEMS OF Λ CDM?

A. Implications for dwarf galaxy formation

We have shown that oscillation in the power spectrum can allow for a significant tensor-to-scalar-ratio, alleviating the tension with the standard Λ CDM + power-law spectrum model at very large scales, while still satisfying the e-folding requirement. However, there are also apparent departures from Λ CDM for small-scale structure which can be mitigated by the same mechanism. First, there is the “missing satellites” problem, which refers to the fact that cosmological dissipationless N-body simulations predict a much larger number of dwarf satellite galaxies around the Milky Way than are actually observed [59].

The second problem is that in the centers of baryon-poor galaxies the measured dark matter density is systematically lower than predicted by dark matter-only Λ CDM simulations. In many cases, such as in low surface brightness galaxies and field dwarfs with rotation [60–65], this lower density is due to the presence of a constant dark matter density core (the so-called “core-

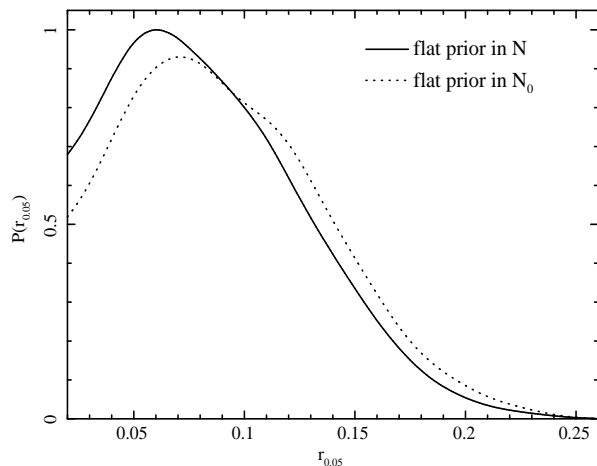


FIG. 8. Posterior in r assuming two different e-folding priors. Solid line corresponds to a flat prior in N , which gives the number of e-foldings under the assumption the oscillation in the potential continued with undiminished amplitude to the end of inflation. Dotted line corresponds to a flat prior in N_0 , which approximates the number of e-foldings in the scenario that the oscillation in the potential died out rapidly after the modes observed in the CMB exited the horizon. The former scenario gives smaller probability for large r -values, since it excludes regions of parameter space for which oscillations dominate near the end of inflation, resulting in a large or infinite number of e-foldings. However, $r = 0.2$ is located outside the 99% confidence region for the flat N prior, and outside the 95% confidence region for the flat N_0 prior and thus is disfavored in either scenario.

cusps” problem). The problem seems to extend to the lowest masses, and recent work with dwarfs in the Local Group [66–68] and further away [69] show that they are also systematically under-dense compared to simple expectations from Λ CDM, the so-called “too big to fail” problem [70].

It is possible that the simple Λ CDM expectations are incorrect and feedback from supernovae [71], reionization [72], and other effects of star formation [73] all change these expectations dramatically. There is no consensus regarding either of these two problems. Reionization will prevent small satellites from being bright enough to be observed but it is unclear whether this by itself explains the luminosity distribution of the known satellites. There are Λ CDM-based solutions for the second problem but none that alleviate the problem for all the different types of galaxies.

Here we ask whether the reduced power on small scales alleviates some of these issues. If the matter power spectrum is suppressed at the scales $k \sim 10 \text{ Mpc}^{-1}$, then this would change the formation of dwarf galaxies. Halos would collapse at lower redshift, resulting in lower central dark matter densities. This reduced power may also change the way baryonic feedback operates in low-luminosity systems. This reduction in power has been considered in light of the BICEP2 result: if one allows a constant negative running $\alpha \approx -0.02$ in the power spec-

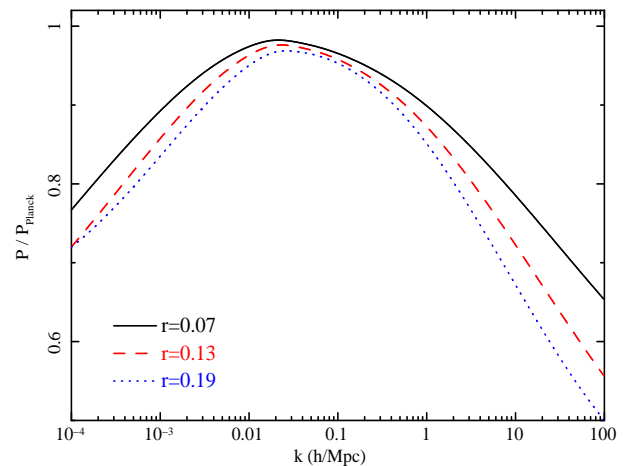


FIG. 9. The ratio of the matter power spectrum of axion monodromy compared to that of the base Planck model. Curves shown correspond to the best-fit model with the tensor-to-scalar ratio r fixed to 0.07 (dark line), 0.13 (red dashed line), and 0.19 (blue dotted line). Note there is a reduction in power of $\sim 20\text{--}35\%$ at $k = 10 \text{ Mpc}^{-1}$, roughly the scale relevant to dwarf galaxy formation.

trum to allow for the high tensor-to-scalar ratio $r = 0.2$ reported by BICEP2, [74] found using N-body simulations that the too-big-to-fail problem is significantly alleviated, although not eliminated entirely (it should be noted that baryonic effects are not included in this work). The reduction in power at dwarf scales in this “BICEP2 cosmology” [75] is approximately 40% compared to vanilla Λ CDM. However, as mentioned before, a constant running violates the e-folding constraint. Since the model considered in this work achieves significant running with the requisite amount of e-foldings, we consider here the suppression of power at small scales from our model. Again, we do not assume the BICEP2 result, but rather consider three different r -values 0.07, 0.13, and 0.19.

