Cosmological implications of ultralight axionlike fields

Vivian Poulin,[1] Tristan L. Smith,[2] Daniel Grin,[3] Tanvi Karwal,[1] and Marc Kamionkowski[1]
[1]*Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland 21218, USA*
[2]*Department of Physics and Astronomy, Swarthmore College, 500 College Avenue, Swarthmore, Pennsylvania 19081, USA*
[3]*Department of Physics and Astronomy, Haverford College, 370 Lancaster Avenue, Haverford, Pennsylvania 19041, USA*

(Received 9 July 2018; published 16 October 2018)

Cosmological observations are used to test for imprints of an ultralight axionlike field (ULA), with a range of potentials $V(\phi) \propto [1-\cos(\phi/f)]^n$ set by the axion-field value ϕ and decay constant f. Scalar field dynamics dictate that the field is initially frozen and then begins to oscillate around its minimum when the Hubble parameter drops below some critical value. For $n=1$, once dynamical, the axion energy density dilutes as matter; for $n=2$ it dilutes as radiation and for $n=3$ it dilutes faster than radiation. Both the homogeneous evolution of the ULA and the dynamics of its linear perturbations are included, using an effective fluid approximation generalized from the usual $n=1$ case. ULA models are parametrized by the redshift z_c when the field becomes dynamical, the fractional energy density $f_{z_c} \equiv \Omega_a(z_c)/\Omega_{\rm tot}(z_c)$ in the axion field at z_c, and the effective sound speed c_s^2. Using *Planck*, BAO and JLA data, constraints on f_{z_c} are obtained. ULAs are degenerate with dark energy for all three potentials if $1+z_c \lesssim 10$. When $3\times 10^4 \gtrsim 1+z_c \gtrsim 10$, f_{z_c} is constrained to be $\lesssim 0.004$ for $n=1$ and $f_{z_c} \lesssim 0.02$ for the other two potentials. The constraints then relax with increasing z_c. These results have implications for ULAs as a resolution to cosmological tensions, such as discrepant measurements of the Hubble constant, or the EDGES measurement of the global 21 cm signal.

DOI: 10.1103/PhysRevD.98.083525

I. INTRODUCTION

The nature of the dark matter (DM) and dark energy (DE) that dominate our universe today is one of the biggest mysteries of modern cosmology. The dominant paradigm is the ΛCDM model, in which DM is a cold, gravitationally interacting particle, while DE is a pure cosmological constant. Remarkably, this simple model is consistent with precise measurements of the cosmic microwave background (CMB) anisotropies by *Planck* [1], but remains purely parametric.

Ultralight axionlike (ULA) fields arise generically in string theory [2,3]. They may be cosmologically relevant, contributing to the cold dark matter (CDM) *and* DE in our universe (see [3] and references therein). These models have also been invoked to solve tensions within the ΛCDM model, calling on the presence of an early dark energy (EDE) phase [4,5].

For example, increasingly precise measurements of the local expansion rate have led to a potentially significant disagreement (see, e.g., Ref. [6]) between measurements of the Hubble constant inferred from the CMB [1] at high redshifts and Cepheid variables/supernovae at low redshifts [7]. Additionally, if the recently claimed measurement of 21-cm absorption at $z\sim 20$ by the EDGES experiment [8] withstands experimental scrutiny [9], the presence of such early cosmological structure [10] sets a lower bound on the ULA mass of $\sim 10^{-21}$ eV [11], if ULAs compose all of the dark matter.

The apparent anomalously low baryon temperature measured by EDGES could indicate that the expansion history at high redshifts could differ from standard assumptions. These observations could be explained through the cosmological effects of a collection of scalar fields, as envisioned in the "string-axiverse" scenario [2,12–14]. These fields would also affect a variety of cosmological observables, such as CMB and matter power-spectra [15,16] and characterizing their impact is therefore crucial in order to disentangle the axiverse over some other DM or DE scenario.

In this paper we explore the observational implications of a cosmological scalar field with a potential of the form $V_n(\phi) \propto [1-\cos(\phi/f)]^n$ that becomes dynamical at a range of times, which arises nonperturbatively and breaks the approximate ULA shift symmetry. The standard axion potential is obtained in the $n=1$ case, while higher-n potentials may be generated by higher-order instanton corrections [17].

Here ϕ denotes the field value and f the ULA decay constant. These fields become dynamical as the Hubble

parameter decreases, eventually settling down at the minima of their potentials. Up to the point when the fields become dynamical (i.e., during the period of "slow-roll" evolution) their equations of state are dark-energy like: $w_a \simeq -1$.

Soon after the field becomes dynamical it starts to oscillate and, when averaged over the oscillation period, has an equation of state equal to $w_a \simeq (n-1)/(n+1)$ for a potential of the form $V_n(\phi) \propto \phi^{2n}$ [18]. As the field oscillates, its energy density dilutes as cold dark matter (CDM) for $n = 1$, for $n = 2$ it dilutes as radiation and for $n = 3$ it dilutes faster than radiation. With a statistical ensemble of such fields (i.e., the "axiverse") the universe may have gone through several periods of "anomalous" expansion, alleviating the coincidence problem today [4,19–22], and possibly reducing the Hubble constant tension [4] and explaining the anomalously low baryon temperature inferred by the EDGES experiment [5]. This general scenario may also provide a way to connect the physics of cosmic inflation to our current period of accelerated expansion [21].

