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

New bounds for axions and axion-like particles with keV-GeV masses
Marius Millea, Lloyd Knox, and Brian D. Fields
Phys. Rev. D 92, 023010 — Published 23 July 2015
DOI: 10.1103/PhysRevD.92.023010
New Bounds for Axions and Axion-Like Particles with keV-GeV Masses

Marius Millea and Lloyd Knox

Department of Physics, One Shields Avenue, University of California, Davis, California 95616, USA

Brian D. Fields

Departments of Astronomy and of Physics, University of Illinois, Urbana-Champaign, IL, USA

We give updated constraints on hypothetical light bosons with a two-photon coupling such as axions or axion-like particles (ALPs). We focus on masses and lifetimes where decays happen near big bang nucleosynthesis (BBN), thus altering the baryon-to-photon ratio and number of relativistic degrees of freedom between the BBN epoch and the cosmic microwave background (CMB) last scattering epoch, in particular such that $N_{\text{eff}}^{\text{CMB}} < N_{\text{eff}}^{\text{BBN}}$ and $\eta_{\text{CMB}} < \eta_{\text{BBN}}$. New constraints presented here come from Planck measurements of the CMB power spectrum combined with the latest inferences of primordial $\eta_{4\text{He}}$ and D/H abundances. We find that a previously allowed region in parameter space near $m = 1$ MeV and $\tau = 100$ ms, consistent with a QCD axion arising from a symmetry breaking near the electroweak scale, is now ruled out at $> 3\sigma$ by the combination of CMB+D/H measurements if only ALPs and three thermalized neutrino species contribute to $N_{\text{eff}}$. The bound relaxes if there are additional light degrees of freedom present which, in this scenario, have their contribution limited to $\Delta N_{\text{eff}} = 1.1 \pm 0.3$. We give forecasts showing that a number of experiments are expected to reach the sensitivity needed to further test this region, such as Stage-IV CMB and SUPER-KEKB, the latter a direct test insensitive to any extra degrees of freedom.

I. INTRODUCTION

Recent improvements in both measurements of the Cosmic Microwave Background (CMB) angular power spectrum and inferences of primordial elemental abundances formed during Big Bang Nucleosynthesis (BBN) motivate us to reconsider bounds on hypothetical scenarios which alter the expansion rate or inject energy into the plasma around these two epochs. One of the simplest scenarios involves a radiatively decaying particle, for which axions or so called “axion-like particles” (ALPs) provide a theoretically well motivated candidate. Axions arise from perhaps the most elegant solution to the strong CP problem as the pseudo Nambu-Goldstone boson of a new spontaneously broken symmetry [1–4]. Axions have a mass and standard model couplings controlled by a single parameter: the energy scale of the symmetry breaking. ALPs form a more general class of particles where the mass and couplings are independent. Such models can arise from other new symmetries which are spontaneously broken [5, 6], and in string theory [7].

In the parameter space of interest here, ALPs are weakly interacting, making them difficult to detect in the laboratory. Cosmological constraints serve as a natural complement as the weak coupling generally leads to later decays allowing the particles to become non-relativistic and pick up energy compared to the plasma, leading to observable consequences upon their decay. Astrophysical bounds are also important as ALPs provide a new method for energy release from stars and supernovae. Some early calculations and compilation of cosmological, astrophysical, and laboratory bounds on ALPs include those from Massó and Toldrà [8, 9]. Cosmological bounds were recently updated by Cadamuro et al. [10] which considered only axions, and Cadamuro and Redondo [11] who extended this more generally to ALPs.

Among other significant advances, Cadamuro and Redondo [11] used newer data, treated out-of-equilibrium decays more carefully, and performed precise calculations of the implications for BBN. These and other known ALP bounds are tabulated in [12–14]. Our work updates these by 1) using the latest inferences of primordial element abundances, 2) using the latest measurements of the CMB power spectrum measurements from Planck, 3) having improved calculations of CMB spectral distortions.

We also highlight the importance of a region in parameter space near $m = 1$ MeV and $\tau = 100$ ms which we term the MeV-ALP window. This region is interesting because it previously has evaded all known constraints (also noted in [12, 15]) and, as we show, can correspond to a particular axion model we will call the DFSZ-EN2. Additionally, in this mass window the symmetry breaking scale is much lower than that in the often-considered “invisible” axion models. A main conclusion of this paper is that the possibility of the DFSZ-EN2 or any other ALP hiding in plain sight in the MeV-ALP window is ruled out by the newer bounds presented here. We show, however, that these bounds are model dependent and the MeV-ALP window can be reopened if there is other exotic radiation present.

The paper is structured as follows. In Sec. II we discuss in more detail the scenario and its implications for cosmology. Sec. III describes the new and tabulated constraints which we use. In Sec. IV A we further discuss the MeV-ALP window and in Sec. IV D we give forecasts for future probes.
II. THE SCENARIO

We begin by discussing the cosmological impact of ALPs, which we define as any particle with a mass $m_\phi$ and two-photon coupling $g_{\phi\gamma\gamma}$. Following standard conventions in the ALP literature, the effective Lagrangian is

$$\mathcal{L} = \frac{1}{2}[(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}m_\phi^2 \phi^2] - \frac{g_{\phi\gamma\gamma}}{4} \phi F_{\mu\nu} \tilde{F}^{\mu\nu},$$

where $F$ is the electromagnetic field strength tensor, $\tilde{F}$ its dual, and $\phi$ the ALP field. We often describe the two dimensional parameter space with $m_\phi$ and, in place of $g_{\phi\gamma\gamma}$, the lifetime for decay into photons

$$\tau_{\phi\gamma} \equiv \Gamma_{\phi\gamma}^{-1} = \frac{64\pi}{m_\phi^2 g_{\phi\gamma}^2}.\quad (2)$$

Two processes drive the cosmological evolution of the ALP energy density. The first is the Primakoff interaction which allows for conversion between photons and ALPs in the presence of a charged particle $q$, via $\gamma q \leftrightarrow \phi q$. Because it is a four-point interaction, the scattering rate for the Primakoff process will always begin in equilibrium and freeze out at later times. The second process is

$$\rho_{\phi\gamma} = \frac{\eta_\gamma}{\eta_b}.$$ (3)

The other quantity is the baryon-to-photon ratio $\eta$, or equivalently the energy density in baryons $\Omega_b h^2$,

$$\eta = \frac{n_b}{n_\gamma} = 2.74 \times 10^{-8} \Omega_b h^2.\quad (4)$$

We sometimes superscript these quantities with BBN or CMB depending on the epoch at which they are evaluated, although we note there is no exact definition as they can change with time in this scenario.

ALP decays can only reduce or leave unchanged the value of $N_{\text{eff}}^{\text{CMB}}$. If the decays occur while the neutrinos are still fully coupled then $N_{\text{eff}}^{\text{CMB}}$ does not change. Later decays increase the temperature of the plasma relative to the neutrinos and thus decrease $N_{\text{eff}}^{\text{CMB}}$. The value of $N_{\text{eff}}^{\text{BBN}}$ can also be decreased in the same manner if decay happens before BBN, although this region in parameter space is fairly small. A much larger region of parameter space corresponds to ALPs decaying after BBN and increasing $N_{\text{eff}}^{\text{BBN}}$ by simply contributing to the relativistic energy density during BBN. ALP decays affect $\eta$ similarly to $N_{\text{eff}}$ by causing its value to be less after the decay. The difference is that with $\eta^{\text{CMB}}$ is held constant, $\eta$ before the decay is now increased. Assuming that the physics of BBN is unchanged, both an increase in $\eta^{\text{BBN}}$ and an increase in $N_{\text{eff}}^{\text{BBN}}$ serve to increase the amount of primordial helium produced. For deuterium, however, there is partial cancellation, leaving the primordial abundance only slightly reduced. This cancellation is key to allowing the ALP scenario at all in light of the very tight constraints on primordial deuterium (see Sec. III B). The increase in primordial helium coupled with the decrease in $N_{\text{eff}}^{\text{CMB}}$ moves along a degeneracy direction for the CMB constraints, and is hence also generally allowed.