In Figure 9 we plot the relative power obtained by dividing the matter power spectrum in our model by the best-fit Planck power spectrum with zero running. In this plot we use the fitting functions of [76] for the transfer function. For each r -value considered, we find the best-fit solution using the same method outlined in Section IV, except with r fixed to the given value. We plot the best-fit cases where N lies in the fiducial range 50–60. Note that for each case, the difference in power compared to the base Planck model comes not only from the primordial power spectrum parameters, but also because of differences in the transfer function which is sensitive to cosmological parameters, particularly Ω_m and H_0 . For $r = 0.19$ we found that most of our solutions had $p > 3$ and thus in many cases, a local minimum formed in the potential rendering the number of e-foldings infinite. We did find one solution in the desired range, which is shown in the figure.

At dwarf scales ($k \sim 10 \text{ Mpc}^{-1}$), we find an 22% reduc-

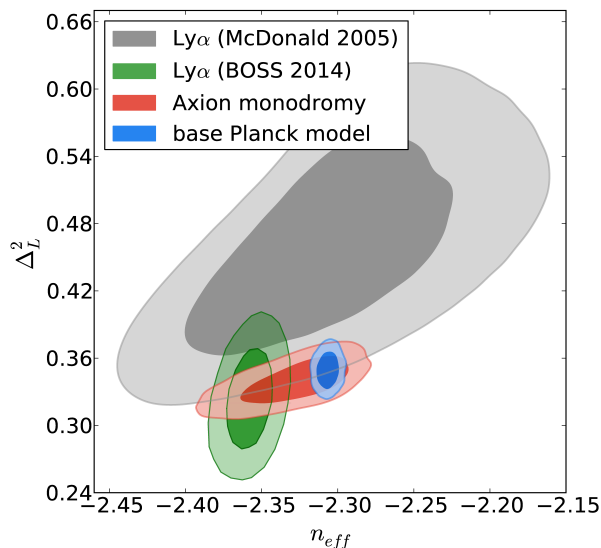


FIG. 10. Constraints on the amplitude Δ_L^2 and slope n_{eff} of the linear matter power spectrum at $z = 3$ and $k = 0.009$ $(\text{km/s})^{-1}$ where k is given in velocity space. Grey contours are from Lyman- α forest data analyzed in [77]; green contours are from Lyman- α forest data from the BOSS survey [78]; red and blue contours are from the marginalized posterior probability for the axion monodromy model and the base Planck model (with zero running), respectively. Note that the more recent BOSS data are in tension with the base Planck model at greater than 2-sigma, whereas they are consistent with axion monodromy to within 1-sigma.

tion in power for $r = 0.07$, 28% reduction for $r = 0.13$, and 33% reduction for $r = 0.19$. This is not as drastic a reduction as in the BICEP2 cosmology ($\approx 40\%$), and recall in addition that $r = 0.19$ is disfavored by the Planck data (Figure 4). Nevertheless, a reduction in power of $\sim 20\text{-}30\%$ is entirely consistent with the data and e-folding constraint, and may be expected to alleviate the too-big-to-fail problem, particularly when considered in combination with baryonic feedback effects. One can also see from Figure 9 that even greater suppression occurs at smaller scales, and this has a bearing on the missing satellites problem. In addition, the corresponding suppression of substructure would substantially reduce the expected dark matter annihilation signal [74].

B. Comparison to power spectrum constraints from Lyman- α forest data

Aside from dwarf galaxies, another approach to detect oscillations in the power spectrum is to observe at high redshift where the matter power spectrum is close to linear at small scales. The Lyman- α forest observed in quasar spectra can reveal structure down to ~ 100 kpc scales in the approximate redshift range $z = 2$ to $z = 4$, which is in the quasi-linear regime. Past studies have found constraints on the running of the spectral

index, though not with the accuracy required for a positive detection of running of order $\alpha \approx -0.01$ [79, 80]. However, more recent quasar observations by the Baryon Oscillation Spectroscopic Survey (BOSS) collaboration [78, 81] have dramatically enlarged the available dataset of quasars with measured Ly α forest power spectra. In Figure 10 we plot inferred probability contours in the amplitude Δ_L^2 and slope n_{eff} of the linear matter power spectrum at $z = 3$ and $k = 0.009$ km/s (corresponding to roughly 1 Mpc scales). The grey contours are from [77], while the green contours are from the more recent data from [82]. The red and blue contours are from derived posterior probability distributions for axion monodromy and the base Planck model respectively. From this figure it is evident that while the older data cannot distinguish between the two models, the new BOSS data favors axion monodromy: the base Planck model prediction for n_{eff} is more than 2-sigma apart from that of [82], while the axion monodromy prediction is consistent to within 1-sigma. More generally, this comparison indicates that a negative running in the power spectrum may now be favored by recent Ly α forest data, at the level required to accommodate a substantial tensor-to-scalar ratio.