Here, we extend previous work in several significant ways. First, we present a fluid approximation that parametrizes the ULA dynamics for arbitrary n in terms of the redshift when the field becomes dynamical, z_c, and the fractional energy density in the axion field at z_c, $f_{z_c} \equiv \Omega_a(z_c)/\Omega_{\rm tot}(z_c)$. A key result of this work is the inclusion of ULA perturbations using an effective fluid approach for $n = 2$ and $n = 3$. These perturbations can be approximately described by a time-averaged fluid component with a time and scale dependent effective sound speed [3,15,23–27] within the "generalized dark matter" parametrization [28].

Past applications of this effective fluid approach were restricted to a scalar field of mass m in a quadratic potential. The effect of anharmonicities on the background has been explored (e.g., Ref. [29]), and in Ref. [30,31], a preliminary effective fluid treatment of anharmonic scalar fields was considered. Anharmonic effects would also alter the imprint of ULA DM on the Lyman-α forest power spectrum, with potentially significant implications for the implied lower limit to the ULA mass, if it composes all or most of hte cosmological DM [32]. Similar results are obtained by taking the Schrödinger limit of the Klein-Gordon equation for small length scales, as shown in Ref. [33]. Here, we generalize past work systematically to anharmonic potentials ($n = 2$, and 3), deriving a new straightforward expression for the sound speed $c_{\rm eff}$ which is easy to compute once the behavior of the homogeneous field is known. Moreover, we derive a mapping between this parametrization and the ULA mass, decay constant and initial field value. We show that our fluid formalism is adequate for $n \leq 3$, but breaks down for larger values of n for which the period of oscillation is never much shorter than a Hubble time.

Using *Planck*, measurements of the baryon acoustic oscillations (BAO) and the Joint Light-Curve Analysis (JLA) data [34], we place constraints on ULAs in the $n = 1$, 2 and 3 models. Using a Markov Chain Monte Carlo (MCMC) analysis, we are able to fully explore degeneracies between the ULA parameters and the standard cosmological parameters. We derive constraints on f_{z_c} as a function of z_c. We find in particular that f_{z_c} becomes partially degenerate with dark energy for all three potentials once $1 + z_c > 10$. When $3 \times 10^4 \lesssim 1 + z_c \lesssim 10$, we find that f_{z_c} is constrained to be $\lesssim 0.004$ for matter-dilution and $f_{z_c} \lesssim 0.02$ for the other two potentials. The constraints then relax with increasing z_c, but we demonstrate that current measurements of the CMB[1] require that f_{z_c} be less than unity as early as $z_c = 10^{10}$. Remarkably, we find that the details of the ULA dynamics could distinguish its effects from other cosmological components, even if the ULA time-averaged equation of state is equal to zero (CDM-like) or $1/3$ (radiationlike).

The organization of this paper is as follows. In Sec. II, we review the basics of the cosmological dynamics of ULAs by laying out the equations for the homogeneous field dynamics and introducing the dynamics of the perturbed field. We also present our fluid approximation and how it maps to the ULA theory parameters. Equipped with this formalism, we describe in Sec. III A the rich dynamics of ULA perturbations. Then, in Sec. IV we calculate the CMB and matter power-spectra that arise in our scenario using a modified version of the CLASS Boltzmann code[2] [35–38]. In Sec. V, we use the MONTEPYTHON[3] [39] MCMC package to obtain constraints on our scenario. We discuss implications for cosmological tensions in Sec. VI. We conclude in Sec. VII. In Appendix A, we obtain the generalized effective fluid equations for anharmonic potentials and the effective sound speed for arbitrary n, a result which may be of interest beyond the specific ULA scenario considered here. We compare our fluid formalism to exact solutions of the Klein-Gordon (KG) equations in Appendix B.

II. THE COSMOLOGICAL DYNAMICS OF ULAS

A. Background dynamics

The background dynamics of a ULA have a simple description. The field is initially pinned at some value due to Hubble friction. Once the expansion rate drops below some critical value (related to the mass of the ULA), the field is free to evolve to the minimum of the potential. It then oscillates around the bottom of its potential such that its energy density is diluted due to the subsequent expansion.

The homogeneous Klein-Gordon (KG) equation of motion for the field is given by

$$\ddot{\phi} + 3H\dot{\phi} + \frac{dV_n(\phi)}{d\phi} = 0. \tag{1}$$

[1]Naturally, alternative probes such as BBN can constrain the parameter space further at such times.

[2]http://class-code.net.

[3]http://baudren.github.io/montepython.html.

The ULA potential is given by

$$V_n(\phi) = \Lambda^4(1-\cos\phi/f)^n, \tag{2}$$

where f is the energy scale at which the global $U(1)$ related to axions is spontaneously broken. The ULA homogeneous energy-density and pressure are

$$\rho_a = \frac{1}{2}\dot{\phi}^2 + V_n(\phi), \tag{3}$$

$$P_a = \frac{1}{2}\dot{\phi}^2 - V_n(\phi). \tag{4}$$

The Hubble equation can be written

$$H = H_0 E(a) = H_0\sqrt{\Omega_m(a)+\Omega_r(a)+\Omega_\Lambda+\Omega_a(a)}, \tag{5}$$

