Axion Cosmology and the Energy Scale of Inflation

Mark P. Hertzberg\textsuperscript{1,*}, Max Tegmark\textsuperscript{1}, and Frank Wilczek\textsuperscript{1,2}

\textsuperscript{1}Dept. of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and
\textsuperscript{2}Peierls Center for Physics, Oxford University, OX13NP, UK

We survey observational constraints on the parameter space of inflation and axions and map out two allowed windows: the classic window and the inflationary anthropic window. The cosmology of the latter is particularly interesting: inflationary axion cosmology predicts the existence of isocurvature fluctuations in the CMB, with an amplitude that grows with both the energy scale of inflation and the fraction of dark matter in axions. Statistical arguments favor a substantial value for the latter, and so current bounds on isocurvature fluctuations imply tight constraints on inflation. For example, an axion Peccei-Quinn scale of $10^{16}$ GeV excludes any inflation model with energy scale $> 3.8 \times 10^{14}$ GeV ($r > 2 \times 10^{-9}$) at 95\% confidence, and so implies negligible gravitational waves from inflation, but suggest appreciable isocurvature fluctuations.

\section{I. INTRODUCTION}

Early universe inflation and the QCD axion provide explanations for otherwise mysterious features of the universe. Here we argue that assuming both at once leads to very significant constraints on their central parameters, and to highly falsifiable predictions.

\subsection{A. Energy Scale of Inflation}

Inflation is the leading paradigm for early universe phenomenology \cite{214,215}. Its mechanism and the values of its central parameters are unknown, however. One central parameter is the energy scale of inflation $E_I$, defined as the fourth root of the inflationary potential energy density, evaluated when the modes that re-enter the horizon today left the horizon during inflation. $E_I$ is subject to both theoretical and observational constraints, as illustrated in Figs. 1 & 2.

A multitude of inflation models involving a broad range of energy scales have been discussed in the literature, including chaotic inflation \cite{216}, brane inflation \cite{217,218} and others \cite{219}. However, very high $E_I$ has been argued to be theoretically problematic, at least for single field slow-roll inflation, because it involves super-Planckian displacements of the inflaton field $\phi$ \cite{220}. From Ref. \cite{217} the region in which $\phi$ moves at least two Planck masses is $E_I \gtrsim 2.4 \times 10^{16}$ GeV. The intuition that high $E_I$ is problematic seems borne out in many string theory models; an example is D-brane models \cite{218}.

Very low $E_I$ has been argued to be theoretically problematic also. Naive consideration of families of potential energy functions suggests that $E_I \lesssim 2 \times 10^{16}$ GeV ($r \lesssim 0.01$) is non-generic \cite{221}. One of the most striking successes of high $E_I$ potentials is that they can naturally predict $n_s \sim 0.96$, and generic low-energy potentials fail to make this prediction. Low $E_I$ potentials have the slow-roll parameter $\epsilon$ exponentially small, so that the observation $n_s = 1 - 6\epsilon + 2\eta = 0.960 \pm 0.013$ \cite{221} implies $\eta = -0.02 \pm 0.0065$, leaving us wondering why $\eta$ is so small when it could just as well have been of order $-1$. This problem is not alleviated by anthropic considerations \cite{222}.

By using the observed value of density fluctuations, and setting $\epsilon < 10^{-4}$ as the boundary, $E_I \lesssim 6.7 \times 10^{15}$ GeV defines this problematic low-scale region. These theoretical issues for inflation are indicated by the vertical regions in Fig. 1. Although there are inflation models in the literature at energy scales both above and below this naive window, the debate about whether they are generic continues.

With theory in limbo, we turn to observational guidance. High $E_I$ implies a large amplitude for primordial gravitational waves (GWs). $E_I > 3.8 \times 10^{15}$ GeV ($r > 0.22$) is ruled out by WMAP5 plus BAO and SN data \cite{224}, as indicated by the orange region of Fig. 2. Possible future searches for primordial GWs have rightly been a focus of attention.\textsuperscript{1} In this article we emphasize the additional information that can be learned from isocurvature fluctuations.

\subsection{B. Axion Physics}

The QCD Lagrangian accommodates a gauge invariant, Lorentz invariant, renormalizable term $\propto \theta \mathbf{E} \cdot \mathbf{B}$, with $\theta \in [-\pi, \pi]$, that manifestly breaks P and T symmetry. Precision bounds on the electric dipole moment of the neutron constraint $|\theta| \lesssim 10^{-10}$. The striking smallness of this parameter, which the standard model leaves unex-

\footnote{Electronic address: mphertz@mit.edu}
plained, defines the strong P and T problem (a.k.a. “CP problem”). After introducing a new asymptotic (or alternatively, classical) Peccei-Quinn (PQ) symmetry [13] which is spontaneously broken, the effective $\theta$ becomes a dynamical variable, and relaxes toward extremely small values. The consequent approximate Nambu-Goldstone boson is the axion [14][15].