More quantitatively, we must track the evolution of the phase space distribution function $f_\phi$, which is governed by the Boltzmann equation

$$\frac{df_\phi}{dt} = (C_q + C_\gamma)(f_\phi^{eq} - f_\phi),\quad (5)$$

where the $f_\phi$’s are comoving, $f_\phi^{eq}$ is a Bose-Einstein distribution, and a dependence on the comoving momentum $p$ is omitted for brevity. When electrons and positrons are the only charged particles present, the scattering rate for the Primakoff interaction, $C_q$, is given by

$$C_q \approx n_e p g_{\phi\gamma}^2 \frac{\alpha}{16} \log \left\{ 1 + \frac{4E(m_e + 3T)^2}{m_\gamma^2 [m_e^2 + (m_e + 3T)^2]} \right\},\quad (6)$$

number of relativistic species, $N_{\text{eff}}$. As usual, this is taken so that in the limit of complete neutrino annihilation, each neutrino species (with antineutrinos) contributes 1 to $N_{\text{eff}}$, making the total relativistic energy density,

$$\rho_{\text{rel}} = \rho_\gamma \left[ 1 + \frac{7}{8} \left( \frac{4}{11} \right)^{3/2} \right].$$

Later decays increase the temperature of the plasma relative in energy density relative to the plasma. Decay happens when the two-photon interaction becomes effective, $T_{\text{de}} > m_\phi$, there is an out-of-equilibrium decay. In this case, upon becoming non-relativistic ALPs cease to decouple via the two-photon interaction. The details of the time ordering in which ALPs 1) freeze-out from the plasma while the ALPs are decoupled, to a large degree only these two in- and out-of-equilibrium scenarios are important.

In discussing the cosmological impact of ALPs, it is useful to define two quantities. The first is the effective
where $n_{ep}$ is the number density of electrons and positrons, $E$ is the ALP energy, and $m_{\gamma} = eT/3$ is the plasmon mass in an electron-positron plasma [16]. As shown by Cadamuro and Redondo [11], an accounting of other charged particles gives an approximation for the Primakoff freeze-out temperature of

$$T_{fo} \approx 123 \left(\frac{g_{*}(T_{io})}{g_{t}(T_{io})}\right) \left(\frac{10^{-9}\text{GeV}^{-1}}{g_{\phi\gamma}}\right)^{2} \text{GeV},$$

where $g_{*}$ is the number of relativistic degrees of freedom in the plasma, and $g_{t}$ the number of charged relativistic degrees of freedom [17].

The two-photon scattering rate, $C_{\gamma}$, can be calculated exactly at all times [11],

$$C_{\gamma} = \frac{m_{\phi}}{E\tau_{\phi\gamma}} \left[ \frac{2T}{p} \ln \frac{E+T}{p} - \frac{E}{T} \ln \frac{E+T}{T} \right].$$

Here we have ignored the plasmon mass, which inhibits decays when $m_{\phi} < 2m_{\gamma}$, as this effect is not important in the parameter range of interest. We have also assumed that photon rethermalization is instantaneous, which is an excellent approximation at early times. We consider thermalization at later times in Sec. III A 1.

In practice, we begin evolving the Boltzmann equations at a temperature $T_{0}$ when electrons and positrons are the only remaining charged particles and we can thus use Eqn. 6 for the Primakoff scattering rate. If the Primakoff process has yet to freeze out at $T_{0}$, as given by Eqn. 7, we take as the initial conditions for $f_{\phi}$ a Bose-Einstein distribution at temperature $T_{0}$. If freeze-out is earlier, we use the conservation of comoving entropy to calculate the increase in photon temperature after Primakoff freeze-out. The result is that ALPs are now at a reduced relative temperature of $T_{0}(g_{*}(T_{0})/g_{*}(T_{fo}))^{1/3}$.

We now describe key regions and contours in the ALP parameter space using Fig. 1 as a guide. An important quantity governing the ALP evolution is the temperature, $T_{re}$, at which the two-photon interaction re-equilibrates. Eqn. 5 shows that the ALP distribution $f_{\phi}$ at a given momentum $p$ becomes equal to its equilibrium value $f^{eq}_{\phi}$ when the scattering rate, $C_{\gamma}(p)$, is in the order of the Hubble constant. Now consider only those momenta which contribute dominantly to the total energy density, since this is the quantity we are interested in. For $m_{\phi} \gtrsim T$, these are $p \approx 0$. In this case, $C_{\gamma}$ vastly simplifies to $1/T_{re}$, and thus recoupling occurs when the Hubble time reaches the lifetime. For $m_{\phi} \lesssim T$, the important momenta are instead $E \approx p \approx T$. Ignoring the term in brackets in Eqn. 8 which can be shown to depend only logarithmically on temperature for these momenta, and using $H = 1.66\sqrt{g_{*}T^{2}}$, we arrive at (in Planck units),

$$T_{re} = \left( \frac{1}{\tau_{\phi\gamma} 1.66\sqrt{g_{*}}} \right)^{1/2} \quad m_{\phi} > T_{re}$$

$$T_{re} = \left( \frac{m_{\phi}}{\tau_{\phi\gamma} 1.66\sqrt{g_{*}}} \right)^{1/3} \quad m_{\phi} < T_{re}$$

where $H$ is the Hubble constant. Now consider only those momenta which contribute dominantly to the total energy density, since this is the quantity we are interested in. For $m_{\phi} \gtrsim T$, these are $p \approx 0$. In this case, $C_{\gamma}$ vastly simplifies to $1/T_{re}$, and thus recoupling occurs when the Hubble time reaches the lifetime. For $m_{\phi} \lesssim T$, the important momenta are instead $E \approx p \approx T$. Ignoring the term in brackets in Eqn. 8 which can be shown to depend only logarithmically on temperature for these momenta, and using $H = 1.66\sqrt{g_{*}T^{2}}$, we arrive at (in Planck units),

$$T_{re} = \left( \frac{1}{\tau_{\phi\gamma} 1.66\sqrt{g_{*}}} \right)^{1/2} \quad m_{\phi} > T_{re}$$

$$T_{re} = \left( \frac{m_{\phi}}{\tau_{\phi\gamma} 1.66\sqrt{g_{*}}} \right)^{1/3} \quad m_{\phi} < T_{re}$$

Fig. 1 shows constant $T_{re}$ contours in black for three values of $T_{re}$ corresponding to neutrino decoupling, and the start and end of BBN. Here and throughout this paper, we take these temperatures to be $1\text{MeV}$, $200\text{KeV}$, and $20\text{KeV}$ respectively. Fig. 1 also shows constant mass contours at these three values. The line implicitly formed by $T_{re} = m_{\phi}$ divides regions NEQ and EQ where the ALP becomes non-relativistic before or after the two-photon interaction re-equilibrates, respectively. Note that in region NEQ we have $m_{\phi} > T_{re}$, which is only part of the requirement for an out-of-equilibrium decay. The other is that $T_{io} > m_{\phi}$, however it turns out that this is always satisfied for any combination of mass and lifetime in region NEQ; that is to say, ALPs can never decay by the Primakoff process alone.

The evolution of the energy densities in the relevant components of the plasma for a typical in-equilibrium decay (region EQ) is shown in the third panel of Fig. 2. If ALPs decay after neutrinos are decoupled, conservation of comoving entropy implies that the temperature
of the neutrinos relative to the photons after the ALP decay and after electrons and positrons have annihilated is $(4/13)^{1/3}$, as compared to $(4/11)^{1/3}$ in the standard scenario. This means that $N_{\text{CMB}}^{\text{eff}}$ will be reduced by $(13/11)^{4/3}$ and, at fixed $\eta_{\text{CMB}}$ the value of $\eta$ prior to the decay is increased by a factor of 13/11. Assuming three neutrinos, this gives $N_{\text{CMB}}^{\text{eff}} \approx 2.44$. Because the exact timing with respect to electron-positron annihilation is unimportant for the final temperature, the entirety of the photons (or equivalently the decrease in $N$) are of interest. Thus, contours (Sec. III A 1) depend on the fractional energy injection after a certain temperature are $\eta_{\text{CMB}}$. The BBN data, however, are sensitive to the exact time of decay via sensitivity to $N_{\text{CMB}}^{\text{BBN}}$ and $\eta_{\text{BBN}}$. Since decay time when in equilibrium is controlled only by $m_\phi$, we find characteristic constant-mass contours in this region in the BBN constraints in Figs. 3, 4, and 5.