There are a number of other approaches to probing the small-scale power spectrum besides Lyman- α forest data. Down the road, 21-cm observations are a promising approach to constraining the power spectrum at even smaller scales [83] since, unlike the Lyman- α forest, much higher redshifts can be probed where the data is not as limited at large k by thermal broadening. In the meantime, resolving the discrepancy between SPT and ACT on whether negative running is preferred in the CMB at high ℓ will be an important step. Within the coming decade, advanced redshift surveys such as WFIRST [84] and LSST may offer the best window onto the power spectrum at scales smaller than CMB [57], with the caveat that galaxy bias must be properly taken into account. In any event, a non-trivial primordial power spectrum remains an intriguing possibility for alleviating small-scale problems. This will be true particularly if alternate forms of dark matter (WDM, self-interacting) are ruled out as solutions to the too-big-to-fail and/or missing satellites problems.

VI. CONCLUSIONS

In this paper we have investigated whether inflation models can allow for a large gravitational wave background while remaining consistent with the CMB power spectrum at horizon scales (low ℓ). We have shown that axion monodromy inflation accomplishes this naturally through Planck-scale corrections, producing a gentle oscillation in the inflaton potential (eq. 2). This generates a running spectral index in the power spectrum while still achieving enough e-foldings to solve the horizon problem, in stark contrast to the usual constant running model. We have fit our model to a combination of Planck, ACT,

SPT, and WMAP low- ℓ polarization CMB data together with a prior on the number of e-foldings. The best-fit model parameters are given in Table I, while inferred probability distributions in the inflation model parameters are shown in Figure 1.

We find a best-fit tensor-to-scalar ratio $r = 0.07^{+0.05}_{-0.04}$ and thus the predicted imprint of gravitational waves on the CMB is within reach of B-mode polarization experiments. It is possible that these primordial B-modes have already been observed by the BICEP2 experiment; however if one assumes negligible foreground contamination, the BICEP2 result $r \approx 0.2$ is disfavored at the 99% confidence level. This is primarily a consequence of the e-folding constraint and the requirement to fit the spectral index at high ℓ in addition to the low- ℓ power spectrum. Since a running spectral index is the most straightforward way to reconcile the BICEP2 result with Planck, it is significant that attempting to implement running in the underlying inflation theory disfavors such a large r , as we have shown here. While it is possible that a dramatic change in the inflaton potential at small scales (high k) can alter the e-folding constraint to allow for a higher r , we find it more likely that the tensor-to-scalar ratio is simply smaller than the best-fit BICEP2 result suggests, particularly in light of the uncertainties about foreground contamination by dust polarization in the BICEP2 field.

In addition to the large gravitational wave amplitude, the (best-fit) axion monodromy model makes two corresponding predictions: first, despite the additional tensor power on horizon scales, the overall power at low multipoles is reduced as a consequence of the running spectral index, providing a better fit to the CMB power spectrum at large scales. The second prediction is that the matter power spectrum is suppressed at the scale of dwarf galaxies, and thus axion monodromy can alleviate some of the small-scale problems of Λ CDM—in particular, the too-big-to-fail and missing satellites problems. We find

that our best-fit models reduce the power at scales relevant to dwarf galaxy formation ($k \sim 10 \text{ Mpc}^{-1}$) by as much as $\sim 35\%$ depending on the assumed r -value, with the greatest reduction achieved at large r . However, 20–30% suppression is more likely, which will alleviate the too-big-to-fail problem and may solve it entirely when combined with baryonic feedback effects, as discussed in [74]. Additionally, we find that axion monodromy is preferred by recent Ly α forest data over the base Planck model without running (Figure 10).

If axion monodromy (or a similar oscillating large-field model) accounts for the reduced power at large and small scales, then the tensor-to-scalar ratio is likely to lie in the range $r \approx 0.03 - 0.12$ (68% CL) and hence the imprint of gravitational waves on the CMB will be observable by B-mode experiments in the future [85]. With this comes the tantalizing prospect of constraining physics at the Planck scale through sky surveys, as we have demonstrated here. Thus, future CMB experiments, in combination with probes of the power spectrum at small scales, may settle the issue of whether Planck-scale physics manifest in inflation can reconcile the standard Λ CDM cosmology with data at all observable scales of the Universe.

ACKNOWLEDGMENTS

QM would like to thank James Bullock, Shahab Joudaki, Jose Ceja and Shea Garrison-Kimmel for their encouragement and feedback at the beginning stages of this project.

We gratefully acknowledge a grant of computer time from XSEDE allocation TG-AST130007. MK was supported in part by NSF grant PHY-1214648.

This research was also supported, in part, by a grant of computer time from the City University of New York High Performance Computing Center under NSF Grants CNS-0855217, CNS-0958379 and ACI-1126113.

-
- [1] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, *et al.*, ArXiv e-prints (2013), arXiv:1303.5076 [astro-ph.CO].
 - [2] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and *et al.*, ArXiv e-prints (2013), arXiv:1303.5083 [astro-ph.CO].
 - [3] D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, M. R.olta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, **148**, 175 (2003), astro-ph/0302209.
 - [4] G. Hinshaw, D. N. Spergel, L. Verde, R. S. Hill, S. S. Meyer, C. Barnes, C. L. Bennett, M. Halpern, N. Jarosik, A. Kogut, E. Komatsu, M. Limon, L. Page, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, **148**, 135 (2003), astro-ph/0302217.
 - [5] B. Feng, M. Li, R.-J. Zhang, and X. Zhang, Phys. Rev. D **68**, 103511 (2003), astro-ph/0302479.
 - [6] M. Bastero-Gil, K. Freese, and L. Mersini-Houghton, Phys. Rev. D **68**, 123514 (2003), hep-ph/0306289.
 - [7] D. J. Chung, G. Shiu, and M. Trodden, Phys. Rev. D **68**, 063501 (2003), astro-ph/0305193.
 - [8] M. Kawasaki, M. Yamaguchi, and J. Yokoyama, Phys. Rev. D **68**, 023508 (2003), hep-ph/0304161.
 - [9] P. Hunt and S. Sarkar, Phys. Rev. D **76**, 123504 (2007), arXiv:0706.2443.
 - [10] H. V. Peiris, E. Komatsu, L. Verde, D. N. Spergel, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, E. Wollack, and E. L. Wright, **148**, 213 (2003), astro-ph/0302225.
 - [11] C. R. Contaldi, M. Peloso, L. Kofman, and A. Linde, **7**, 002 (2003), astro-ph/0303636.