The simplest axion models contain only one phenomenologically significant parameter: $f_a$, the scale at which the PQ symmetry breaks. The zero temperature Lagrangian for the complex field $\phi = \rho e^{i\theta}/\sqrt{2}$ is

$$\mathcal{L} = \frac{1}{2} f_a^2 (\partial_\mu \theta)^2 + \frac{1}{2} (\partial_\mu \rho)^2 - 2\Lambda^4 \sin^2(\theta/2) - \lambda (|\phi|^2 - f_a^2/2)^2$$

($\Lambda \approx 78$ MeV, $\rho$ is irrelevant at low energies). Accelerator bounds require $f_a$ to be well above the electroweak scale, and stellar astrophysics constraints place considerably higher limits. Given that electroweak values for $f_a$ are ruled out, economy suggests that $f_a$ could be associated with unification or Planck scales, rather than the “desert” of particle physics or super-Planckian scales, as indicated by the horizontal regions in Fig. [1]. This intuition for $f_a$ seems borne out in string theory, where $f_a$ for the energy scale of inflation $E_I$ and the axion PQ scale $f_a$. For $E_I \gtrsim 2.4 \times 10^{16}$ GeV, the inflaton must undergo super-Planckian excursions in field space (in single field models). For $E_I \lesssim 6.7 \times 10^{15}$ GeV, generic inflation potentials fail to reproduce the observed nearly scale invariant power spectrum. For $f_a \gtrsim 2.4 \times 10^{18}$ GeV, the PQ breaking scale is super-Planckian. For $f_a \lesssim 10^{15}$ GeV (and $f_a \gg T_{\text{TeV}}$), the PQ breaking is in the “desert” of particle physics and non-trivial to achieve in string theory. This leaves the region labeled “naive window”. Why is universe nearly scale-invariant? Why is universe nearly scale-invariant? Why is universe nearly scale-invariant?

Any realistic inflationary anthropic scenario and below this line is the classic scenario. The region in which there is too much isocurvature, $z_{\alpha} > 0.072$, depends on the axion fraction $R_{\alpha} \equiv \xi_{\alpha}/\xi_{\text{CDM}}$ of the CDM; the purple region applies for any $R_{\alpha}$, the blue region for $R_{\alpha} > 0.25\%$ (which is expected at 95% confidence), and the cyan region for $R_{\alpha} = 100\%$. The green region has too much axion CDM; $\xi_{\alpha} > 2.9$ eV. Each constraint is divided into two parts: the darker part is for a conservative value $\chi = 1/20$ and the lighter part is for a moderate value $\chi = 1$. The orange region has excessive GWs amplitude: $Q_I > 9.3 \times 10^{-6}$. The yellow region has too much axion interaction in stars (darker is firmly ruled out, lighter is for some analyses). The brown region is excluded by the laboratory ADMX search. The dashed cyan, orange, and brown lines are future targets for isocurvature, GWs, and ADMX searches, respectively.
C. Cosmological Observables, Summary

Quantum fluctuations in an effective adiabatic field give rise to the standard adiabatic fluctuations that have grown into our cosmologically observed large-scale structure. If the PQ symmetry undergoes spontaneous symmetry breaking before the end of inflation, quantum fluctuations in the consequent light axion field give rise to isocurvature fluctuations. The amplitude of the isocurvature fluctuations grows with $E_I$, so upper bounds on the amplitude of isocurvature fluctuations imply upper bounds on $E_I$.

The purpose of this article is to delineate these bounds, extending earlier work on this subject such as 17, 21–25. The bounds depend sensitively on the fraction $R_a$ of CDM in the form of axions, which in turn depends not only on $f_a$, but also on the local initial misalignment angle $\theta \in [-\pi, \pi]$. These constraints are shown in Figs. 2 for different choices of the axion CDM fraction. We will estimate this fraction using statistical arguments.

In Section II we calculate the production and late time abundance $\xi_a$ of axions and the amplitude $\alpha_a$ of isocurvature fluctuations, as well as reviewing the amplitude $Q_I$ of primordial GWs. These three observables depend on two micro-physical parameters: the PQ scale $f_a$ (or equivalently, the $T = 0$ axion mass $m_a$) and the energy scale of inflation $E_I$ (or equivalently, the Hubble scale of inflation $H_I$), and on one “environmental” parameter: the misalignment angle $\theta$. We summarize the final formulae here:

$$\xi_a = \frac{\Lambda}{\langle \delta T/T \rangle^2} \frac{2}{25} \left( f_a/m_{Pl} \right)^2 \left( 2 \sigma_a^2 + \sigma_\theta^2 \right) \chi F, \quad (1)$$

$$\alpha_a = \frac{8}{25} \frac{(\Lambda/\langle \delta T/T \rangle^2)}{(\langle \delta T/T \rangle^2)^2} \sigma_a^2 \left( f_a/m_{Pl} \right)^2 \chi^2 F^2, \quad (2)$$

$$Q_I = \frac{H_I}{5\pi m_{Pl}} = \frac{E_I^2}{5\sqrt{3} \pi m_{Pl}^2}, \quad (3)$$

where

$$F \approx \begin{cases} 2.8 \left( \frac{\Lambda}{\langle \delta T/T \rangle^2} \right)^{2/3} \left( \frac{f_a}{m_{Pl}} \right)^{7/6} \frac{\Lambda}{\langle \delta T/T \rangle^2} & \text{for } f_a \lesssim \hat{f}_a, \\ 4.4 \left( \frac{f_a}{m_{Pl}} \right)^{3/2} & \text{for } f_a \gtrsim \hat{f}_a, \end{cases} \quad (4)$$

$$\sigma_\theta = \frac{\gamma H_I}{2\pi f_a} = \frac{E_I^2}{2\sqrt{3} \pi f_a m_{Pl}}, \quad (5)$$

(definitions are given below).