The second panel of Fig. 2 shows instead a typical out-of-equilibrium decay (region NEQ). In this region an important quantity is the fractional increase in the energy of the photons (or equivalently the decrease in $N_{\text{CMB}}^{\text{eff}}$) once the ALP decays at $t \sim \tau_{\phi \gamma}$. Because the ALP here is non-relativistic, this is,

$$\frac{1}{N_{\text{CMB}}^{\text{eff}}} \sim \frac{\rho_\phi}{\rho_\gamma} \bigg|_{t=\tau_{\phi \gamma}} \sim \frac{m_\phi a^{-3}}{a^{-4}} \bigg|_{t=\tau_{\phi \gamma}} \sim m_\phi \sqrt{\Phi_{\text{CMB}}} \quad (10)$$

where we have assumed that the ALP does not remain non-relativistic for too long before decay hence the universe is radiation dominated and $a(t) \sim \sqrt{t}$. This assumption is true for models right on the edge of the allowed region, thus we find characteristic contours of constant $m_\phi \sqrt{T_{\gamma \gamma}}$ for cosmological constraints in region NEQ as seen in Figs. 3, 4, and 5.

As we will see, another set of important characteristic contours are those of constant “fractional energy injection after a certain temperature”. For example, CMB and BBN constraints are only sensitive to energy injected after the temperature at which neutrinos decouple, as any earlier injection is “invisible” because it is rethermalized among all components. Bounds from CMB spectral distortions (Sec. III A 1) depend on the fractional energy injected only after the temperature when photon number-changing reactions freeze-out. In region EQ, the fractional energy injected does not depend on either mass or lifetime, and the amount after a certain temperature is controlled only by the mass. In region NEQ, the story is slightly more subtle. Similarly as in Eqn. 10, we are looking to evaluate,

$$\frac{\rho_\phi}{\rho_\gamma} \bigg|_{t=\tau_{\phi \gamma}} \sim \frac{m_\phi a^{-3}}{a^{-4}} \bigg|_{t=\tau_{\phi \gamma}}$$

except now at some $T = T_*$ rather than at the decay time. Because of the ALP decay, however, it is not the case that $a \sim 1/T$, and in fact the second Friedmann equation shows that the correction due to the decay gives a factor which exactly cancels the factor of $m_\phi$ in Eqn. 11, leaving only sensitivity to $T_{\phi \gamma}$. Essentially, more massive ALPs lead to more fractional energy injection, but delay the time it takes to reach a certain temperature, leaving the same amount of energy injected after that temperature. Thus, contours of constant fractional energy injection after a certain temperature are $m = \text{const}$ in region EQ and $\tau = \text{const}$ in region NEQ. The final key region in Fig. 1, region EARLY, is delineated by such a contour, and corresponds to zero energy injection after neutrino decoupling and thus no current cosmologically observable imprints.

The Primakoff freeze-out temperature plays a smaller role in the ALP evolution than the recoupling temperature, although we briefly mention two effects stemming from events happening in the gap between freeze-out and recoupling (in regions of parameter space where the gap exists). We first note that if no reheating of the plasma occurs in this gap and ALPs are relativistic, they still track their equilibrium abundance even though they are decoupled. One possible reheating occurring during the gap is the QCD phase transition which imprints a sharp feature our constraints. During this transition the temperature of the ALPs roughly doubles along with the rest of the plasma if they are still coupled, leading to an energy density roughly twenty times larger. Another possibility is reheating from electron-positron annihilation, a case in which the ALPs can re-equlibrate relativistically thereafter. Such a scenario happens near the top-left part of Fig. 1 and is shown in the fourth panel of Fig. 2. For deuterium and helium constraints shown in Fig. 5, near this region we can find both lines of constant $T_{\gamma \gamma}$ and $T_{\phi \gamma}$.

### III. CONSTRAINTS

#### A. Cosmic Microwave Background

1. Frequency Spectrum

The measurement of the CMB frequency spectrum by COBE/FIRAS places very tight bounds on spectral distortions away from a black-body spectrum [18], limiting possible energy injection into the plasma [19].

The effects of energy injection depend crucially on when it occurs, with the time-line roughly divided into three eras. In the earliest era, reactions that change photon number are fast and any injection of photon energy is quickly rethermalized. This leads to only an adjustment of the temperature, hence this is called the “T era”. Decays during this era do not cause any distortions in the frequency spectrum observed today and are thus allowed. At about $T \approx 750$ eV, double Compton (DC) and Bremsstrahlung (BR) reactions become too slow to significantly change photon number. Injected photons now lead to a spectrum with a chemical potential, giving this the name the “$\mu$ era”. Finally, Compton scattering freezes out at $T \approx 25$ eV, making even the redistribution of energy inefficient. Initially we are left with a Comptonized spectrum described by a $y$ parameter, and at even later times photons from decay can be seen directly, contributing to the extra-galactic background. We do not
consider $y$ distortion bounds in this work as they rule out only very late decays which are outside of the parameter range of interest.

As shown by Hu and Silk [20], if $\mu(z) \ll 1$ for the duration of the energy injection and if Compton scattering is fast compared to DC and BR, the evolution of $\mu$ obeys,

$$
\frac{d\mu}{dt} = \frac{dp_s}{dt} - \mu \left( \frac{1}{t_{DC}} + \frac{1}{t_{BR}} \right) \tag{12}
$$

where $1/t_{DC}$ and $1/t_{BR}$ are the scattering rates given by,

$$
t_{DC} = \frac{1.9 \times 10^{35}}{\Omega_b h^2 (1 - Y_p / 2)} \left( \frac{T}{\text{Kelvin}} \right)^{-9/2} \tag{13}
$$

$$
t_{BR} = \frac{8.9 \times 10^{26}}{(\Omega_b h^2)^{3/2} (1 - Y_p / 2)} \left( \frac{T}{\text{Kelvin}} \right)^{-13/4} \tag{14}
$$

and $dp_s/dt$ is a source term given by,

$$
dp_s = \frac{3\pi^4 \zeta(3)}{2\pi^6 - 810\zeta(3)^2} \left( 3 \frac{dp_\phi}{\rho_\gamma} - 4 \frac{dn_\gamma}{n_\gamma} \right) \tag{15}
$$

$$
\approx \frac{1}{2.14} \left( 3 \frac{dp_\phi}{\rho_\gamma} - 4 \frac{dn_\gamma}{n_\gamma} \right) \tag{16}
$$

which is the chemical potential induced by adding a comoving photon energy density $d\rho_\gamma$ and number $dn_\gamma$. Since ALPs decay to two photons, we have $d\rho_\gamma = -d\rho_\phi$ and $dn_\gamma = -2dn_\phi$.

In practice, we first evolve Eqn. 5, which assumes instantaneous thermalization. Then we take the resulting solution for $d\rho_\phi$, $dn_\phi$, and expansion history, and numerically integrate Eqn. 12 to arrive at the actual thermalization history. This procedure is correct to first order in the energy injection. We then compare with the COBE/FIRAS data which gives a 95% upper bound of,

$$
|\mu| < 9 \times 10^{-5} \tag{17}
$$

The excluded region is shown in orange in Fig. 3 and is labeled “CMB Spectral Distortions”. Because Eqn. 15 shows the source term for $\mu$ depends on the fractional energy density injected into the plasma after DC and BR freeze-out, the arguments of Sec. II imply that the exclusion region should be one of constant decay time, and indeed it is parallel to $\nu$-dec and BBN start/end lines. The exact division between allowed and disallowed scenarios is close to the transition from the $T$ to the $\mu$ era, as expected because almost any chemical potential is ruled out while almost any injection in the $T$ era is unobservable. These same constraints were calculated by Cadamuro and Redondo [11] under a further approximation of Eqn. 12 where DC and BR are taken to be infinitely fast until they instantaneously freeze-out at $T \approx 750$ eV. In the “previous constraints” panel of Fig. 4 we reproduce their result, showing that qualitatively this is a very good approximation. The use of Eqn. 12 becomes more important, however, for constraining scenarios with even smaller chemical potentials generated deeper into what is currently called the $T$-era.
future missions which are predicted to reach such sensitivity are PIXIE and COrE [21, 22], for which we give forecasts in Sec. IV D.

2. Angular Power Spectrum

Measurements of CMB anisotropies have been recently improving, both from the ground [23–26] and from space [27, 28]. Better angular resolution and lower noise have tightened up small-scale constraints where the CMB is most sensitive to changes in $N_{\text{eff}}$ and $Y_p$, both of which
are altered by the decay of ALPs.