where $\Omega_X \equiv \rho_X/\rho_{\rm crit}$ and $\rho_{\rm crit} = 3H_0^2M_P^2$, where $M_P \equiv (8\pi G)^{-1/2}$ is the reduced Planck mass. In order to solve these equations numerically it is useful to redefine the variables so that they are dimensionless. If we define $\Theta \equiv \phi/f$, $m \equiv \Lambda^2/f$, $\alpha \equiv f/M_P$, $x \equiv H_0 t$, and $\mu \equiv m/H_0$ these equations can be written

$$V_n(\Theta) = \mu^2\alpha^2(1-\cos\Theta)^n, \tag{6}$$

$$\Theta'' = -3E\Theta' - \alpha^{-2}\frac{dV_n}{d\Theta}, \tag{7}$$

$$\Omega_a(a) = \frac{1}{3}\left[\frac{1}{2}\alpha^2\Theta'^2 + V_n(\Theta)\right], \tag{8}$$

where a prime indicates a derivative with respect to x.

Before the field starts to oscillate it undergoes "slow-roll" evolution (that is, $\dot{\phi}^2/2 \ll V$ and the dynamics are dominated by Hubble friction) which we will refer to as an "early dark energy" (EDE) phase. To obtain a useful parametrization for all the models under consideration, we have found an analytic approximation to the initial field evolution. First, we expand the potential to linear order around the initial field value Θ_i to obtain a solution for the field evolution (assuming that $\Theta_i' \to 0$ as $x \to 0$):

$$\begin{aligned}\Theta(x) &\simeq \Theta_i + \frac{\sin(\Theta_i)({}_0F_1[\frac{1}{2}(3p+1);\mathcal{A}x^2]-1)}{n\cos\Theta_i + n - 1},\\ &\simeq \Theta_i - \frac{\mu^2 n x^2\sin\Theta_i(1-\cos\Theta_i)^{n-1}}{2(3p+1)} + \mathcal{O}(\mathcal{A}^2x^4)\end{aligned} \tag{9}$$

where ${}_0F_1$ is the confluent hypergeometric function and

$$\mathcal{A} \equiv \frac{1}{4}\mu^2 n(1-\cos\Theta_i)^{n-1}(1-n\cos\Theta_i - n), \tag{10}$$

and where Θ_i is the initial value of the field at $x = 0$ and $a'/a = p/x$ so that during radiation domination $p = 1/2$ and during matter domination $p = 2/3$. When numerically solving for the evolution of the homogeneous scalar field, we take the initial field value to be $0 < \Theta_i < \pi$ and the initial velocity of the field is determined by the curvature of the potential at Θ_i through Eq. (9). We set $p = 1/2$ since the field is always initialized during radiation domination.

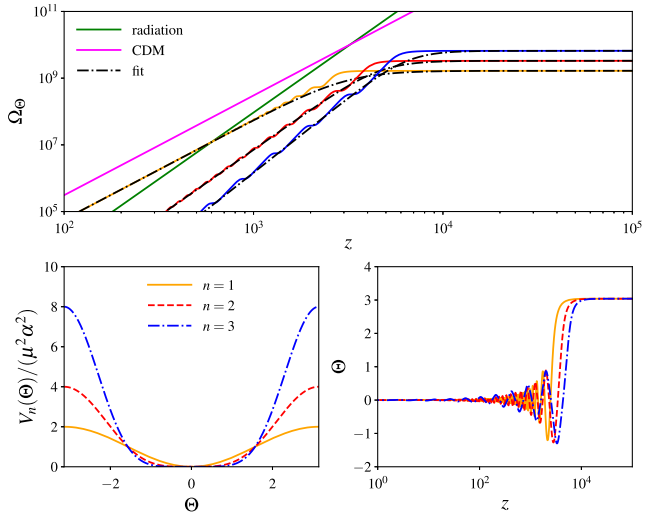

FIG. 1. The evolution of the background field with $\mu = 10^6$, $\alpha = 0.05$, and $\Theta_i = \pi - 0.1$ for the three forms of the axion potential explored in this paper.

After a period of slow-roll evolution, the field transitions to an oscillatory phase with a decreasing amplitude due to the dilution of the field's energy density from expansion. The potential during the oscillating phase takes the form $V_n(\Theta) \simeq 2^{-n}\mu^2\alpha^2\Theta^{2n}$ so that for $n = 1$ the field undergoes simple harmonic oscillation with a frequency which is independent of its amplitude and for $n > 1$ the oscillations are anharmonic and the frequency depends on the amplitude. We show the evolution of Θ for the three forms of the potential considered here in Fig. 1.

Once oscillating, over timescales shorter than a Hubble time the field evolves according to the equation of motion

$$\Theta'' + \alpha^{-2}\frac{dV_n}{d\Theta} = 0. \tag{11}$$

Furthermore if we assume that the oscillation frequency $\varpi \gg H$, the total energy will be approximately conserved over several oscillations so that we have

$$\frac{1}{2}\Theta'^2 + \alpha^{-2}V_n(\Theta) = \alpha^{-2}V_n(\Theta_m), \tag{12}$$

where Θ_m is the maximum field value, reached when $\Theta' = 0$. We can use the virial theorem to write $\langle 1/2\Theta'^2\rangle = n\alpha^{-2}\langle V_n\rangle$ so that

$$\langle\Omega_\Theta\rangle \simeq \frac{1}{3}\frac{\mu^2\alpha^2}{2^n}\Theta_m^{2n} \simeq \Omega_{a,0}a^{-3(1+w_n)}, \tag{13}$$