The most recent observational bounds from WMAP5 combined with other data are 11

$$\xi_a \leq 2.9 \text{ eV}, \quad \alpha_a < 0.072, \quad Q_I < 9.3 \times 10^{-6}, \quad (6)$$

thereby constraining the two micro-physical parameters $E_I$ and $f_a$.

These expressions for $\xi_a$ and $\alpha_a$ only apply if the PQ symmetry undergoes spontaneous symmetry breaking before inflation and is not restored thereafter. This is true if $f_a$ exceeds the Gibbons-Hawking temperature during inflation and the maximum post-inflationary thermalization temperature, as we discuss in Section II C. Inefficient thermalization leads to constraints displayed in Fig. 2 (top), while efficient thermalization leads to constraints displayed in Fig. 2 (bottom).

In Section II we present a statistical estimate for the axion abundance to place additional constraints on out parameter space. In Section IV we conclude by discussing the implications for inflationary model building and future prospects.

II. AXION COSMOLOGY

In this section, we review the production of axions in the early universe, their abundance in the late universe, and the amplitude of isocurvature fluctuations, following Refs. 17–21, and explain and derive eqs. 1–5. We focus on axion production from the so called “vacuum misalignment” mechanism only. This provides the most conservative constraints. Additional production mechanisms, such as cosmic string decay, are subject to larger theoretical uncertainties (e.g., see 26).
A. Onset of Axion Production

In an expanding flat FRW background at temperature $T$ with Hubble parameter $H(T)$, the phase field $\theta$ of broken PQ symmetry satisfies the equation of motion

$$\ddot{\theta} + 3H(T)\dot{\theta} - \frac{\nabla^2 \theta}{a^2} = -\frac{1}{f_a^2} \frac{\partial}{\partial \theta} V(\theta, T), \quad (7)$$

where dots indicate derivatives with respect to coordinate time. Here $V(\theta, T)$ is the temperature dependent potential induced by QCD instantons. At zero temperature, $V(\theta, 0) = \Lambda^4 (1 - \cos \theta)$, where $\Lambda \approx 78$ MeV sets the scale of the vacuum energy of QCD. For small values of the axion field, the potential is approximately harmonic:

$$V(\theta, T) \approx \frac{1}{2} m_a(T)^2 f_a^2 \theta^2. \quad (8)$$

The mass is temperature dependent, with high and low $T$ limits given by

$$m_a(T) \approx m_a(0) \begin{cases} b \left( \frac{\Lambda Q}{T} \right)^4 & \text{for } T \gtrsim \Lambda_Q, \\ 1 & \text{for } T \lesssim \Lambda_Q, \end{cases} \quad (9)$$

where $\Lambda_Q \approx 200$ MeV is the scale at which QCD becomes strongly coupled, $b = \mathcal{O}(10^{-2})$ depends on detailed QCD physics, and $m_a(0)$ is the zero temperature axion mass, related to the PQ scale $f_a$ and $\Lambda$ by $m_a(0) = \Lambda^2 / f_a$. The temperature dependence is illustrated in Fig. 3.

In the early universe, the axion field is effectively massless, and so the right hand side of eq. (7) is negligible. Hence the zero mode of the axion field is essentially frozen due to Hubble friction. When the temperature $T$ drops below $T_{osc}$, defined by

$$3H(T_{osc}) \approx m_a(T_{osc}), \quad (10)$$

the axion field will begin to oscillate, producing axions. Since this occurs during the radiation dominated era, we have

$$H(T)^2 = \frac{1}{3m_T^2} \frac{\pi^2}{30} g_*(T) T^4, \quad (11)$$

where $m_T \approx 2.4 \times 10^{18}$ GeV and $g_*(T)$, the effective number of relativistic degrees of freedom, depends on whether $T_{osc}$ occurs before or after the QCD phase transition: $g_*(T_{osc}) = 61.75$ for $T_{osc} \gtrsim \Lambda_Q$ and $g_*(T_{osc}) = 10.75$ for $T_{osc} \lesssim \Lambda_Q$. Eqs. (9) (11) allow us to solve for $T_{osc}$ in terms of $f_a$:

$$T_{osc} \approx \begin{cases} 0.36 \Lambda_Q^{2/3} \Lambda^{1/3} \left( \frac{m_{Pl}}{f_a} \right)^{1/6} & \text{for } f_a \lesssim \hat{f}_a, \\ 0.55 \left( \frac{m_{Pl}}{f_a} \right)^{1/2} & \text{for } f_a \gtrsim \hat{f}_a, \end{cases} \quad (12)$$

where $\hat{f}_a$, which reflects the break in eq. (9), is defined in eq. (10) below.

B. Density of Axions

At the onset of production (when $T = T_{osc}$) the axion energy density is

$$\rho_a(T_{osc}) \approx \frac{1}{2} m_a(T_{osc})^2 f_a^2 \langle \theta^2 \rangle f(\theta_i^2) \chi, \quad (13)$$

FIG. 3: Axion mass $m_a$ (solid) and axion energy density $\rho_a$ (dashed), as a function of temperature $T$ (or time $t$). The red and blue curves are for PQ breaking scales $f_a = 10^{13}$ GeV and $f_a = 10^{15}$ GeV, respectively. The arrow indicates the effect of increasing $f_a$. The decreasing orange curve is the Hubble parameter $3H$. We have taken $\langle \theta^2 \rangle = \pi^2 / 3$, although this is modified by the intervention of inflation, as explained in the text. For simplicity, we have here only kept track of the variation with temperature in the number of relativistic degrees of freedom before ($g_\ast = g_{\ast S} = 61.75$) and after ($g_\ast = g_{\ast S} = 10.75$) the QCD phase transition, and taken $\chi = 1$.