A fully general treatment would include ALPs in the set of Boltzmann equations for calculating the CMB power spectrum, but it turns out this is not necessary for the scales that are currently well measured. For these scales, all of the physical effects of ALPs are in fact identical to changes in $N_{\text{eff}}$ and $Y_p$, as long as we assume adiabatic initial conditions. This is essentially because decays must happen early enough, as enforced by the spectral distortion bound discussed in the previous section, which requires the decay happen by $T \approx 750\, \text{eV}$ or equivalently $z \approx 3 \times 10^6$. The angular scales well constrained by CMB measurements, roughly $\ell \lesssim 3000$, correspond to physical scales which do not begin to enter the horizon until about $z \approx 3 \times 10^6$. At this point two things are different in the ALP scenario as opposed to the standard case 1) the amplitude of neutrino density perturbations upon horizon entry is reduced relative to the photons and 2) the expansion rate is different. However, both are exactly captured in the standard scenario by changing $N_{\text{eff}}$. No other scale-dependent changes to spatial perturbations are possible because the relevant scales are still outside of the horizon by the time of the decay. Finally, we note that the altered helium abundance is taken as an input to the CMB spectrum calculation, and in the ALP scenario it is now just at a different value.

In their work, Cadamuro and Redondo [11] took as a CMB constraint a lower bound on $N_{\text{eff}}$ as given by WMAP7. The Planck data tighten this constraint and are also sensitive to $Y_p$. We use the joint constraint on these two parameters given by the combination of Planck-WP+highL from [27], approximating the likelihood as Gaussian and taking just the mean and covariance. In practice, we first calculate $N_{\text{eff}}^{\text{CMB}}$ and $Y_p$ for a given mass and lifetime (here ignoring uncertainties in $Y_p$ due to $\eta$ and nuclear reaction rate uncertainties which are unimportant at the level of the CMB constraint). We then calculate the $\chi^2$ of these values against the mean and covariance of $N_{\text{eff}}^{\text{CMB}}, Y_p$ from the Planck chain. Masses and lifetimes excluded at $> 3\sigma$, or equivalently $\chi^2 > 11.8$, are shown in cyan in Fig. 3 and labeled “CMB Anisotropies.” This region covers a large part of the parameter space corresponding to out-of-equilibrium decays. These decays tend to greatly reduce $N_{\text{eff}}^{\text{CMB}}$ because ALPs can live long enough to become non-relativistic without becoming Boltzmann suppressed, thus greatly increasing their total energy injection. The majority of the remaining parameter space corresponds to in-equilibrium decay, which is allowed because here there is a lower limit of $N_{\text{eff}}^{\text{CMB}} \approx 2.44$ as discussed in Sec. II, compatible with CMB constraints. Fig. 6 shows typical values for $N_{\text{eff}}^{\text{CMB}}$ and $Y_p$ in the ALP scenario as compared to the CMB constraints.

**B. Primordial Abundance Inferences**

The primordial production of the light nuclides during BBN occurs from $t \sim 1\, \text{sec}$ to $\sim 3\, \text{minutes}$, corresponding to $T \sim 1\, \text{MeV}$ to $100\, \text{keV}$ [reviewed recently in, e.g., 30–32]. Observations of primordial light element abundances thus probe new physics at play during this epoch [e.g., 33]. For our case of an ALP, light element production is affected by changes to the cosmic expansion rate during BBN and to the extrapolation of $\eta$ back from the CMB epoch. For the majority of the parameter space where ALPs decay after neutrino decoupling, they are still present during BBN and hence increase $N_{\text{BBN}}^{\text{CMB}}$. However, if ALPs are massive enough, they could decay before BBN, leaving a large part of the parameter space corresponding to out-of-equilibrium decays. These decays tend to greatly reduce $N_{\text{eff}}^{\text{CMB}}$ because ALPs can live long enough to become non-relativistic without becoming Boltzmann suppressed, thus greatly increasing their total energy injection. The majority of the remaining parameter space corresponds to in-equilibrium decay, which is allowed because here there is a lower limit of $N_{\text{eff}}^{\text{CMB}} \approx 2.44$ as discussed in Sec. II, compatible with CMB constraints. Fig. 6 shows typical values for $N_{\text{eff}}^{\text{CMB}}$ and $Y_p$ in the ALP scenario as compared to the CMB constraints.
FIG. 5. The colored contours show the prediction for each of the labeled quantities as a function of different values of ALP mass and lifetime. The dotted/dashed/solid lines give 1/2/3 $\sigma$ contours given the measurements for these quantities discussed in Sec. III. No lines are visible on the lithium plot because the entire parameter space is excluded at $>3\sigma$ (our scenario does not alleviate the lithium problem). We do not give contours for the $N_{\text{CMB}}^\text{eff}$ plot because the CMB constraint is highly degenerate with $Y_p$. For the D/H panel, the colored contours are calculated assuming a best-fit $\eta$ from the CMB, and uncertainties in $\eta$ and nuclear reaction rates are taken into account in producing the $\sigma$ contours (see Sec. III B for discussion).

and the expansion rate. Additionally, because the decay decreases $\eta$, fixing $\eta_{\text{CMB}}$ to the observed value generally leads to an increased $\eta_{\text{BBN}}$. Both effects serve to increase $Y_p$, but partially cancel for D/H and $^7$Li. The increase in $N_{\text{BBN}}^\text{eff}$ increases D/H while the increase in $\eta$ has the opposite effect, with the latter about twice as large, leaving an overall reduced D/H. For $^7$Li it is instead the former which wins out. The light-element trends in the mass-lifetime planes of Fig. 5 bear out these expectations, as we now see in detail.

In practice we have modified the AlterBBN code of Arbey [34] to include changes to the expansion history and $\eta$ due to ALPs. Our code assumes the photon spectrum is instantaneously rethermalized, in effect ignoring the possibility that high energy photons from the decay can break apart already-formed nuclei. Bounds due
to this phenomenon constitute so called “photo-erosion” bounds, discussed in e.g. [35] and references therein. We will consider them separately at the end of this section.

Deuterium is observable at \( z \sim 3 \) in QSO absorption systems, via the \( \sim 82 \text{ km/s} \) isotope shift between D and H Lyman absorption lines. Recent D/H measurements have been reported [36],
\[
\frac{D}{H} = (2.53 \pm 0.04) \times 10^5, \tag{18}
\]
which represents a factor \( \sim 3 \) improvement in precision. These bounds are now so tight as to place the measurement errors on level footing with uncertainties associated with nuclear reaction rates and with a determination of \( \eta \) from the CMB.

For convenience, constraints on deuterium are often given with the uncertainty from nuclear reaction rates marginalized into the error bars, with the implicit instruction to be compare to theoretical predictions calculated assuming the best-fit values of the rates. One can similarly also marginalize over the uncertainty from \( \eta \), which we now show how to perform. Because the ALP scenario changes \( N_{\text{eff}} \) which is degenerate with \( \eta \) within the CMB data, extra care needs to be taken that with respect to these quantities (there are no such analogous parameters within the nuclear reaction rate data).