which shows that, due to the expansion of the universe, the maximum field value will decrease as $\Theta_m \propto a^{-3/(1+n)}$. As in Ref. [18], we find that the axion energy density is constant at early times and decays as $a^{-3(1+w_n)}$ with

$$w_n \equiv \frac{n-1}{n+1}. \tag{14}$$

With this we will parametrize the axion energy density by

$$\Omega_a(z) = \frac{2\Omega_a(z_c)}{[(1+z_c)/(1+z)]^{3(w_n+1)}+1}, \tag{15}$$

which has an associated equation of state

$$w_a(z) = \frac{1+w_n}{1+[(1+z)/(1+z_c)]^{3(1+w_n)}} - 1, \tag{16}$$

and which asymptotically approaches -1 as $a \to 0$ and w_n for $z \ll z_c$. We show a comparison between the exact axion energy density and our parametrization in Fig. 1. This shows that when $n = 1$, the homogeneous axion energy density dilutes like matter once the field is dynamical. On the other hand it dilutes like radiation when $n = 2$. When $n \geq 3$, dilution is faster than radiation.[4]

B. On the validity of the fluid approximation

From Eq. (12) we can compute the time for one full oscillation:

$$T \simeq 4H_0^{-1}\int_0^{\Theta_m} \frac{d\Theta}{\sqrt{2\alpha^{-2}[V(\Theta_0)-V(\Theta)]}}, \tag{17}$$

$$= 4H_0 \frac{\sqrt{\pi}2^{\frac{n-1}{2}}\Theta_m^{1-n}\Gamma(1+\frac{1}{2n})}{\mu\Gamma(\frac{n+1}{2n})} \tag{18}$$

This leads to an angular frequency

$$\varpi = \varpi_0 a^{-3w_n}, \tag{19}$$

$$\varpi_0 = H_0 \frac{\sqrt{\pi}2^{-\frac{n^2+1}{2n}}\Omega_{\Theta,0}^{\frac{n-1}{2n}}\Gamma(\frac{n+1}{2n})(\alpha\mu)^{1/n}}{\alpha\Gamma(1+\frac{1}{2n})}. \tag{20}$$

This shows that the angular frequency is only constant if $n = 1$; for $n > 1$ the oscillation frequency *decreases* in time [42]. In particular, the fluid approximation is only accurate if $\varpi/H \gg 1$ and, assuming that the axion field never dominates the energy budget, we have

$$\frac{\varpi}{H} \propto \begin{cases} a^{(5-n)/(1+n)} & a < a_{\rm eq}, \\ a^{6/(1+n)-3/2} & a > a_{\rm eq}, \end{cases} \tag{21}$$

where $a_{\rm eq} \equiv \Omega_{r,0}/\Omega_{m,0}$ is the value of the scale-factor at matter/radiation equality. This ratio increases with time for $n < 5$ during radiation domination and for $n < 3$ for matter domination. During a period of accelerated expansion the ratio will *decrease* in time for any positive value of n. Therefore, in this work we limit our study to $n \leq 3$ such that if $\varpi/H \gtrsim 1$ at the start of the oscillatory phase, then the ratio will remain large up until almost today, when the latest epoch of cosmic acceleration began.

C. Perturbed dynamics in the fluid formalism: A first look

Linear perturbations to the axion field will develop and evolve according to the perturbed Klein-Gordon equation. However, these equations are computationally expensive to solve and would not allow us to scan over the parameters of the ULA theory and the standard cosmological parameters. Since the oscillations of the scalar field generally occur with periods much shorter than a Hubble time, much of the dynamics can be captured by averaging over the oscillations and dealing with fluid equations [18]. The equations governing the evolution of density and bulk velocity perturbations can be written in terms of fluid variables in the synchronous gauge as [28]

$$\dot{\delta}_a = -(1+w_a)\left(\theta_a + \frac{\dot{h}}{2}\right) - 3(c_s^2 - w_a)\mathcal{H}\delta_a - 9(1+w_a)(c_s^2 - c_a^2)\mathcal{H}\frac{\theta_a}{k^2}, \tag{22}$$

$$\dot{\theta}_a = -(1-3c_s^2)\mathcal{H}\theta_a + \frac{c_s^2 k^2}{1+w_a}\delta_a, \tag{23}$$

where in these equations the dot refers to a derivative with respect to conformal time. From the background dynamics, w_a is known. Note that the effective sound speed $c_s^2 \equiv \delta p/\delta\rho$, is possibly different from unity for an ULA, and the adiabatic sound speed

$$c_{\rm a}^2 \equiv \frac{\dot{P}_a}{\dot{\rho}_a} = w_a - \frac{\dot{w}_a}{3(1+w_a)\mathcal{H}}. \tag{24}$$

The adiabatic sound speed is straight forward to calculate since it depends only on background quantities. Using the initial EDE evolution of the field given in Eq. (9) and assuming $\dot{\phi}_i = 0$,[5] one can show that $c_{\rm a}^2 \simeq -7/3$ [15] during slow-roll for any form of the potential.

[4] A qualitatively similar stiff dilution phase occurs in *complex* scalar-field dark-matter models, with the relevant phenomenology discussed in Refs. [40,41].

[5] In our model, this is naturally realized because of the large Hubble friction at early times.