$^2$ $\Lambda$ is set by $\Lambda_Q$ and quark masses: $\Lambda^2 = \sqrt{\frac{\pi}{12}} f_a m_q$, $z = m_a/m_d \approx 0.56$.

$^3$ If there are $N$ distinct vacua for $\theta$, then we should replace $f_a$ by $f_a/N$ here and throughout the article. However, any $N > 1$ models are expected to have a large overabundance of energy density from domain walls, unless inflation intervenes.

$^4$ We assume here that the universe before BBN is adequately described by conventional physics. See Ref. [27] for other scenarios.
Here \( (\theta^2) \) is the spatial average over our Hubble volume of the square of the initial misalignment. In terms of its mean \( \theta_i \) and standard deviation \( \sigma_0 \), \( (\theta^2) = \theta_i^2 + \sigma_0^2 \). If the axion field is established before (or during the early stages of) inflation, then spatial variations in \( \theta \) are smoothed out over our Hubble volume \( (\nabla^2 \theta/a^2 \rightarrow 0) \). Then \( \theta_i = (\theta) \) in our Hubble volume is an angle drawn from a uniform distribution: \( \theta_i \in [-\pi, \pi] \), with a small variance that we discuss in the next subsection. On the other hand, if the axion field is established after inflation, then \( \theta_i = (\theta) = 0 \), with variance \( \sigma_0^2 = \pi^2/3 \) due to small scale variations. \( f(\theta^2) \) is a fudge factor acknowledging anharmonicity in the axion potential; for \( \theta_i \rightarrow 0 \), \( f(\theta^2) \rightarrow 1 \). Finally, \( \chi \) is a dimensionless correction factor due to temperature dependence during formation. In our numerics, we take \( \Lambda_Q = 200 \text{ MeV} \), \( b = 0.018 \), and absorb all theoretical uncertainties into \( \chi \). A conservative value is \( \chi = 1/20 \) and a more moderate value is \( \chi = 1/5 \); both values are reported in Fig. 2.

By following the redshift as the universe expands from axion-formation to today, we can convert this initial energy density into a prediction for the present density, as illustrated in Fig. 3. Since we are focusing on the zero-mode, the axions form a non-relativistic Bose-Einstein condensate. At late times (say today’s temperature \( T_0 \)), the axion energy per photon is

\[
\xi_a = \frac{\rho_a(T_0)}{n_s(T_0)} = \frac{m_a(T_0)}{m_a(T_{osc})} s(T_0) \frac{n_s(T_0)}{n_s(T_{osc})} ,
\]

where we have exploited the fact that entropy density \( s(T) \) in a comoving volume is conserved. The entropy density is given by

\[
s(T) = \frac{2\pi^2}{45} g_{*s}(T) T^3,
\]

where \( g_{*s}(T) \) is the effective number of relativistic degrees of freedom for \( s(T) \). Note that \( g_{*s}(T_{osc}) = g_{*s}(T_{osc}) \), \( g_{*s}(T_0) \) comes from photons and neutrinos: \( g_{*s}(T_0) = 2 + \frac{7}{8} \times 6 \times \frac{4}{11} = 3.91 \), and \( m_a(T_0) \approx m_a(0) \). Since the number density of photons \( n_s(T_0) = 2\zeta(3) T_0^3 / \pi^2 \) depends on temperature in the same way that the late time axion energy density \( \rho_a(T_0) \) does, the quantity \( \xi_a \) is a temperature-independent, or equivalently time-independent, measure of the axion abundance. In contrast, the commonly used quantities \( \Omega_a \) and \( h^2 \Omega_a \) do not tell us anything fundamental about our universe, since like \( T \), they are effectively alternative time variables that evolve as our universe expands. These different measures of axion density are related by \( \Omega_a h^2 \approx 0.0019 (\xi_a/1 \text{ eV})(T_0/1 \text{ K})^3 \), which at the present epoch \( (T_0 = 2.725 \text{ K}) \) reduces to \( \Omega_a h^2 \approx \xi_a/26 \text{ eV} \) \((h \text{ is the dimensionless Hubble parameter}).

Combining all this information yields eq. (11). Note that higher \( f_a \) (for a fixed value of \( (\theta^2) \)) corresponds to higher axion energy density, as seen in Fig. 3. The reason for this is that higher \( f_a \) corresponds to lower \( m_a(0) \), so that the onset of axion production, when \( 3H(T) \) has fallen to \( m_a(T) \), occurs later. Hence there is less redshifting of the axion energy density after production (furthermore, \( \rho_a(T_{osc}) \) is higher if \( f_a < f_a \)).

We locate the boundary between the low and high \( f_a \) limits by equating the two expressions for \( \xi_a \). They match when \( f_a = \tilde{f}_a \), where

\[
\tilde{f}_a \equiv 0.26 \left( \frac{\Lambda}{\Lambda_Q} \right)^2 m_{\nu_i} .
\]

The observed total density of cold dark matter in our universe \( \xi_{CDM} \approx 2.9 \text{ eV} \) implies \( \xi_a \leq 2.9 \text{ eV} \).\footnote{See Refs. [28, 29] for precise estimates of axion abundance.}

\[
\xi_a \leq 2.9 \text{ eV} .
\]

C. Fluctuations from Inflation

During inflation, the universe undergoes an approximately de Sitter phase with Hubble parameter \( H_I \). Quantum fluctuations during this phase induce several kinds of cosmological fluctuations.