We begin by considering the joint likelihood for the CMB, D/H, and nuclear reaction rate measurements, which can be written as
\[
-\log L = \frac{(DH(m_{\phi}, \tau_{\gamma}, \eta, \alpha_i) - DH)^2}{2\sigma_{\text{MEAS}}^2} + \log L_{\text{CMB}}(\eta, N_{\text{eff}}, \Omega) + \log L_{\text{NUCL}}(\alpha_i, \Omega') \tag{19}
\]
where \( DH \pm \sigma_{\text{MEAS}} = (2.53 \pm 0.04) \times 10^5 \) as per Eqn. 18, \( \eta \) and \( N_{\text{eff}} \) are evaluated at the CMB epoch but we omit the label for brevity, \( N_{\text{eff}} = N_{\text{eff}}(m_{\phi}, \tau_{\gamma}) \) is uniquely set by the mass and lifetime, \( \alpha_i \) are parameters describing the nuclear reaction rates, and \( \Omega \) and \( \Omega' \) are any remaining cosmological and nuisance parameters. \( N_{\text{eff}} \) is important and appears explicitly because it is both dependent on the ALP parameters and its measurement from the CMB is significantly degenerate with \( \eta \). We next analytically marginalize over all parameters other than \( m_{\phi} \) and \( \tau_{\gamma} \) under the assumption that these other parameters have Gaussian posterior likelihoods and that D/H depends linearly on \( \eta \). This gives
\[
-\log L = \frac{(DH(m_{\phi}, \tau_{\gamma}, \bar{\eta} + r \sigma_{\eta} N_{\text{eff}} - N_{\text{eff}}, \bar{\eta}_i) - DH)^2}{2[\sigma_{\text{MEAS}}^2 + \sigma_{\text{NUCL}}^2 + \sigma_{\text{ETA}}^2]} \tag{20}
\]
with
\[
\sigma_{\text{ETA}}^2 = \left( \frac{dD}{d\eta} \sigma_{\eta} \right)^2 (1 - r^2) \tag{21}
\]
where \( \bar{\eta} \pm \sigma_{\eta} \) and \( \bar{N}_{\text{eff}} \pm \sigma_{N_{\text{eff}}} \) are the mean and standard deviation of the posterior likelihoods from the CMB with all other parameters marginalized over, and \( r \) is the correlation coefficient between \( \eta \) and \( N_{\text{eff}} \). The presence of \( r \) in this equation can be understood by considering the (unrealistic) \( r = 1 \) case, which would imply that CMB measurements could turn a fixed \( N_{\text{eff}} \) into a perfect determination of \( \eta \); the quantity above at which the D/H prediction is evaluated, \( \bar{\eta} + r \sigma_{\eta}(N_{\text{eff}} - \bar{N}_{\text{eff}})/\sigma_{N_{\text{eff}}} \), is the mean of this determination. Because in our case \( N_{\text{eff}} \) is fixed by the mass and lifetime, it would mean \( \eta \) is also fixed, leading to no extra uncertainty in D/H. In reality, from the Planck chains we find the two parameters are correlated at \( r \approx 0.4 \). We make one further approximation which is that neither the D/H derivative nor the nuclear reaction rate uncertainty depends on the values of mass and lifetime or the fact that \( \eta \) evolves with time in the ALP scenario, which we have checked is sufficient. We find \( \sigma_{\text{NUCL}} = 4.5 \times 10^{-7} \) using AlterBBN which takes the \( \alpha_i \) to be principal components in the nuclear reaction rate parameter space [37]. Numerically evaluating the D/H derivative and taking posterior likelihoods from Planck+WP+highL, we find \( \sigma_{\text{ETA}} = 6.9 \times 10^{-7} \). When added in quadrature these lead to an effective deuterium constraint of
\[
\frac{D}{H} = (2.53 \pm 0.091) \times 10^5 \tag{22}
\]
which is meant to be compared to a theoretical prediction calculated assuming the best-fit nuclear reaction rates and $\eta = \bar{\eta} + r_{\gamma d}(N_{\text{eff}} - N_{\text{eff}}^{\text{BBN}})/\sigma_{\text{BBN}}$.

The effects of these D/H constraints on the $(m_\phi, \tau_\gamma)$ plane appear in Fig. 5. We see that the effect of an ALP is always to decrease D/H due to the ALP’s effective increase of $\bar{\eta}_{\text{BBN}}$ winning out over the increase in $N_{\text{eff}}^{\text{BBN}}$. Moreover, we see that the high precision of the D/H measurements leads to a tight constraint on the ALP space in all regions where the decays occur after neutrino decoupling. Indeed, D/H is now a very powerful probe of ALPs.

The primordial $^4\text{He}$ abundance is inferred astronomically from observations of emission spectra of highly ionized gas in primitive nearby dwarf galaxies, i.e., in low-metallicity extragalactic HII regions. The primordial abundance is traditionally inferred by extrapolation to zero metallicity. To derive helium and metal abundances from the observed spectra requires characterization of the thermodynamic properties of the emitting gas (i.e., temperature, density). The analysis of [40] derives these quantities simultaneously in a self-consistent manner, and finds a primordial abundance

$$Y_p = 0.2465 \pm 0.0097$$

where the uncertainty is quantified with an MCMC analysis. We adopt this as our fiducial primordial $^4\text{He}$ constraint. Using a similar data set, Izotov and Thuan [38] give a helium constraint of, $Y_p = 0.254 \pm 0.003$ is derived in a less conservative manner. In Fig. 5 we see that, as expected, the effect of an ALP is to increase $Y_p$ for almost all of the parameter space where the decays occur after neutrinos decouple. However, the constraints are not as strong as those of D/H. There is an island of parameter space around $m_\phi > 100$ MeV and $\tau_\gamma \sim 1$ sec where $Y_p$ decreases. This region corresponds to decays happening between neutrino decoupling and the start of BBN. Here we have $\bar{\eta}_{\text{BBN}} = \eta_{\text{CMB}}$ and $N_{\text{eff}} < 3$, the latter of which serves to decrease $Y_p$. It is interesting to note this low helium region does not extend along the entire constant decay-time contour, cutting off once we enter the in-equilibrium decay side. This occurs because in-equilibrium decays reduce the ALP energy density more slowly than do out-of-equilibrium ones, and thus the entire decay cannot fit in the short time between neutrino decoupling and BBN.

Finally, the primordial $^7\text{Li}$ abundance is inferred from observations of the atmospheres of low-metallicity (extreme Population II) stars in the Galactic stellar halo. Down to a metallicity of $(\text{Fe/H}) \sim 10^{-2.8}(\text{Fe/H})_{\odot}$, these stars have lithium abundances that are the same to within a small scatter consistent with observational errors. The independence of Li with Fe in this “Spite plateau” indicates that lithium is primordial [39], and implies a primordial abundance

$$\frac{\text{Li}}{\text{H}} = (1.6 \pm 0.3) \times 10^{-10}$$

[40]. At lower metallicity, however, the Li/H abundance scatter increases dramatically, but always below the Spite plateau value. This suggests that in these very metal-poor stars some lithium destruction has occurred; the reason for this remains unclear.

The astronomically-inferred lithium abundance in Eqn. 24 is inconsistent with the primordial value expected from standard BBN theory combined with CMB determinations of $\eta$. The observed Li/H value is low at the $\sim 5\sigma$ level. This is the “lithium problem” [reviewed in, e.g., 32]. Stellar astrophysics uncertainties may be the origin of the problem, but solutions to date require fine tuning and do not explain the observed Li/H “meltdown” at very low metallicities. A more radical and intriguing solution is the presence of new physics during or after BBN. The challenge for such scenarios is to reduce lithium substantially without drawing other light elements—particularly deuterium—from their concordant primordial abundances. The ALP scenario tends to aggravate the problem by increasing lithium slightly, as seen in Fig. 5, and as expected due to the ALP effect on $\eta$. Thus, awaiting a resolution to the lithium problem, we do not consider lithium bounds.

In closing we note that our calculations have neglected the effects of photoerosion of the light elements. This occurs 1) when the ALP mass exceeds light-element binding energies $m_\phi \gtrsim B \sim 10$ MeV, and 2) for decay time scales long enough so that the decay photons interact with light elements before thermalization. This leads to some deuterium destruction via $\gamma d \rightarrow np$, but a net production due to e.g., $\gamma ^4\text{He} \rightarrow dd$. Thus constraints arise from D/H, $Y_p$, and $^4\text{He}/^4\text{He}$ [41–43]. These were recently computed for purely electromagnetic decays by Cyburt et al. [35], assuming that the decay photons provide a negligible contribution to the energy density and thus expansion rate. In this case, the constraints are only important for $\tau_X \gtrsim 10^4$ sec and $m_X \gtrsim 10^5$ eV, with $X$ the decaying particle. This regime shows a $Y_p$ drop due to photoerosion and a corresponding D/H increase. These trends could potentially bring $Y_p$ and D/H predictions back into agreement with observations, but would require a more detailed calculation. Since the regions of parameter where this can happen are already ruled out by CMB observations, we ignore the effects of photo-erosion.

C. Laboratory

Laboratory bounds on ALPs come from a variety of different experimental setups. At lower masses, roughly $m \lesssim eV$, some examples include photo-regeneration experiments (“shining light through walls”), microwave cavities, and helioscopes [for a review, see 12–14]. For the larger masses considered here, the best constraints come from electron-positron colliders and beam dumps.

The presence of the ALP-photon interaction allows for the possibility of single-photon final states at electron-positron colliders. Early interpretation of searches for
these events in terms of constraints on the ALP coupling $g_{\phi \gamma}$ was done by Massó and Toldrà [8]. Both Kleban and Rabadan [44] and Mimasu and Sanz [15] have further shown the ability of current and future colliders to improve these bounds. Here we reproduce the constraint from LEP given by Mimasu and Sanz [15] of,

\[ g_{\phi \gamma} < 4.5 \times 10^{-4} \text{ GeV}^{-1}, \]  

valid in the entire lifetime range considered here. The excluded region is labeled “Collider” and shown in magenta in Fig. 3.