In the approximation for w_a given by Eq. (16) the adiabatic sound-speed during the EDE period is given by

$$c_{\rm a}^2 = -\frac{3n+1}{n+1}, \tag{25}$$

it then evolves to w_a once the field starts oscillating. At early times, except for the case $n = 2$, this parametrized adiabatic sound-speed differs from the exact value of $-7/3$ (with a range $-7/3 \leq c_{\rm a}^2 \leq -5/2$). We have checked that given that both the exact and parametrized $c_{\rm a}^2$ are negative and of order unity, our parametrization gives a good approximation to the exact evolution of the perturbations. We show a comparison between the exact mode evolution and the approximate mode evolution in Appendix B.

Finally, in order to utilize the GDM equations of motion, we must determine c_s^2. During the EDE phase $c_s^2 = 1$ for a slowly rolling scalar field, but deviates strongly from 1 once the field starts oscillating. We discuss our derivation of the time-averaged effective sound speed in Appendix A. We find that for a ULA potential which takes the form $V \propto \phi^{2n}$ around the minimum:

$$c_s^2 = \frac{2a^2(n-1)\varpi^2 + k^2}{2a^2(n+1)\varpi^2 + k^2}, \tag{26}$$

with the frequency ϖ given by Eq. (19). We discuss the dynamics of perturbations in Sec. III A. Before entering into these details, we relate our parametrization to the ULA theory parameters.

Note that c_s^2—the sound speed of the ULA in its local rest-frame—is never negative, since during the Hubble friction dominated evolution $c_s^2 = 1$, and then once it is oscillating it is given by our Eq. (26), which clearly shows it is always positive. On the other hand, the adiabatic sound speed, $c_{\rm a}^2$, can be negative. The adiabatic sound speed does not govern the speed at which perturbations in the field propagate, so a negative value does not indicate any issues with stability. Instead the adiabatic sound speed gives the relative rate of change of the pressure to the density—so if one increases while the other decreases the adiabatic sound speed can be negative. This is exactly what happens during the Hubble dominated evolution of the field when the field's pressure increases while its energy density decreases.

D. Approximate translation between model and theory parameters

The axion model is fully specified by four "theory" parameters: the potential-index n, the initial field value $\Theta_i \equiv \phi_i/f$, the mass parameter $\mu \equiv m/H_0$, and the coupling parameter $\alpha \equiv f/M_P$. Our model is also described by four parameters: the redshift z_c when the field begins to oscillate, the energy density of the field $\Omega_\phi(z_c)$ at z_c, the time-averaged equation of state w_n during oscillations, and the scale dependence ϖ_0 of the effective sound-speed. The equation of state w_n and the index n are related through Eq. (14) and ϖ_0 is related to α and μ through Eq. (20). The last two parameters are related by more involved expressions, as we now discuss.

First, note that we can use Eqs. (9) and (15) to relate $\Omega_a(z_c)$ to μ, α, and Θ_i by computing the energy density in the axion field at very early times:

$$\Omega_a(z_c) = \frac{1}{6}\alpha^2\mu^2(1-\cos\Theta_i)^n. \tag{27}$$

We can obtain an approximate expression for z_c by noting that the field starts to oscillate soon after the field evolves away from its initial value, Θ_i. We can compute the time at which the field starts to evolve using the approximate evolution of $\Theta(x)$ given in Eq. (9). We define x_c as the time at which the field evolves to some fraction of its initial value, $\Theta(x_c) = \mathcal{F}\Theta_i$:

$$x_c \equiv \frac{(1-\cos\Theta_i)^{\frac{1-n}{2}}}{\mu}\sqrt{\frac{(1-\mathcal{F})(6p+2)\Theta_i}{n\sin\Theta_i}}. \tag{28}$$

We can relate this to z_c by using the fact that during radiation or matter domination the Hubble parameter is given by $E \simeq p/x$ so that

$$E(z_c) \simeq \frac{p}{x_c}, \tag{29}$$

where, as before, $p = 1/2$ for $z_c > z_{\rm eq} = \Omega_{r,0}/\Omega_{M,0} \simeq 10^{-5}$ and $p = 2/3$ for $z_c < z_{\rm eq}$. We compare our approximate x_c to the full field evolution by solving Eq. (7) and find that z_c is most accurately approximated when we choose $\mathcal{F} = 7/8$. Note that our approximate solution for the field evolution in Eq. (9) fails in the limit $\Theta_i \to \pi$ and we have found that this mapping can reproduce the full dynamics up until $\Theta_i \simeq 3$. Also note that for $n = 1$ and $\Theta_i \ll 1$ our results give $H(z_c) \simeq m$ which agrees with previous work [25].

This mapping can be used to go from our model parameters to the theory parameters. Assuming that the field makes up a small fraction of the total energy density at z_c, we can use Eq. (29) to determine x_c and then Eq. (28) provides a relationship between μ and Θ_i. Given $\Omega_\phi(z_c)$, Eq. (27) provides a relationship between μ, α, and Θ_i. Combining these together we can then write ϖ_0 as a function of z_c, and Θ_i:

$$\varpi_0(z_c,\Theta_i,n) = H_0\mu(z_c,\Theta_i,n)(1-\cos\Theta_i)^{\frac{n-1}{2}}\mathcal{G}(z_c,n), \tag{30}$$

$$\mathcal{G}(z_c,n) \equiv \frac{\sqrt{\pi}\Gamma\left(\frac{n+1}{2n}\right)}{\Gamma\left(1+\frac{1}{2n}\right)}2^{-\frac{n^2+1}{2n}}3^{\frac{1}{2}\left(\frac{1}{n}-1\right)} \times (1+z_c)^{\frac{6}{n+1}-3}\left[(1+z_c)^{\frac{-6n}{n+1}}+1\right]^{\frac{1}{2}\left(\frac{1}{n}-1\right)}. \tag{31}$$

This shows that, in principle, the homogeneous and perturbative effects of this field on cosmological observation can give us enough information to reconstruct all of the theory parameters. Said another way, an estimate of z_c and $\Omega_a(z_c)$ from the homogeneous dynamics of the field will determine the evolution of perturbations up to the unknown initial field value, Θ_i; an estimate of ϖ_0 from the perturbations then determines Θ_i.