- Adiabatic density fluctuations are generated, with an approximately scale-invariant spectrum. The measured amplitude is \( Q \approx 1.98 \times 10^{-5} \)\footnote{Here \( Q = \frac{4}{3} \Delta R(k = 0.002 \text{ Mpc}^{-1}) \) of Ref. [11].}.

- Primordial gravity waves are generated, with an approximately scale-invariant spectrum whose amplitude is given in eq. (3). WMAP5 plus BAO and SN data imply the bound: \( Q_1 < 9.3 \times 10^{-6} \) (95%). Since \( Q \) is measured and \( Q_1 \) is set by \( E_I \), the tensor to scalar ratio \( r = (Q_1/Q)^2 \) is often used to characterize the scale of inflation. Using the same notation as [11], it is bounded by \( r < 0.22 \) (95%).

- Any other light scalar fields, such as the axion, are imprinted with fluctuations during inflation, similarly to gravitons. The power spectrum of a canonically normalized scalar field \( \phi \), such as \( \phi_a \), in de Sitter space has a scale-invariant spectrum (e.g., see [31])

\[
\langle |\delta \phi_a(k)|^2 \rangle = \left( \frac{H_I}{2\pi} \right)^2 \frac{1}{k^3/2\pi^2} .
\]
It is essentially a thermal spectrum at Gibbons-Hawking temperature $T_{GH} = H_f / 2\pi$. Fluctuations in the misalignment angle in $k$-space are scaled as $\sigma_\theta = \sigma_a / f_a$, per eq. (5). We write the corresponding fluctuations in real space as $\sigma_a = \gamma H_f / 2\pi$, where $\gamma = O(1)$ is a dimensionless constant. Ref. [11] effectively takes $\gamma = 1$ and Ref. [22] argues that observations are sensitive to length scales corresponding to $\gamma \approx 4$, while in our figures we have taken a moderate value of $\gamma = 2$. These fluctuations provide a lower bound on $\xi_a$ and, as we discuss in the next subsection, on isocurvature fluctuations.

In the preceding discussion, we assumed the existence of a light axion field during inflation. This is true only if PQ symmetry is broken before inflation. Furthermore, if PQ symmetry is restored after inflation, the fluctuations will be washed out. PQ symmetry can be restored either by the maximum thermalization temperature after inflation, or by the maximum thermalization temperature during inflation, or by fluctuations in the axion energy density which could cause PQ symmetry restoration. Inflationary expansion implies a mass and a significant energy density (see Fig. 3), but any such fluctuations cannot alter the total energy density of the universe. Hence, these are isocurvature fluctuations.

To quantify the amplitude of isocurvature fluctuations, it is useful to introduce the fractional change in the number density to entropy density ratio:

$$S_i = \frac{\delta(n_i / s)}{n_i / s} = \frac{\delta n_i}{n_i} + 4\rho_r T / T.$$  (19)

For adiabatic fluctuations, $S_i = 0$. We assume that this is true for all species other than the axion. Isocurvature fluctuations in the total energy density involve a sum over all massive species and radiation:

$$0 = \delta \rho_{\text{tot}} = m_a \delta n_a + \sum_{i \neq a} m_i \delta n_i + 4\rho_r \frac{\delta T}{T}.  \quad (20)$$

These two equations will be used to obtain an expression for the corresponding temperature fluctuations.

Initially the energy density of the axion field is a small fraction of the ambient total density, so eq. (20) gives $\delta T / T \ll \delta n_a / n_a$, and $S_a = \delta n_a / n_a$. Since $n_a \propto \theta^2$ (ignoring anharmonic effects), this implies

$$S_a = \frac{\theta^2 - \langle \theta^2 \rangle}{\langle \theta^2 \rangle}. \quad (21)$$

Assuming $\delta \theta \equiv \theta - \langle \theta \rangle$ is Gaussian distributed, we can calculate $\langle S_a^2 \rangle$ in terms of $\theta_i \equiv \langle \theta \rangle$ and $\sigma_\theta^2 = \langle (\delta \theta)^2 \rangle$, as

$$\langle S_a^2 \rangle = \frac{2\sigma_\theta^2 (2\theta_i^2 + \sigma_\theta^2)}{\theta_i^4 + \sigma_\theta^4}. \quad (22)$$

---

7 The maximum thermalization temperature should not be confused with the reheating temperature, which can be somewhat lower [22]. The maximum thermalization temperature is the maximum temperature of the thermal bath post-inflation, while the reheating temperature is the temperature at the end of the reheating phase, i.e., at the beginning of the radiation era.

8 Later, around the onset of the matter dominated era, these isocurvature fluctuations are converted to adiabatic fluctuations, responsible for the familiar gravitational structures in our universe.

9 This is a good assumption in the regime where the axions comprise a significant fraction of the dark matter, i.e., $\theta_i^2 \gg \sigma_\theta^2$. 

---
Note that if $\theta_0^2 \ll \sigma_0^2$ then $\langle S_0^2 \rangle = 2$, while if $\theta_0^2 \gg \sigma_0^2$ then $\langle S_0^2 \rangle = 4\sigma_0^2/\theta_0^2$.