Additional constraints come from beam dump experiments, where ALPs would be produced in the beam dump, penetrate through shielding, then decay to photons which can be detected by a downstream detector. We use the constraints from Bjorken et al. [45] which find that at 95% confidence,

\[ m_\phi \tau_{\phi \gamma} > 1.4 \text{ keV sec.} \]  

This is labeled “Beam Dump” and shown in red in Fig. 3.

D. Globular Clusters and SN1987A

ALPs offer a new means for energy loss from stars if they can both be produced in stellar interiors and have sufficiently weak interaction strengths to subsequently escape. In the case of SN1987A, the energy loss can affect the duration of the neutrino pulse from the handful of neutrinos which were detected, placing constraints on the ALP interaction strength [8, 9]. These bounds are reproduced in Fig. 3. Energy loss can also affect the duration of the red giant phase and of the horizontal branch, leading to a different observed ratio of such stars in globular clusters. We use the bounds from Cadamuro and Re-dondo [11] based on arguments of Raffelt and Dearborn [46] and Raffelt [47]. These are also reproduced in Fig. 3.

IV. DISCUSSION

A. The MeV-ALP Window

An interesting feature of the exclusion regions prior to this work is the allowed window bounded on all sides near \( m \sim 1 \text{ MeV} \) and \( \tau \sim 100 \text{ ms} \) corresponding to an ALP decay during BBN. We will henceforth call this the MeV-ALP window. It can be seen in the left panel of Fig. 4 as well as in Hewett et al. [12], Mimasu and Sanz [15]. Further interest is driven by the fact that a particle in this window could actually be a DFSZ axion, in which case its symmetry breaking scale is close to the electroweak scale. One of the main conclusions of this work is to show that this region is now, in fact, ruled out by the combination of CMB+D/H measurements.

We first briefly review two relevant generic axion models, referred to as the KSVZ and DFSZ models. Both models introduce a new global \( U(1) \) symmetry which is approximately broken at some energy scale \( f_\phi \) giving rise to an axion with mass \( m_\phi \). The symmetry breaking scale is related to the axion mass by non-perturbative effects and given by

\[ m_\phi = \frac{\sqrt{2}}{1 + z} \frac{m_\pi f_\pi}{f_\phi}, \]  

where \( m_\pi \) is the pion mass, \( f_\pi \) its decay constant, \( z = m_u/m_d \) the ratio of up to down quark masses. Axion models differ in what other new fields are introduced to implement the symmetry breaking and how these, as well as standard model fields, transform under the new \( U(1) \). The KSVZ model [48, 49] has the standard model fermions neutral, whereas in the DFSZ model [50, 51] they can carry \( U(1) \) charge. These model dependent choices in turn affect the axion’s effective photon coupling which arises from fermion loops, ultimately leading to a consistency relation between axion mass and photon lifetime which can be written as,

\[ \tau_{\phi \gamma} = \left( \frac{E}{N} - \frac{2(4 + z)}{3(1 + z)} \right)^{-1} \frac{\sqrt{\pi}}{1 + z} \frac{16\pi^{3/2}}{f_\pi m_\pi} \frac{m_\phi}{\alpha}^{2} \]  

where \( E \) and \( N \) are the electromagnetic and color anomalies of the axial current associated with the axion, and can vary according to specific axion model. In evaluating this relation, we will adopt fixed values of \( m_\pi = 135 \text{ MeV}, f_\pi = 92 \text{ MeV}, \) and \( z = 0.56 \), ignoring small uncertainties that lead to roughly a 10% uncertainty in the axion mass [10, 52].

The KSVZ model has \( E/N = 0 \), so that we have,

\[ \tau_{\phi \gamma} = 6.20 \times 10^{4} \left( \frac{m_\phi}{\text{eV}} \right)^{-5} \]  

The particular DFSZ model we consider here has \( E/N = 2 \), hence we refer to it as the “DFSZ-EN2” model, with consistency relation,

\[ \tau_{\phi \gamma} = 2.66 \times 10^{5} \left( \frac{m_\phi}{\text{eV}} \right)^{-5} \]  

These two consistency relations are shown as the dashed lines in Fig. 3. Other values for \( E/N \) are possible, but the DFSZ-EN2 has the distinction of having a particularly weak coupling because \( E/N \approx 2(4 + z)/3(1 + z) \) and so these terms nearly cancel in Eqn. 28 [53, 54]. This weaker coupling means that the DFSZ-EN2 is consistent with the collider bounds over a larger range of masses as compared to the KSVZ. Ultimately it is that the consistency relation passes through the MeV-ALP window which motivates our interest in this model. The lower mass limit for the DFSZ-EN2 in the MeV-ALP window is around \( m_\phi \sim 200 \text{ keV} \), corresponding to \( f_\phi \sim 30 \text{ GeV} \), less than an order of magnitude from the electroweak scale \( v_{\text{weak}} \sim 246 \text{ GeV} \) where the axion was initially thought to lie.
While it is interesting that this mass range for the DFSZ-EN2 was previously allowed, this part of parameter space for the DFSZ-EN2, and more generally the entire MeV-ALP window, is now ruled out by the combination of CMB+D/H data. This region corresponds to in-equilibrium decays hence it gives $N_{\text{eff}}^{\text{CMB}} = 2.44$. The decay happens essentially in the middle of BBN, increasing $N_{\text{eff}}^{\text{BBN}}$ and $\eta_{\text{BBN}}$, which, as discussed previously, increases $Y_p$ and decreases D/H. The decrease in $N_{\text{eff}}^{\text{CMB}}$ and increase in $Y_p$ moves along the degeneracy direction for CMB measurements, and is allowed even by our updated CMB constraints coming from Planck (see Fig. 6). It is in combination with the D/H constraints that the MeV-ALP window is closed, with the best fitting model within the window ruled out at about 3.5σ. If we replace the CMB constraint from Planck with previous measurements from the combination of WMAP, ACT, and SPT, the window is ruled out at a similar significance. This is despite the 20% tighter $\eta$ constraint from Planck because the central value also shifts lower, increasing D/H back towards the measured value. Conversely, replacing the D/H measurement with previous bounds does open the MeV-ALP window again, as seen in the left panel of Fig. 4. It is thus the new bounds from Cooke et al. [36] that are the key improvement.

While the MeV-ALP region is now excluded by the CMB+D/H measurements, this result depends on the assumption of having no extra radiation besides neutrinos and the ALP. In some scenarios, for example as predicted by the string axiverse [7], it is natural to have many ALPs, some of which could also contribute to $N_{\text{eff}}$ but be light enough to remain otherwise invisible. Motivated by this possibility, we explore constraints when in addition to the ALP mass and lifetime, we also allow an arbitrary addition to $N_{\text{eff}}$. Sufficiently tight constraints around the standard value (black dot) could rule out the ALP scenario even in the presence of extra radiation.

B. A Loophole in the Presence of Extra Radiation

While the MeV-ALP region is now excluded by the CMB+D/H measurements, this result depends on the assumption of having no extra radiation besides neutrinos and the ALP. In some scenarios, for example as predicted by the string axiverse [7], it is natural to have many ALPs, some of which could also contribute to $N_{\text{eff}}$ but be light enough to remain otherwise invisible. Motivated by this possibility, we explore constraints when in addition to the ALP mass and lifetime, we also allow an extra arbitrary addition to $N_{\text{eff}}$.

The MeV-ALP region is ruled out largely because it predicts too low an abundance of primordial deuterium.
An addition to $N_{\text{eff}}^{\text{DBN}}$ increases D/H and can bring it back into agreement with measurements. The penalty is a further increase in $Y_p$, but because the helium constraints are not as tight as D/H, an allowed window now opens up again.

We explore this window with Markov Chain Monte Carlo (MCMC) [55]. We run two MCMC chains, one for the ALP case where both mass and lifetime are free parameters, and another for the DFSZ-EN2 case with only the mass free and the lifetime given by Eqn. 30. In both cases we also leave free the quantity we call $\Delta N_{\text{eff}}$ which controls any extra relativistic energy density at some early time before BBN, meaning that $N_{\text{eff}}^{\text{preBBN}} = 3 + 4/7 + \Delta N_{\text{eff}}$. The ALP contributes 4/7 because it is one bosonic degree of freedom and is fully thermalized in all regions of parameter space explored by the chain. The likelihood includes all of the bounds in Fig. 3. In either the ALP or DFSZ-EN2 cases, we find that the MeV-ALP window is again allowed and the best-fitting model consistent with all of the data at $\lesssim 1\sigma$.