Finally, we can use these expressions to relate the theory parameters to the model parameters. In particular Eqs. (28) and (30) show that z_c and ϖ_0 are both determined by μ and Θ_i. These can then be combined to give an estimate of the fractional contribution of the ULA to the total energy density f_{z_c} at z_c. Recall that these expressions have assumed $f_{z_c} \ll 1$ and our analytic expressions for the field evolution are only accurate for $\Theta_i \lesssim \pi - 0.1$.

III. DETAILED STUDY OF ULA PERTURBATIONS IN THE FLUID APPROXIMATION

A. Setup and initial conditions of perturbations

As explained previously, we solve for the ULA dynamics using the GDM equations of motion [28], which require the specification of the ULA equation-of-state w_a, the adiabatic sound speed $c_{\rm a}^2$, and effective sound speed c_s^2. During slow roll, generic scalar fields have that $w_a \simeq -1$, $c_{\rm a}^2 \simeq -7/3$, and $c_s^2 = 1$. Since $w_a \simeq -1$ the linear perturbation equations written in terms of the velocity perturbation θ_a are unstable. To deal with this we solve the evolution of the perturbations in terms of the heat-flux, $u_a \equiv (1+w_a)\theta_a$ [15].

$$\dot{\delta}_a = -\left[u_a + (1+w_a)\frac{\dot{h}}{2}\right] - 3(c_s^2 - w_a)\mathcal{H}\delta_a - 9(c_s^2 - c_{\rm a}^2)\mathcal{H}\frac{u_a}{k^2}, \tag{32}$$

$$\dot{u}_a = -(1-3c_s^2)\mathcal{H}u_a + 3\mathcal{H}(w_a - c_{\rm a}^2)u_a + c_s^2 k^2 \delta_a. \tag{33}$$

In practice, when $z > z_c$, we set $w_a \simeq -1$, $c_s^2 = 1$ and $c_{\rm a}^2$ is given by Eq. (25). During the oscillatory phase, when $z < z_c$, c_s^2 is given by the time and scale-dependent effective sound speed in Eq. (26), $c_{\rm a}^2$ is given by Eq. (24) with w_a given by Eq. (16). Abrupt changes in these quantities can lead to the generation of transients in numerical solutions. We have verified that these had no significant effects on the predicted power spectra used to constrain this model. A comparison between the approximate and exact ULA evolution is discussed in Appendix B and shows very good agreement.

In general, adiabatic initial conditions on super-Hubble scales are expected when the perturbations within each component are due to a single degree of freedom (e.g., slight time delay in the decay of the inflaton field) and lead to simple relations of the type

$$\frac{\delta_i(\tau,\vec{x})}{1+w_i} = \frac{\delta_{i'}(\tau,\vec{x})}{1+w_{i'}} = -\frac{h}{2}, \tag{34}$$

where i and i' are two species and $h \sim (k\tau^2)$ corresponds to the growing mode solution of a fourth order linear differential equation for the trace of the metric perturbation in the synchronous gauge [43]. For a species with zero nonadiabatic sound speed, this would typically be enough. However, a fluid with $c_s^2 \neq c_a^2$ does not generally obey such relations. In the ULA scenario considered here the ULA component is always subdominant on superhorizon scales and at early times. In that case, the ULA perturbations fall inside the gravitational potential wells created by the radiation component, such that there is a generic attractor solution [44]

$$\delta_a = -\frac{C}{2}(1+w_a)\frac{4-3c_s^2}{4-6w_a+3c_s^2}(k\tau)^2, \tag{35}$$

$$u_a = -\frac{C}{2}(1+w_a)\frac{c_s^2}{4-6w_a+3c_s^2}(k\tau)^3 k, \tag{36}$$

where C is the initial amplitude and τ is the conformal time. Note that we take $\delta_a = u_a = 0$ initially since these quantities are quickly driven to the attractor solution [44].

More generally, if the axion symmetry-breaking scale $f > H_I$ (where H_I is the inflationary Hubble parameter), axions will carry isocurvature perturbations, as a light relic present during the inflationary era (see Ref. [45] and references therein). Indeed, this scenario may be natural in light of relic-density bounds on the complementary $f < H_I$ scenario [46], though in practice, isocurvature signatures may be quite small for observationally allowed values of H_I.

Isocurvature perturbations will change the height of the Sachs-Wolfe plateau and alter the phases of CMB acoustic peaks. Limits to isocurvature perturbation from CMB data are now quite stringent, and constrain the ratio f/H_I, with implications for the amplitude of inflationary gravitational waves. The complementarity between isocurvature and tensor modes in axion models is explored more fully (for the harmonic limit of the $n = 1$ potential) in Ref. [45]. In future work, we plan to explore the phenomenology of and constraints to ULA isocurvature perturbations for the much more general class of models considered here.