The most important axion induced temperature fluctuations are those on the largest scales. Such fluctuations enter the horizon well into the matter dominated era, where $\rho_r$ can be ignored. This implies

$$
\left( \frac{\delta T}{T} \right)_{\text{iso}} \approx -\frac{\xi_a}{3 \xi_m} S_a, \tag{23}
$$

where $\xi_m$ is the total matter energy density per photon, whose measured value is $\xi_m = 3.5 \text{ eV}$.

Following [11], we define $\alpha_a$ to be the fractional contribution to the CMB temperature power spectrum due to axion isocurvature:

$$
\alpha_a \equiv \frac{\langle (\delta T/T)_{\text{iso}}^2 \rangle}{\langle (\delta T/T)_{\text{tot}}^2 \rangle}. \tag{24}
$$

Using the relationship between $S_a$ and $\delta T/T$ in eq. (23) and the preceding expression for $\langle S_0^2 \rangle$ we obtain

$$
\alpha_a = \frac{8}{25} \frac{(\xi_a/\xi_m)^2}{\langle (\delta T/T)_{\text{tot}}^2 \rangle} \frac{\sigma_0^2(20^2 + \sigma_0^2)}{\theta_0^2 + \sigma_0^2}^2. \tag{25}
$$

COBE measured (and WMAP confirmed) the root-mean-square total temperature fluctuations to be $\langle (\delta T/T)_{\text{tot}}^2 \rangle \approx (1.1 \times 10^{-5})^2$, averaged over the first few $l$. Using the expression for $\xi_a$ given in eq. (1) we can write the isocurvature fluctuations as in eq. (2). This must be consistent with the latest observational bound $\alpha_a < 0.072$. Here we have used $\alpha_0$ of Ref. [11], which assumes isocurvature fluctuations are uncorrelated from curvature ones.

### III. DIRECT AND STATISTICAL CONSTRAINTS

It is conceivable that the axion abundance is negligible (but see the following subsection). This scenario (case (i)) requires $\theta_0^2 \ll \sigma_0^2$. By demanding $\alpha_a < 0.072$ (the current isocurvature bound) and using eq. (2), we obtain the most conservative bound, studied in Ref. [17], corresponding to the purple region marked “Any $R_a$” in Fig. 2.

At the other extreme, if axions are the dominant form of dark matter in the universe (case iii), then $\theta_0^2 \gg \sigma_0^2$. Again demanding $\alpha_a < 0.072$ in eq. (2), with $\theta_i$ determined from eq. (1) with $\xi_a = \xi_{\text{CDM}}$, the excluded region expands to include the cyan region marked “$R_a = 100\%$” in Fig. 2 (as well as the blue region marked “$R_a > 0.25\%$”).

Each of these three regions are bifurcated by a line. In all three cases, the rightmost part gives the most robust constraint, coming from a conservative value $\chi = 1/20$, while the leftmost part extend the constraints using a more moderate (and more speculative) value $\chi = 1$. This comes from our uncertainty in the total axion abundance.

#### A. Statistics of a Two-Component Model

The viability of large $f_a$ axion cosmology depends on taking selection effects seriously, since they produce a higher dark matter density $\xi$ than observed in most Hubble volumes. In particular, the density of a typical galaxy on $\rho \sim \xi^4$. Taking into account that denser galaxies have fewer stable solar systems due to close encounters with other stars, etc., it has been found that typical stable solar systems in large $f_a$ axion models reside in Hubble volumes where $\xi$ is comparable to the observed value.

Here we draw out a statistical implication for the predicted axion abundance, if there is a second contributor to the dark matter density. Consider the hypothesis that the total CDM ($\xi_{\text{CDM}}$) is comprised of axions ($\xi_a$) and some other component, say WIMPs ($\xi_W$): $\xi_{\text{CDM}} = \xi_a + \xi_W$. The unknown separate axion and WIMP abundances should be drawn from prior distributions determined by underlying micro-physical theories. For axions in the large $f_a$ regime, above any inflation temperatures, this scenario implies that the initial misalignment angle $\theta_i$ is uniformly distributed. In the regime where the axion abundance is non-negligible ($\theta_i^2 \gg \sigma_i^2$), but still sufficiently small that we can ignore anharmonic effects ($\theta_i^2 < 1$), we have $\xi_a \propto \theta_i^2$. Since $\theta_i$ is uniformly distributed, it is simple to show that

$$
p_{\text{prior}}(\xi_a) \propto \frac{1}{\sqrt{\xi_a}}. \tag{26}
$$

In contrast, we do not have a reliable prior distribution $p_{\text{prior}}(\xi_W)$ for the WIMP.

As discussed in Ref. [20], selection effects depend only on the sum $\xi_{\text{CDM}} = \xi_a + \xi_W$, so the total joint distribution for axions and WIMPs is

$$
p(\xi_a, \xi_W) \propto p_{\text{prior}}(\xi_a)p_{\text{prior}}(\xi_W)p_{\text{sel}}(\xi_{\text{CDM}}) \tag{27}
$$

As demonstrated in Ref. [20], the observed value of CDM $\xi_{\text{CDM}} \approx 2.9 \text{ eV}$ is nicely consistent with this distribution. Given this, we can focus on the remaining one-dimensional distribution for the axion:

$$
p(\xi_a) \propto p_{\text{prior}}(\xi_a)p_{\text{prior}}(\xi_{\text{CDM}} - \xi_a). \tag{28}
$$

Unless $p_{\text{prior}}$ is sharply peaked at $\xi_{\text{CDM}}$, the axion prior (when integrated) disfavors very small values of $\xi_a$. For
example, let us take the WIMP prior to be uniform. We can then make a prediction for the axion abundance at, say, 95% confidence. Defining $\xi_a$ implicitly through

$$\int_0^{\xi_a} p(\xi_a) d\xi_a = 0.05 \quad (29)$$

and solving eq. (29) using eq. (26), we find $\hat{\xi}_a = (0.05)^2 \xi_{\text{CDM}} = 0.25\% \xi_{\text{CDM}}$. This says that it statistically unlikely – at the 95% level – for axions to comprise less than 0.25% of the CDM of the universe (case (ii)).