The mass posterior distributions for the ALP and DFSZ-EN2 chains are shown in Fig. 7. In both cases masses below 200 keV are excluded. Masses above 1 MeV in DFSZ-EN2 case are excluded by the collider bound, but are allowed in the ALP case because models can evade this constraint by having a smaller photon coupling. We also show a forecast for a next generation electron-positron collider SUPER-KEKB after two years of integration (discussed in Sec. IV D) which can probe down to almost exactly the 200 keV minimum.

The corresponding likelihoods for $N_{\text{eff}}$ are given in the right panel. These show that the data accommodate the ALP scenario by initially having $N_{\text{eff}}^{\text{preBBN}} \sim 4.7$ and diluting this down to $N_{\text{eff}}^{\text{CMB}} \sim 3.4$ via the ALP decay. Constraints on the extra radiation are $\Delta N_{\text{eff}} = 1.13 \pm 0.30$. Thus, one or (marginally) two extra neutrino-like particles allow for an ALP in the MeV-ALP window. Alternatively, the decay of such an ALP can hide the existence of one or two additional neutrino-like particles from the tight CMB constraints which would otherwise rule them out. A similar loophole for extra radiation has been proposed by Ho and Scherrer [56].

We next test the extent to which the data prefer these extended models. We perform a simple test using best-fit $\chi^2$ values given in Tab. 11. If the $\chi^2$ for the extended model decreases significantly as compared to $\Lambda\text{CDM}$, then roughly that model is preferred. Although all bounds from Fig. 3 are included in the fit, we only give $\chi^2$ for those which are not hard cutoffs. When using the combination of Planck+D/H+$Y_p$, we find the baseline data choice slightly disfavors both the DFSZ-EN2 and ALP models. The only case where there is a preference for the extend model is in the ALP case when using the helium constraint from Izotov and Thuan [38]. Here we find an improvement in $\chi^2$ of 5.81 when we have added 3 new free parameters, something we expect to happen by chance only 12% of the time. If the high helium value inferred by Izotov and Thuan [38] is confirmed, then this scenario is a natural explanation as it can increase helium compared to the standard value while keeping the deuterium abundance and $N_{\text{eff}}^{\text{CMB}}$ roughly unchanged.

### C. A Simple Expression for Exclusion Bounds

Given the improved constraints from CMB+D/H measurements, we suggest a simple expression for ALP bounds which can be adopted by those who prefer a simpler picture than the many probes shown in Fig. 3. The CMB+D/H data alone now essentially rule out any energy injection after neutrino decoupling, giving allowed parameters of

$$\frac{m_\phi}{eV} > 10^7 \quad \text{and} \quad \frac{\tau_{\phi\gamma}}{\text{sec}} < 10^{-2}.$$  \hfill (31)

This assumes no extra radiation besides ALPs, and is valid roughly until masses become small enough or lifetimes long enough that decays happen after CMB last scattering. These late decays are analyzed in more detail by Cadamuro and Redondo [11], who find approximately

$$\frac{m_\phi}{eV} < 10^1 \quad \text{or} \quad \frac{\tau_{\phi\gamma}}{\text{sec}} > 10^{24}.$$  \hfill (32)

are once again allowed.

### D. Forecasts

Measurements relevant for placing bounds on ALP parameters have been recently improving and will continue to do so in the near future. It is expected that several probes will soon have the sensitivity to further test the MeV-ALP window. In this section we compute forecasts for some of them.

Currently CMB anisotropies alone are not enough to rule out the MeV-ALP window where there is a maximum of $N_{\text{eff}} = 2.44$ and an increase in $Y_p$, but which lie along the CMB degeneracy direction and are thus allowed. Abazajian et al. [57] show that a Stage-IV CMB experiment could measure $N_{\text{eff}}^{\text{CMB}}$ to within 0.02 at 1-$\sigma$. Given such tight constraints, the arguments of Sec. III A 2 may need to be revisited; while it is true that the ALP decays before any modes relevant for the CMB enter the horizon, the difference is only an order of magnitude in scale factor. Assuming any such corrections do not provide loopholes, if a Stage-IV CMB measurement found a value of $N_{\text{eff}}$ consistent with the standard value of 3.046, the MeV-ALP window would be strongly ruled out [58]. For the CMB constraints, however, there will always be the possibility of extra radiation exactly canceling the dilution due to the ALP decay (as in Sec. IV B).

Helium and deuterium measurements will continue to improve as more systems are discovered and systematic errors are better understood. Additionally, D/H measurements can be significantly improved by
better measuring nuclear reaction rates in the laboratory. Fig. 8 shows that in the D/H-\(Y_p\) plane, the \(\Lambda\)CDM+\(\Delta N_{\text{eff}}\)+ALP scenario is not continuous with the standard model. This allows for sufficiently tight \(Y_p\) and D/H constraints around the standard values to rule out the presence of ALPs, independent of assumptions about extra radiation. We find that the minimum requirement for all points in the \(\Lambda\)CDM+\(\Delta N_{\text{eff}}\)+ALP chain to be ruled out at \(>3\sigma\) by the combination of \(Y_p\) and D/H measurements is a factor of two improvement in the D/H error and a factor of three improvement in the \(Y_p\) error bar. We note the former is possible by eliminating the uncertainty due to nuclear reaction rates alone.

On the laboratory side, in Sec. III C we used the constraint from LEP which limited \(g_{\gamma\gamma} \lesssim 4.5 \times 10^{-4}\) GeV\(^{-1}\). If a search for single-photon events were performed using the entire 1000 fb\(^{-1}\) of currently existing KEKB data and the standard model background was found, forecasts from Kleban and Rabdan [44] show that the constraints could improve to \(g_{\gamma\gamma} \lesssim 10^{-6}\) GeV\(^{-1}\). Similar improvement could come from reinterpreting the constraints on dark photons from 500 fb\(^{-1}\) of BABAR data given in [59] in terms of ALPs. Attempting either of these is outside of the scope of this paper, but could make significant improvements in the mass bounds shown in Fig. 7. SUPER-KEKB, an ongoing upgrade to KEKB, plans to improve on the integrated luminosity of KEKB by a factor of ten with two years of integration, and by a factor of fifty with ten years [60]. Taking the constraint on \(g_{\gamma\gamma}\) to scale with the square root of the integrated luminosity, we find that within the first two years, SUPER-KEKB can rule out the last remaining part of the MeV-ALP window through which the DFSZ-EN2 passes (the mass limit forecast is shown in Fig. 7). The full ten year forecast is exactly enough to close the MeV-ALP window entirely, bringing the collider bound up to \(g_{\gamma\gamma} \lesssim 10^{-6}\) GeV\(^{-1}\) where the SN1987a constraint begins. This simple forecast is in broad agreement with a more sophisticated calculation given by Mimasu and Sanz [15].

Finally, PIXIE is a proposed mission which would greatly improve constraints on CMB spectral distortions [21]. Expected bounds on the \(\mu\) parameter are,

\[
|\mu| < 5 \times 10^{-8}. \tag{33}
\]

Bounds due to spectral distortion constitute constant decay-time boundaries (Sec. II), and we find that the PIXIE forecast moves up the exclusion region by only a factor of five in decay time as compared to FIRAS. This is much less than needed to reach the MeV-ALP window. Evidently it is very difficult to constrain decays happening very much into the \(T\) era using spectral distortions.