- [12] M. J. Mortonson, C. Dvorkin, H. V. Peiris, and W. Hu, *Phys. Rev. D* **79**, 103519 (2009), arXiv:0903.4920 [astro-ph.CO].
- [13] D. K. Hazra, A. Shafieloo, G. F. Smoot, and A. A. Starobinsky, **8**, 048 (2014), arXiv:1405.2012.
- [14] L. Lello and D. Boyanovsky, **5**, 029 (2014), arXiv:1312.4251.
- [15] Y.-S. Piao, B. Feng, and X. Zhang, *Phys. Rev. D* **69**, 103520 (2004), hep-th/0310206.
- [16] J. White, Y.-l. Zhang, and M. Sasaki, *Phys. Rev. D* **90**, 083517 (2014), arXiv:1407.5816.
- [17] A. Ashoorioon, K. Dimopoulos, M. M. Sheikh-Jabbari, and G. Shiu, *Physics Letters B* **737**, 98 (2014), arXiv:1403.6099 [hep-th].
- [18] C. Dvorkin, M. Wyman, D. H. Rudd, and W. Hu, *ArXiv e-prints* (2014), arXiv:1403.8049.
- [19] L. A. Anchordoqui, *ArXiv e-prints* (2014), arXiv:1407.8105.
- [20] K. M. Smith, C. Dvorkin, L. Boyle, N. Turok, M. Halpern, G. Hinshaw, and B. Gold, *Physical Review Letters* **113**, 031301 (2014), arXiv:1404.0373.
- [21] P. A. R. Ade, R. W. Aikin, D. Barkats, *et al.*, *Physical Review Letters* **112**, 241101 (2014), arXiv:1403.3985.
- [22] R. Flauger, J. C. Hill, and D. N. Spergel, **8**, 039 (2014), arXiv:1405.7351.
- [23] M. J. Mortonson and U. Seljak, *ArXiv e-prints* (2014), arXiv:1405.5857.
- [24] Planck Collaboration, R. Adam, P. A. R. Ade, N. Aghanim, M. Arnaud, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, and *et al.*, *ArXiv e-prints* (2014), arXiv:1409.5738.
- [25] D. H. Lyth, *Physical Review Letters* **78**, 1861 (1997), hep-ph/9606387.
- [26] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and *et al.*, *ArXiv e-prints* (2013), arXiv:1303.5082 [astro-ph.CO].
- [27] R. Easther and H. V. Peiris, **9**, 010 (2006), astro-ph/0604214.
- [28] G. Ballesteros, J. A. Casas, J. R. Espinosa, R. Ruiz de Austri, and R. Trotta, **3**, 018 (2008), arXiv:0711.3436 [hep-ph].
- [29] G. Ballesteros and J. A. Casas, *ArXiv e-prints* (2014), arXiv:1406.3342.
- [30] D. K. Hazra, A. Shafieloo, G. F. Smoot, and A. A. Starobinsky, *Phys.Rev.Lett.* **113**, 071301 (2014), arXiv:1404.0360 [astro-ph.CO].
- [31] A. Ashoorioon and A. Krause, *ArXiv High Energy Physics - Theory e-prints* (2006), hep-th/0607001.
- [32] E. Silverstein and A. Westphal, *Phys. Rev. D* **78**, 106003 (2008), arXiv:0803.3085 [hep-th].
- [33] N. Kaloper and L. Sorbo, *Physical Review Letters* **102**, 121301 (2009), arXiv:0811.1989 [hep-th].
- [34] L. McAllister, E. Silverstein, and A. Westphal, *Phys. Rev. D* **82**, 046003 (2010), arXiv:0808.0706 [hep-th].
- [35] N. Kaloper, A. Lawrence, and L. Sorbo, **3**, 023 (2011), arXiv:1101.0026 [hep-th].
- [36] T. Kobayashi and F. Takahashi, **1**, 026 (2011), arXiv:1011.3988 [astro-ph.CO].
- [37] K. Freese, J. A. Frieman, and A. V. Olinto, *Physical Review Letters* **65**, 3233 (1990).
- [38] S. Dimopoulos, S. Kachru, J. McGreevy, and J. G. Wacker, **8**, 003 (2008), hep-th/0507205.
- [39] Y. Wan, S. Li, M. Li, T. Qiu, Y. Cai, and X. Zhang, *Phys. Rev. D* **90**, 023537 (2014), arXiv:1405.2784.
- [40] I. P. Neupane, *Phys. Rev. D* **90**, 123502 (2014), arXiv:1409.8647.
- [41] R. Flauger, L. McAllister, E. Pajer, A. Westphal, and G. Xu, **6**, 009 (2010), arXiv:0907.2916 [hep-th].
- [42] M. Aich, D. K. Hazra, L. Sriramkumar, and T. Souradeep, *Phys.Rev.* **D87**, 083526 (2013), arXiv:1106.2798 [astro-ph.CO].
- [43] L. McAllister, E. Silverstein, A. Westphal, and T. Wrase, *ArXiv e-prints* (2014), arXiv:1405.3652 [hep-th].
- [44] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, **215**, 203 (1992).
- [45] D. Baumann, *ArXiv e-prints* (2009), arXiv:0907.5424 [hep-th].
- [46] Note that the scale of the potential V_* will not enter into our equations directly, since we only encounter the combination V/V_* . Nevertheless, V_* is not an independent parameter but is rather determined by the normalization of Δ_ϵ^2 , with the result $V_* = \frac{3\pi^2}{2} r A_s$ in the slow roll approximation.
- [47] A. Lewis and S. Bridle, *Phys. Rev. D* **66**, 103511 (2002), astro-ph/0205436.
- [48] A. Lewis, *Phys. Rev. D* **87**, 103529 (2013), arXiv:1304.4473 [astro-ph.CO].
- [49] Planck collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and *et al.*, *ArXiv e-prints* (2013), arXiv:1303.5075 [astro-ph.CO].
- [50] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolte, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright, **208**, 19 (2013), arXiv:1212.5226 [astro-ph.CO].
- [51] K. T. Story, C. L. Reichardt, Z. Hou, *et al.*, *Astrophys. J.* **779**, 86 (2013), arXiv:1210.7231 [astro-ph.CO].
- [52] J. L. Sievers, R. A. Hlozek, M. R. Nolte, *et al.*, **10**, 060 (2013), arXiv:1301.0824 [astro-ph.CO].
- [53] To be precise, a very large number of e-foldings would occur until the field tunnels out of the local minimum via bubble nucleation. Given that relatively few e-foldings would follow this before inflation ends, this scenario would produce an open universe with a very sub-critical energy density, in gross violation of cosmological constraints (see for example [86]).
- [54] E. Di Valentino, S. Galli, M. Lattanzi, A. Melchiorri, P. Natoli, L. Pagano, and N. Said, *Phys. Rev. D* **88**, 023501 (2013), arXiv:1301.7343 [astro-ph.CO].
- [55] J. E. Kim, H. P. Nilles, and M. Peloso, **1**, 005 (2005), hep-ph/0409138.
- [56] R. Kappl, S. Krippendorf, and H. P. Nilles, *Physics Letters B* **737**, 124 (2014), arXiv:1404.7127 [hep-th].
- [57] P. D. Meerburg, *ArXiv e-prints* (2014), arXiv:1406.3243.
- [58] T. Kobayashi, O. Seto, and Y. Yamaguchi, *ArXiv e-prints* (2014), arXiv:1404.5518 [hep-ph].
- [59] A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada, *Astrophys. J.* **522**, 82 (1999), astro-ph/9901240.
- [60] R. Kuzio de Naray, S. S. McGaugh, W. J. G. de Blok, and A. Bosma, **165**, 461 (2006), astro-ph/0604576.
- [61] J. D. Simon, A. D. Bolatto, A. Leroy, L. Blitz, and E. L. Gates, *Astrophys. J.* **621**, 757 (2005), astro-ph/0412035.