By setting $\xi_a = 0.25\% \xi_{\text{CDM}}$, we rule out the blue region marked $R_a > 0.25\%$ in Fig. 2 with high confidence. In other words, without assuming that axions comprise all the CDM, we find that on statistical grounds axions must comprise at least a non-negligible fraction of the universe’s CDM, allowing us to extend the excluded region in Fig. 2 further towards the upper left.

**B. Additional Constraints**

The preceding applies in the $f_a > \text{Max}\{T_{\text{GH}}, T_{\text{max}}\}$ regime, where the initial misalignment angle $\theta_i$ takes on a single constant value in our Hubble volume. For $f_a < \text{Max}\{T_{\text{GH}}, T_{\text{max}}\}$, the misalignment angle varies on cosmologically small scales, with average $\langle \theta^2 \rangle = \pi^2/3$. In this regime the isocurvature fluctuations are negligible. In this case, bounds arise from the requirement that the axion abundance is not greater than the observed total CDM abundance: $\xi_a \leq \xi_{\text{CDM}} \approx 2.9 \, \text{eV}$. Using the upper expression for $\xi_a$ in eq. (1), with $\theta_i^2 + \sigma_0^2 \rightarrow \langle \theta^2 \rangle = \pi^2/3$, we find that $f_a > 2.3 \times 10^{11} \chi^{-6/7} \text{GeV}$ is ruled out. For the conservative value $\chi = 1/20$, this excludes the upper part of the green region marked “Too much CDM” in Fig. 2 and for the moderate value $\chi = 1$ this extends the exclusion to the lower part of the green region.\(^{11}\)

Also, $m_a(0) > 10^4 \, \mu\text{eV}$ is firmly ruled out (and $\delta m_a(0) > 10^3 \, \mu\text{eV}$ for some analyzes), since in this regime the coupling of axions to matter is too large, affecting the physics of stars, such as the cooling of red giants and the neutrino flux from SN 1987A.\(^8\) (yellow region at bottom of Fig. 2) Furthermore, the ADMX search for axion dark matter in a microwave cavity detector has ruled out axions comprising the bulk of the halo dark matter in the following mass window: $1.9 \, \mu\text{eV} < m_a(0) < 3.3 \, \mu\text{eV}$ (brown band in Fig. 2) for so-called KSVZ axions, and the sub-window $1.98 \, \mu\text{eV} < m_a(0) < 2.17 \, \mu\text{eV}$ for so-called DSVF axions\(^{34,35}\). The remaining white region is the allowed “classic window”.

In Fig. 2 (top), corresponding to inefficient thermalization ($\epsilon_{\text{eff}} \approx 0$), the boundary between the anthropic and classic regimes is $f_a = T_{\text{GH}} = H_t/2\pi$. In Fig. 2 (bottom), corresponding to efficient thermalization ($\epsilon_{\text{eff}} = 10^{-3}$, with $\epsilon_{\text{eff}} = 10^{-1.5}$, 1 indicated), the boundary between the two regimes is $f_a = T_{\text{max}} = \epsilon_{\text{eff}} E_I$. Efficient thermalization thus opens up a larger “classic window”, but the “anthropic window” is essentially unaltered.

**C. Effect of Falling Density During Inflation**

In our analysis, we have treated inflation as occurring at one rather well-defined Hubble scale. Although this is a good approximation in some inflation models, there are others giving an appreciable change in $H$ between its value (say $H_I$) when the modes that are now re-entering our horizon left the horizon (55 or so e-foldings before the end of inflation), and its value (say $H_{\text{end}}$) at the end of inflation. This is particularly relevant to Fig. 2 (top), since it implies that the boundary between “anthropic” and “classic” regimes is blurred, since $T_{\text{GH}} = H/2\pi$ is evolving. For high scale inflation, $H$ typically changes by an amount of order the number of e-foldings, i.e., $O(10^2)$, while for low scale inflation models, $H$ typically changes very little.

If we consider $H_{\text{end}} \ll H_I$, then the PQ symmetry can break during inflation. The resulting cosmology could be quite interesting with axion dark matter varying appreciably from one point in our Hubble volume to another, but is ruled out since $Q \sim 10^{-5}$. If PQ breaking occurs very close to 55 or so e-foldings before the end of inflation, then $\theta_i$ can be smoothed out on today’s cosmological scales and make the analysis “anthropic”. Otherwise, we expect the “classic” analysis to apply as usual, providing a ruled out green region in Fig. 2. Hence, we expect such corrections to the constraints to be reasonably minimal.