V. CONCLUSION

We have shown how cosmological, astrophysical, and laboratory bounds can provide complementary constraints in the mass-lifetime parameter space of axions and ALPs. We have updated the work of Cadamuro and Redondo [11] with constraints from the Planck satellite and the latest inferences of primordial D/H and helium abundances, and provided a more detailed calculation of spectral distortions. The most important change is that CMB+D/H constraints now rule out the entire region corresponding to decays happening after neutrino decoupling but before CMB last scattering. This includes closing the MeV-ALP region of parameter space which we have also shown can correspond to a type of DFSZ axion. The presence of additional radiation can relax the exclusion regions and once again allow the MeV-ALP window. Although it is allowed in this case, including such a particle slightly degrades the overall fit to the CMB+BBN data if our most robust data combination is used. Alternatively, this model can provide a natural explanation for a high value of helium such as found by Izotov and Thuan [38]. Forecasts for future primordial abundance measurements and for SUPER-KEKB are promising; both have the ability to test the MeV-ALP window even in the presence of extra radiation. A detection by either would be very exciting. Even the null result, however, would signify a new level of precision in our understanding of the contents of the primordial plasma.
ACKNOWLEDGMENTS

We thank Kevin Cleary, Francis-Yan Cyr-Racine, Rouen Essign, Brent Follin, Nemanja Kaloper, Yury Kolomensky, Markus Luty, Javier Redondo, and McCullen Sandora for helpful conversations and input. We acknowledge support from NSF awards No. 1213801.
bichaeux, M. Rominsky, R. Roser, J. L. Rosner, C. Rott, P. Rubin, N. Saito, M. Sanchez, S. Sarkar, H. Schellman, B. Schmidt, M. Schmitt, D. W. Schmitz, J. Schneps, A. Schopper, P. Schuster, A. J. Schwartz, M. Schwarz, J. Seenan, Y. K. Semertzidis, K. K. Seth, Q. Shafi, P. Shanahan, R. Sharma, S. R. Sharpe, M. Shiozawa, V. Shihetsev, K. Sigurdson, P. Sikivie, J. Singh, D. Sivers, T. Skwarnicki, N. Smith, J. Sobczyk, H. Sobel, M. Soderberg, Y. H. Song, A. Soni, P. Souder, A. Sousa, J. Spitz, M. Stancari, G. C. Stavenga, J. H. Steffen, S. Stepanyan, D. Stoeckinger, S. Stone, J. Strait, M. Strassler, I. A. Sultanai, R. Sundrum, R. Svoboda, B. Szereszinska, A. Szke, T. Takeuchi, P. Tanedo, S. Taneja, J. Tang, D. B. Tanner, R. Tayloe, I. Taylor, J. Thomas, C. Thorn, X. Tian, B. G. Tice, M. Tobar, N. Tolich, N. Toro, I. S. Towner, Y. Tsai, R. Tschirhart, C. D. Tunnell, M. Tzanov, A. Upadhye, J. Urheim, S. Vainshtein, E. Valencia, R. G. Van de Water, R. S. Van de Water, M. Velasco, J. Vogel, P. Vogel, W. Vogelsang, Y. W. Wai, D. Walker, N. Weiner, A. Weitman, R. Wendell, W. Wester, M. Wetstein, C. White, L. Whitehead, J. Whitmore, E. Widmann, G. Wiedemann, J. Wilkerson, G. Wilkinson, P. Wilson, R. J. Wilson, W. Winter, M. B. Wise, J. Wodin, S. Wojcicki, B. Wojtsekhowski, T. Wongirad, E. Worcester, J. Wurtelle, T. Xin, J. Xu, T. Yamanaka, Y. Yamazaki, I. Yavin, J. Yeck, M. Yeh, M. Yokoyama, J. Yoo, A. Young, E. Zimmerman, K. Zioutas, M. Zisman, J. Zupan, and Z. Zwaska, 

\[ \text{Eqn. 6 ignores the dependence of Primakov mass on } g_2, \text{ leading to changes in freeze-out temperature as large as } \sim 25\% \text{ [see 11].} \]

Since the details of Primakov freeze-out are not important to our main conclusions, this is sufficient.

[3] R. Essig, J. A. Jaros, W. Waster, P. Hansson Adrian, S. Andreas, T. Averett, O. Baker, B. Batell, M. Battagliari, J. Beacham, T. Beranek, J. D. Bjorken, F. Bossi, J. R. Boyce, G. D. Cates, A. Celentano, A. S. Chiu, R. Cowan, F. Curciarelli, H. Davoudiasl, P. deNiverville, R. De Vita, A. Denig, R. Dharmapalan, B. Dongwi, B. Dörbig, B. Echenard, D. Espriu, S. Fegan, P. Fisher, G. B. Franklin, A. Gasparian, Y. Gerstein, M. Graham, P. W. Graham, A. Haas, A. Hatzikoutelis, M. Holtrop, I. Irastorza, E. Izaguirre, J. Jaeckel, Y. Kahn, N. Kalantarians, M. Kohl, G. Krujaic, V. Kubarovskiy, H. Lee, A. Lindner, A. Lobanov, W. J. Marciano, D. J. E. Marsh, T. Maruyama, D. McKeen, H. Merkel, K. Moffett, P. Monaghan, G. Mueller, T. K. Nelson, G. R. Neil, M. Orison, Z. Pavlovic, S. K. Phillips, M. J. Pivovarov, R. Poltis, M. Pospelov, S. Rajendran, J. Redondo, A. Ringwald, A. Ritzi, J. Ruz, K. Saenboonruang, P. Schuster, M. Shinn, T. R. Slatyer, J. H. Steffen, S. Stepanyan, D. B. Tanner, J. Thaler, M. E. Tobar, N. Toro, A. Upadhye, R. Van de Water, B. Vlahovic, J. K. Vogel, D. Walker, A. Weitman, B. Wojtsekhowski, S. Zhang, and K. Zioutas, ArXiv e-prints (2013).

[14] K. A. Olive and P. D. Group, Chinese Physics C 38, 090001 (2014).

[15] K. Mimasu and V. Sanz, ArXiv e-prints 1409, 4792 (2014).

[16] M. Bolz, A. Brandenburg, and W. Buchmüller, Nuclear Physics B 606, 518 (2001).

[17] The degrees of freedom $g_2$ are counted identically to $g_*$, but weighted by the square of the charge of each species, in units of the elementary charge. Efn. 6 ignores the dependence of the plasmon mass on $g_2$, leading to changes in freeze-out temperature as large as $\sim 25\%$ [see 11]. Since the details of Primakov freeze-out are not important to our main conclusions, this is sufficient.
G. Weiglein, E. J. Weinberg, L. R. Wiencke, L. Wolfenstein, J. Womersley, C. L. Woody, R. L. Workman, A. Yamamoto, G. P. Zeller, O. V. Zenin, J. Zhang, R.-Y. Zhu, G. Harper, V. S. Lugovsky, and P. Schaffner, Phys. Rev. D 86, 010001 (2012).

53. D. B. Kaplan, Nuclear Physics B 260, 215 (1985).

54. S. L. Cheng, C. Q. Geng, and W.-T. Ni, Phys. Rev. D 52, 3132 (1995).

55. https://github.com/marius311/cosmoslik

56. Their forecast assumes standard BBN, however we believe it unlikely that freeing $Y_p$ could degrade it so much as to allow 2.4.

57. T. B. Collaboration, Physical Review Letters 113 (2014), 10.1103/PhysRevLett.113.201801, arXiv: 1406.2980.

58. T. Abe, I. Adachi, K. Adamczyk, S. Ahn, H. Aihara, T. B. Collaboration, Physical Review Letters 113, 201801, arXiv: 1406.2980.

59. C. M. Ho and R. J. Scherrer, Phys. Rev. D 87, 065016 (2013).

60. K. N. Abazajian, K. Arnold, J. Austermann, B. A. Benson, C. Bischoff, J. Bock, J. R. Bond, J. Borrell, E. Calabrese, J. E. Carlstrom, C. S. Carvalho, C. L. Chang, H. C. Chiang, S. Church, A. Cooray, T. M. Crawford, K. S. Dawson, S. Das, M. J. Devlin, M. Dobbs, S. Dodenhoff, O. Dore, J. Dunkley, J. Errard, A. Frasie, J. Gallicchio, N. W. Halverson, S. Hanany, S. R. Hildebrandt, A. Hincks, R. Hlozek, G. Holder, W. L. Holzapfel, K. Honscheid, W. Hu, J. Hubmayr, K. Irwin, W. C. Jones, M. Kamionkowski, B. Keating, R. Keisler, L. Knox, E. Komatsu, J. Kovac, C.-L. Kuo, C. Lawrence, A. T. Lee, E. Leitch, E. Linder, P. Lubin, J. McMahon, A. Miller, L. Newburgh, M. D. Niemack, H. Nguyen, H. T. Nguyen, L. Page, C. Pryke, C. L. Reichardt, J. E. Ruhl, N. Sehgal, U. Seljak, J. Sievers, E. Silverstein, A. Slosar, K. M. Smith, D. Spergel, S. T. Staggs, A. Stark, R. Stompor, A. G. Viregg, G. Wang, S. Watson, E. J. Wollack, W. L. K. Wu, K. W. Yoon, and O. Zahn, ArXiv e-prints 1309, 5383 (2013).

61. Their forecast assumes standard BBN, however we believe it unlikely that freeing $Y_p$ could degrade it so much as to allow 2.4-