B. Time evolution of ULA perturbations

For a fixed wave number k there are three timescales that are important for the ULA mode evolution: (i) horizon crossing $k = a_k H(a_k)$; (ii) the redshift z_c at which the field starts to oscillate around its minimum; (iii) the time a_s after which the sound speed is equal to the oscillation-averaged ULA equation of state, $k = a_s \varpi(a_s)$. Note that, $a_c \equiv 1/(1+z_c)$ is always smaller than a_s, or in other words the field starts oscillating before its sound speed starts to evolve. This is because, for the field to acquire $c_s^2 < 1$, it must be oscillating. However, for a given k, one can

TABLE I. Theory parameters (determined using the translation equations in Sec. II D) for $z_c = 10^4$, $f_{z_c} = 0.01$, and $\Theta_i = \pi/2$.

Parameter	$n = 1$	$n = 2$	$n = 3$
μ	1.54×10^6	1.09×10^6	8.92×10^5
α	0.124	0.175	0.215
ϖ_0 (Mpc^{-1})	341.7	0.0185	1.11×10^{-4}

potentially have an arbitrary hierarchy between a_k and a_c, and a_k and a_s.

We wish to explore the mode evolution of different Fourier modes for the three forms of the ULA potential at fixed z_c, which we set to be 10^{-4}. We choose a fraction of the total energy density at z_c in the ULA to be $f_{z_c} = 0.01$. In doing so, the ULA never makes a significant contribution to the total energy density of the universe. Since the ULA is always sub-dominant, comparing the evolution of the same wave number leads to equal a_k for each ULA potential.

For each value of n, we can use Eq. (28) to translate our condition on z_c to a relationship between α and Θ_i. Similarly by specifying f_{z_c} we fix the relationship between μ and Θ_i. We are left with one degree of freedom to fully specify the model: the value of the frequency ϖ_0, which enters the effective sound speed after the field starts oscillating and is specified by further fixing Θ_i. In the following discussion we arbitrarily set $\Theta_i = \pi/2$ (choosing another value would not affect our conclusions). The resulting theory and parameter values for the specific model discussed in this section are shown in Table I.

We explore the evolution of three modes: $k_1 = 1$ Mpc^{-1}, $k_2 = 10^{-2}$ Mpc^{-1}, and $k_3 = 10^{-3}$ Mpc^{-1}. We show these modes, along with other important scales, in Fig. 2. From Eq. (26) we can see that if $k \gg a\varpi$ the sound speed goes to 1. Hence, for a fixed k, the time-evolution of the effective sound speed is different for different values of n: c_s^2 goes from one to less than one for $n = 1$, it is a constant for $n = 2$, and it evolves from less than one to one for $n = 3$.

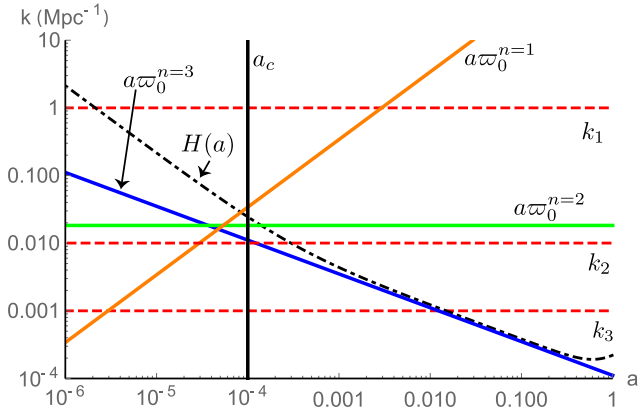

FIG. 2. The evolution of a series of scales associated with ULA perturbations. Note that while $k > a\varpi$ the mode has $c_s^2 \simeq 1$.