**IV. DISCUSSION**

We have surveyed observational constraints on the parameter space $\{E_I, f_a\}$ of inflation and axions, finding that most of it is excluded, leaving only two allowed regions that we term classic and anthropic windows. Part of the classic window $f_a \sim 10^{11} - 10^{12} \text{GeV}$ will be intensely explored by the ongoing ADMX experiment. The region indicated by the arrow to the horizontal brown lines in Fig. 2 to $m_a(0) = 10 \, \mu\text{eV}$ is expected to be explored by the end of ADMX Phase II, and onwards to $m_a(0) = 100 \, \mu\text{eV}$ some years thereafter.\(^{36}\) In this window, comparatively little can be concluded about the scale of in-
flation. From Fig. 2 (top), taking \( m_a(0) = 100 \mu \text{eV} \) and assuming \( \xi_a > 0.25\% \xi_{\text{CDM}} \), we rule out \( 8.4 \times 10^{13} \text{GeV} \lesssim E_I \lesssim 1.3 \times 10^{15} \text{GeV} \).\(^\text{12}\) From Fig. 2 (bottom), the upper end of this ruled out region is reduced due to efficient post-inflation thermalization. Although we can rule out a range of low scale inflation models, these conclusions are not exceedingly strong.

On the other hand, a large \( f_a \) axion has strong implications for inflation. According to both Figs. 2 (top) & (bottom), if \( f_a = 10^{16} \text{GeV} \) then \( E_I \gtrsim 5.5 \times 10^{14} \text{GeV} \) (\( r \gtrsim 9 \times 10^{-9} \)) is ruled out at 95\% confidence for the conservative value \( \chi = 1/20 \), and \( E_I \gtrsim 2.6 \times 10^{14} \text{GeV} \) (\( r \gtrsim 4 \times 10^{-10} \)) is ruled out for the moderate value \( \chi = 1 \). The geometric mean is \( E_I \gtrsim 3.8 \times 10^{14} \text{GeV} \) (\( r \gtrsim 2 \times 10^{-9} \)), which is reported in the abstract. This is incompatible with many models of inflation, including "classic" models with a single slow-rolling scalar field in a generic potential. For example, monomial potentials \( V \propto \phi^p \) predict \( r = 4p/N_e \), where \( N_e \) is the number of e-foldings of inflation from when it generated our horizon scale fluctuations to when it ended. For such models, \( N_e \) around 50 or 60 is expected, so any reasonable \( p \) is ruled out, including \( \phi^3 \) chaotic inflation \(^\text{4}\) and the stringy N-inflation \(^\text{37}\) and Monodromy \(^\text{38}\) models. The same is true for exponential potentials \( V \propto \exp(-\sqrt{2p/\phi}/m_v) \), which predict \( r = 16p \).

Evidently there is considerable tension between the theoretically appealing large \( f_a \) and high-scale inflation scenario (see Fig. 1) and the observational constraints (see Fig. 2). Low-scale inflation may be emerging as favored from recent work in string theory. If we consider the small subspace (see \(^\text{39, 41}\)) of presently constructed string models that both inflate and agree with the observed values of \( Q \) and \( n_s \), we are left with models that tend to be at rather small energies, typically \( r < 10^{-8} \) for D-brane models and various other scenarios such as modular inflation \(^\text{42}\). There are also arguments for very low \( r \) in the simple KKLT framework discussed in Ref. \(^\text{43}\). This allows \( f_a \sim 10^{16} \text{GeV} \) to be marginally consistent with present isocurvature bounds. Although it is highly premature to conclude that very low energy scale is a generic feature of string realizations of inflation, it is intriguing that many string constructions have this feature. (See \(^\text{37, 38}\) for interesting exceptions.)

The Planck satellite, CMBPol, and upcoming sub-orbital CMB experiments should probe well beyond the current bound on GWs of \( r < 0.22 \), perhaps reaching \( r \sim 0.01 \). This is indicated by an arrow toward the vertical dashed red line in Fig. 2. If gravity waves are observed in this regime, then the PQ scale \( f_a \) must be in the classic window.

Our considerations emphasize the fundamental importance of improving bounds on isocurvature fluctuations. For example, an order of magnitude improvement to \( \alpha \sim 0.007 \) would push the isocurvature bounds to the diagonal dashed cyan line in Fig. 2. (We have indicated the improvement for the case where axions comprise all the CDM: \( \xi_a = \xi_{\text{CDM}} \).) Detection of isocurvature fluctuations in this regime has three important implications:

1. It could be interpreted as evidence for the existence of the axion field, and assuming this:

2. It would probe low inflation scales \( E_I \) far beyond the scope of any foreseeable GW measurements.

3. It would be evidence that we live in a highly atypical Hubble volume, i.e., \( \{E_I, f_a\} \) must be in the anthropic window.

Isocurvature modes and tensor modes thus provide complementary constraints on fundamental physics, making it fruitful to study dark matter and inflation in a unified way.

Acknowledgments

We would like to thank Daniel Baumann, Mauro Brigante, Andrea de Simone, Shamit Kachru, David B. Kaplan, Tongyan Lin, Gray Rybka, and Paul Steinhardt, for helpful discussions, and a special thanks to Krishna Rajagopal and Scott Watson for comments on a preliminary version of this paper. M. P. H. and F. W. thank the Department of Energy (D.O.E.) for support under cooperative research agreement DE-FC02-94ER40818. This work was supported by NSF grants AST-0134999 and AST-05-06556, a grant from the John Templeton foundation and fellowships from the David and Lucile Packard Foundation and the Research Corporation.

\(^\text{12}\) The quoted lower end of the ruled out region is the geometric mean of the conservative and moderate \( \xi_a \)-scenarios.