- [62] J. J. Adams, J. D. Simon, M. H. Fabricius, R. C. E. van den Bosch, J. C. Barentine, R. Bender, K. Gebhardt, G. J. Hill, J. D. Murphy, R. A. Swaters, J. Thomas, and G. van de Ven, *Astrophys. J.* **789**, 63 (2014), arXiv:1405.4854.
- [63] G. Gentile, P. Salucci, U. Klein, D. Vergani, and P. Kalberla, **351**, 903 (2004), astro-ph/0403154.
- [64] P. Salucci, M. I. Wilkinson, M. G. Walker, G. F. Gilmore, E. K. Grebel, A. Koch, C. Frigerio Martins, and R. F. G. Wyse, **420**, 2034 (2012), arXiv:1111.1165 [astro-ph.CO].
- [65] S.-H. Oh, W. J. G. de Blok, E. Brinks, F. Walter, and R. C. Kennicutt, Jr., **141**, 193 (2011), arXiv:1011.0899 [astro-ph.CO].
- [66] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, **422**, 1203 (2012), arXiv:1111.2048 [astro-ph.CO].
- [67] E. J. Tollerud, M. Boylan-Kolchin, and J. S. Bullock, **440**, 3511 (2014), arXiv:1403.6469.
- [68] E. N. Kirby, J. S. Bullock, M. Boylan-Kolchin, M. Kaplinghat, and J. G. Cohen, **439**, 1015 (2014), arXiv:1401.1208.
- [69] I. Ferrero, M. G. Abadi, J. F. Navarro, L. V. Sales, and S. Gurovich, **425**, 2817 (2012), arXiv:1111.6609 [astro-ph.CO].
- [70] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, **415**, L40 (2011), arXiv:1103.0007 [astro-ph.CO].
- [71] F. Governato, A. Zolotov, A. Pontzen, C. Christensen, S. H. Oh, A. M. Brooks, T. Quinn, S. Shen, and J. Wadsley, **422**, 1231 (2012), arXiv:1202.0554 [astro-ph.CO].
- [72] J. S. Bullock, A. V. Kravtsov, and D. H. Weinberg, *Astrophys. J.* **539**, 517 (2000), astro-ph/0002214.
- [73] A. Zolotov, A. M. Brooks, B. Willman, F. Governato, A. Pontzen, C. Christensen, A. Dekel, T. Quinn, S. Shen, and J. Wadsley, *Astrophys. J.* **761**, 71 (2012), arXiv:1207.0007 [astro-ph.CO].
- [74] S. Garrison-Kimmel, S. Horiuchi, K. N. Abazajian, J. S. Bullock, and M. Kaplinghat, **444**, 961 (2014), arXiv:1405.3985.
- [75] K. N. Abazajian, G. Aslanyan, R. Easther, and L. C. Price, **8**, 053 (2014), arXiv:1403.5922.
- [76] D. J. Eisenstein and W. Hu, *Astrophys. J.* **496**, 605 (1998), astro-ph/9709112.
- [77] P. McDonald, U. Seljak, R. Cen, D. Shih, D. H. Weinberg, S. Burles, D. P. Schneider, D. J. Schlegel, N. A. Bahcall, J. W. Briggs, J. Brinkmann, M. Fukugita, Ž. Ivezić, S. Kent, and D. E. Vanden Berk, *Astrophys. J.* **635**, 761 (2005), astro-ph/0407377.
- [78] N. Palanque-Delabrouille, C. Yèche, A. Borde, J.-M. Le Goff, G. Rossi, M. Viel, É. Aubourg, S. Bailey, J. Bautista, M. Blomqvist, A. Bolton, J. S. Bolton, N. G. Busca, B. Carithers, R. A. C. Croft, K. S. Dawson, T. Delubac, A. Font-Ribera, S. Ho, D. Kirkby, K.-G. Lee, D. Margala, J. Miralda-Escudé, D. Muna, A. D. Myers, P. Noterdaeme, I. Pâris, P. Petitjean, M. M. Pieri, J. Rich, E. Rollinde, N. P. Ross, D. J. Schlegel, D. P. Schneider, A. Slosar, and D. H. Weinberg, **559**, A85 (2013), arXiv:1306.5896 [astro-ph.CO].
- [79] U. Seljak, A. Slosar, and P. McDonald, **10**, 014 (2006), astro-ph/0604335.
- [80] M. Viel, J. Weller, and M. G. Haehnelt, **355**, L23 (2004), astro-ph/0407294.
- [81] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, and et al., **145**, 10 (2013), arXiv:1208.0022 [astro-ph.CO].
- [82] N. Palanque-Delabrouille, C. Yèche, J. Lesgourgues, G. Rossi, A. Borde, M. Viel, E. Aubourg, D. Kirkby, J.-M. LeGoff, J. Rich, N. Roe, N. P. Ross, D. P. Schneider, and D. Weinberg, ArXiv e-prints (2014), arXiv:1410.7244.
- [83] H. Shimabukuro, K. Ichiki, S. Inoue, and S. Yokoyama, *Phys. Rev. D* **90**, 083003 (2014), arXiv:1403.1605.
- [84] D. Spergel, N. Gehrels, J. Breckinridge, *et al.*, ArXiv e-prints (2013), arXiv:1305.5425 [astro-ph.IM].
- [85] K. N. Abazajian, K. Arnold, J. Austerlmann, *et al.*, ArXiv e-prints (2013), arXiv:1309.5381.
- [86] M. Bucher, A. S. Goldhaber, and N. Turok, *Phys. Rev. D* **52**, 3314 (1995), hep-ph/9411206.

I. APPENDIX: IS THE SLOW ROLL APPROXIMATION VALID IN OUR MODEL?

When there are oscillating features in the power spectrum, the slow roll approximation can be violated, particularly for short-period oscillations. Thus, it is important to check that the slow roll approximation we have used here is valid in the parameter range we are interested in. For the slow roll approximation to hold, both the first slow roll parameter $\epsilon = \frac{d}{dt} \left(\frac{1}{H} \right)$ and the second slow roll parameter $\eta = \frac{1}{H} \frac{d}{dt} \epsilon$ must be small, such that $\epsilon \ll 1, \eta \ll 1$. The former condition is obviously satisfied at CMB scales, as we are many e-foldings before the end of inflation and thus $\epsilon \ll 1$. The second condition is trickier, however, because rapid oscillation can cause η to become large. To estimate η , we must solve the classical equation of motion for $\phi(t)$:

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = 0 \quad (23)$$

Since ϵ is small, we have $H \approx \sqrt{\frac{V}{3}}$. Let us switch variables to $x = t\sqrt{\frac{V_*}{3}}$, where x can be interpreted as the time in units of the Hubble time when the mode $k = k_*$ leaves the horizon. We then find

$$\frac{d^2\phi}{dx^2} + 3\sqrt{\frac{V}{V_*}} \frac{d\phi}{dx} + 3\frac{V_{,\phi}}{V_*} \approx 0. \quad (24)$$

This equation can be solved numerically, using the expression for the (scaled) potential and its derivative from equation 3. Here, since we are interested in an analytic expression for η , we will use a method similar to that of [41] and try a solution of the form $\phi = \phi_0 + a\phi_1$, where ϕ_0 is the unmodulated scalar field, ϕ_1 contains the oscillation and the amplitude a is assumed to be small (as it must be). For ϕ_0 , the slow roll approximation holds very well, as usual, so we can ignore $\ddot{\phi}_0$. Substituting this expression in to equation 24 and separating out the zeroth order and first order terms, we find for the zeroth order equation

$$\frac{d\phi_0}{dx} = \frac{p}{\phi_{min}} \left(1 - \frac{\phi_0}{\phi_{min}}\right)^{\frac{p}{2}-1}. \quad (25)$$

We will not need to solve this equation exactly for the present discussion, but we note that since we are 50-60 e-foldings from the end of inflation, we have $\phi_0 \ll \phi_{min}$ and hence, $\frac{d\phi_0}{dx} \approx \frac{p}{\phi_{min}}$.

Proceeding to the first-order equation, we switch to using ϕ_0 as the independent variable, yielding the equation

$$\frac{1}{3}\phi_1'' + \frac{1}{p}\phi_{min}\phi_1' - \frac{1}{2}\phi_1 = -\frac{\phi_{min}^2}{p^2 f} \cos\left(\frac{\phi_0}{f} + \delta\right). \quad (26)$$

where the prime denotes the derivative $\partial/\partial\phi_0$. To solve this equation we try a linear combination of sines and cosines, and arrive at the solution

$$\phi_1(x) \approx \frac{3f\left(\frac{\phi_{min}}{nf}\right)^2}{\sqrt{\left(1 + \frac{3}{2}f^2\right)^2 + \left(\frac{3\phi_{min}f}{p}\right)^2}} \cos\left(\frac{\phi_0(x)}{f} + \psi\right). \quad (27)$$

where ψ is an unimportant phase factor. We can now estimate the amplitude of the slow roll parameter $\eta = \frac{1}{H}\ddot{\phi} \approx a\frac{d^2\phi_1}{dx^2}/\frac{d\phi_0}{dx}$. Plugging in eqs. 25 and 27, we find

$$\eta_{max} \approx \frac{3a\frac{\phi_{min}}{nf}}{\sqrt{\left(1 + \frac{3}{2}f^2\right)^2 + \left(\frac{3\phi_{min}f}{p}\right)^2}}. \quad (28)$$

This is well approximated as

$$\eta_{max} \approx \frac{3b}{2 + b\cos\delta} \left(1 + 18N_0\frac{f^2}{p}\right)^{-1/2}, \quad (29)$$

where we have made use of equations 11, 12, and 14. Note that if the amplitude b is kept fixed, η can be made large (and hence, slow roll violated) if f is sufficiently small.

For the best-fit solution in Table I, we have $\eta_{max} \approx 0.04$ and thus the slow roll description holds. On the other hand, at the extreme ends of the parameter space we have chosen, η can become large; for example, setting $b = 1$, $f = 0.1$ can give η_{max} comparable to 1. Fortunately, such regions where slow roll breaks down are excluded by the data. To show this robustly, in Figure 11 we plot a derived posterior in η_{max} using equation 28, with the same priors and data as discussed in Section III. For all the points sampled, we have $\eta_{max} \lesssim 0.1$, and hence the slow roll approximation holds good for all allowed regions of parameter space. This justifies our use of the slow roll approximation in deriving the power spectra in Section II B.

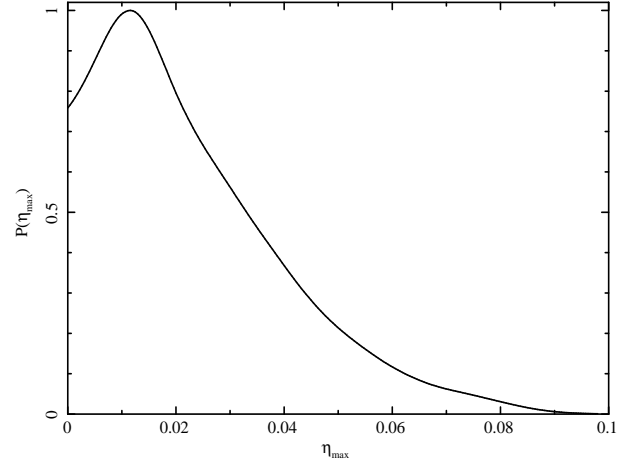


FIG. 11. Posterior in the amplitude of the slow roll parameter η . Since $\eta_{max} \ll 1$ for all allowed regions of parameter space, the slow roll expansion holds to good approximation, justifying the method used to derive the power spectra in Section II B.

II. APPENDIX: EXPRESSIONS FOR THE SPECTRAL INDEX AND RUNNING IN THE SLOW-ROLL APPROXIMATION

To relate the parameters to observable features, it is useful to have expressions for the spectral index and running at the pivot scale. The most useful expressions we derive will be equations 31 and 33, which are given in terms of the fit parameters in the usual case where $a \ll 1$, and the running is dominated by the sinusoidal term.

A. Spectral index

To derive expressions for the spectral index, we can use the relation in terms of the slow roll parameters, $n_s = 1 - 6\epsilon_V + 2\eta_V$. Using the fact that $\epsilon_V = r/16$ and $\eta_V = V_{,\phi\phi}/V$, we find using equation 3,

$$n_s = 1 - \frac{3r}{8} + 2 \left[\frac{p(p-1)}{\phi_{min}^2} (1 - a\sin\delta) - \frac{a}{f^2} \sin\delta \right] \quad (30)$$

where it is understood that we are evaluating the spectral index at the pivot scale $k = k_*$. Substituting equation 10 and keeping terms to first order in a , we find (after substituting equation 11),

$$n_s \approx 1 - \frac{r}{8} \left(1 + \frac{2}{p}\right) - \frac{b}{f} \sqrt{\frac{r}{8}} \sin\delta + \frac{r}{4} \left(1 - \frac{1}{p}\right) b \cos\delta. \quad (31)$$

Note that the oscillation is dominated by the sine term (unless δ is very small such that $\tan\delta \lesssim f\sqrt{\frac{r}{8}}$), and thus the amplitude of the spectral index oscillation is $\Delta n_s \approx \frac{2a}{f^2} \approx \frac{b}{f} \sqrt{\frac{r}{8}}$. Our constraint in n_s is plotted in

Figure 2(a) and is consistent with the result $n_s \approx 0.96$ found by applying the base Planck model.

B. Running of the spectral index

As discussed in Section IV, the best fit parameters are characterized by a substantial running of the spectral index which is dominated by the sinusoidal term. To first order, the running can be calculated from $\alpha_1 \approx -2\zeta_V$ where $\zeta_V = \frac{V_{,\phi} V_{,\phi\phi\phi}}{V^2} = \sqrt{2\epsilon_V} \frac{V_{,\phi\phi\phi}}{V}$ is the third potential slow roll parameter. From this we find (at the pivot scale k_*),

$$\alpha_{1,*} \approx -\frac{rb}{8f^2} \cos \delta. \quad (32)$$

From this formula it is evident that the running oscillates in $\phi \propto \ln k$ with the approximate amplitude $\alpha_{max} \approx \frac{rb}{8f^2}$.

Now we derive the more exact expression for the running using the formula $\alpha = -2\zeta_V + 16\epsilon_V\eta_V - 24\epsilon_V^2$. Expanding this to first order in a and up to second order in r , we obtain

$$\alpha_* \approx -\frac{rb}{8f^2} \cos \delta - \frac{4b}{f} \left(\frac{r}{8}\right)^{3/2} \sin \delta + \frac{r^2}{8} \left(1 - \frac{1}{p} - \frac{3}{4}\right). \quad (33)$$

The last term above is due to the monomial term in the potential, and is clearly of order 10^{-3} if $r \sim 0.1$; the middle term is typically of a similar order. For the best fits, we have $\alpha \sim -10^{-2}$, in which case the above

formula is accurate for up to two significant figures. Our constraint in the running α_* is plotted in Figure 2(b).

III. APPENDIX: STRUCTURE OF THE POSTERIOR DISTRIBUTION

As discussed in Section IV, the posterior distribution has a complicated multi-modal structure, with several non-Gaussian tails running away from the best-fit region. To understand this structure better, in Figure 12 we plot 3-dimensional posteriors with the axes for b and δ shown, color-coded by r , α_* , and f respectively. Note there are three separate “wings” in the distribution running away from the best-fit region, and in all of these wings the tensor-to-scalar ratio r prefers to be small (Figure 12(a)). This is because large r must be accommodated by a large negative running, and in Figure 12(b) one sees that a large negative running only occurs in the best-fit region, whereas it is quite small in the wings. Two of the wings have $b \approx 0$, which then naturally would imply a small running; however, the other wing has a large amplitude $b \gtrsim 1$. How can the running be small in this case? The answer can be seen in Figure 12(c): in most of this region f (corresponding to the period) is quite large, and thus the running is again small. Note that in all the wings f tends to take on either very small ($f \approx 0.1$) or very large values ($f \approx 1$), with intermediate values occurring mostly in the best-fit region. For this reason, if one stays within the best-fit region, the axion decay constant f is somewhat better constrained than is suggested by the posterior in Figure 1 although the errors are still large.

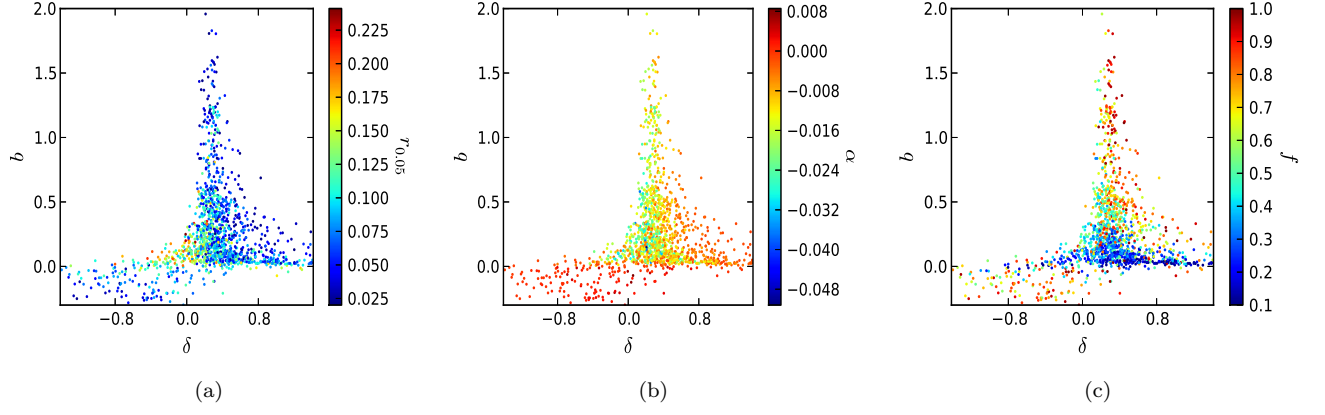


FIG. 12. Joint 3-dimensional posteriors in the amplitude b and phase shift δ , color-coded by the tensor-to-scalar ratio r , running α_* , and axion decay constant f respectively. Note that outside the best-fit region, there are tails extending to large b , or small b and extreme values of δ . In these tails, r (and correspondingly, the running α_*) prefers to be small. Likewise, f tends to take extreme values in the tails while preferring intermediate values in the best-fit region.