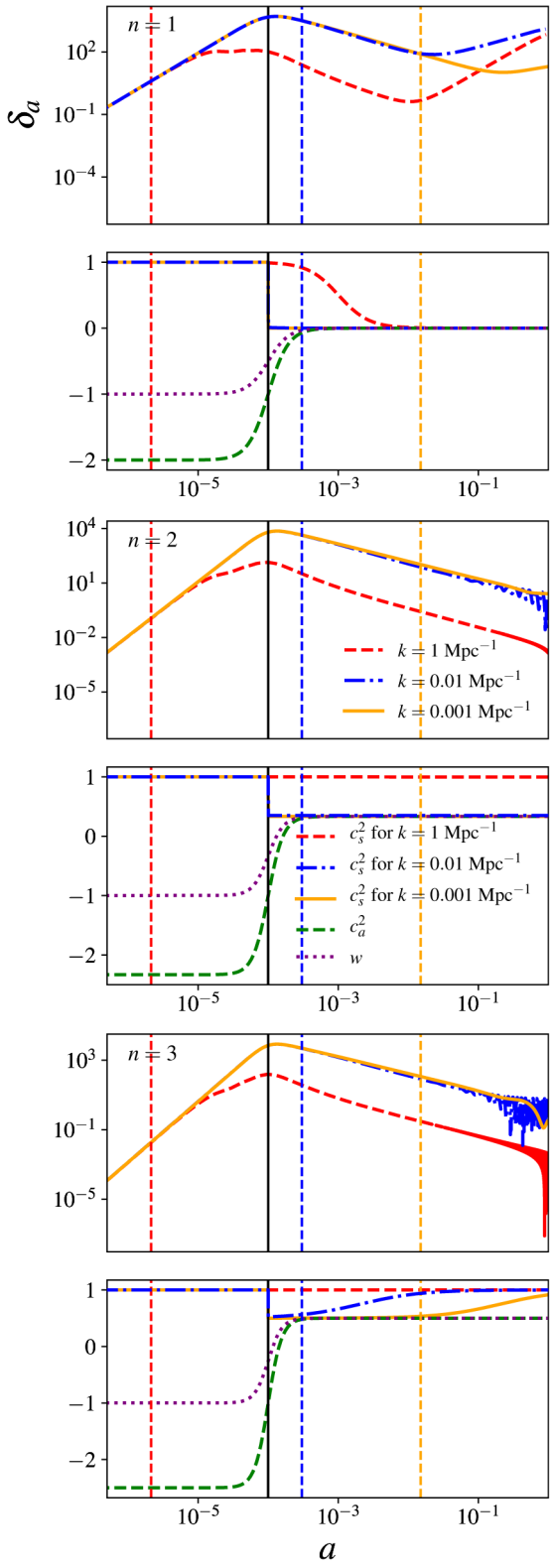

FIG. 3. Evolution of the density contrast for the three forms of the ULA potential considered in this paper and with $z_c = 10^4$, $f_{z_c} = 0.01$, and $\Theta_i = \pi/2$. The vertical dashed black line shows a_c, while the vertical dashed colored lines show Horizon crossing for each mode.

These modes were chosen because they have different hierarchies: k_1 has $a_{k_1} < a_c < a_{s_1}$, k_2 has $a_c < a_{k_2}$ and no a_{s_2}, and k_3 has $a_c < a_{s_3} < a_{k_3}$. In Fig. 3 we show the evolution of the ULA density contrast for these three modes.

At early times, as long as the mode is superhorizon and $a < a_c$, we have $w_a \simeq -1 + c_n(a/a_c)^{3(1+w_n)}$, where c_n is a factor of order unity. The evolution of density perturbations is similar for each ULA potential and each mode, as dictated by the initial behavior in Eqs. (35) and (36). Since both the density contrast and the heat flux are proportional to $1 + w_a$, this shows that for a fixed a_c we expect that the lower values of n will have larger perturbations. This is indeed the case in Fig. 3.

As illustrated by the wave number k_1, modes with $a_k < a_c$ enter the horizon while the field is still undergoing slow-roll, EDE, evolution. This results in a suppression in the growth of the perturbations compared to their superhorizon evolution. Once $a \gtrsim a_c$, the field starts to oscillate in its potential and $w_a \to (n-1)/(n+1)$. As long as $a < a_s$, $c_s^2 = 1$ and the pressure support leads to a strong decrease in the perturbation amplitude. This suppression is present for both superhorizon and subhorizon modes.

Once $a > a_s$, $c_s^2 \to w_a$ and the field's internal pressure support will decrease. In the case where $n = 1$ the field is effectively pressure-free and the density perturbation starts tracking that of CDM. For $n > 1$ some residual pressure support remains, leading to rapid oscillations in the ULA's density perturbations with an oscillation frequency and amplitude that differs for each n and k.

IV. IMPACT OF AN ULTRALIGHT AXION ON THE CMB AND MATTER POWER SPECTRA

We compute the CMB and matter power spectra using CLASS for several values of the potential exponent $n = (1, 2, 3)$ and decay redshift $1 + z_c = (10, 10^5)$. We set the six ΛCDM parameters to their best fit values of *Planck* TT, TE, EE + lowP 2015 [1]. We fix the angular scale of the sound horizon, θ_s, which requires us to adjust the value of H_0 (this is done using a shooting method implemented in CLASS). We set the density of ULAs to its upper limit at 95% C.L. derived in the next Section. The results are shown in Figs. 4 and 5.

A. The CMB power spectra in the presence of a ULA

We start by discussing the case where the dilution starts much before matter-radiation equality, shown in Fig. 4. In that case, the CMB power spectra show clear differences depending on the value of n, i.e., on the properties of the ULA once it starts diluting.

For $n = 1$, the effects of the ULA are very similar to that of an extra matter component. We illustrate this by comparing it to a universe with an additional pure CDM component, represented by the black curve on each panel of Fig. 4. At the highest multipoles, the amplitude of the acoustic peaks is altered by an earlier epoch of matter/radiation equality, changes to the gravitational driving of acoustic oscillations (affecting the Sachs-Wolfe term), and modifications to the blue shift of photons in decaying gravitational wells, the early integrated Sachs-Wolfe (EISW) effect [47,48]. Since we hold the angular sound horizon θ_s fixed the angular scale θ_d of diffusion damping will vary and this leads to an altered damping tail. Keeping θ_s fixed for a flat universe required adjusting the value of Ω_Λ, changes the late integrated Sachs-Wolfe (LISW) plateau visible at low-l's. It is also visible in the EE spectrum as the reionization history is affected by a change in Ω_Λ. Note that the effect of this ULA on the CMB power spectra makes it a viable CDM candidate (see, e.g., Ref. [25]): had we adjusted $\omega_{\rm cdm}$ accordingly, the remaining effects would have been due to a suppression of the matter power spectrum on small scales (which we comment on later) and therefore almost invisible in the CMB, aside from a moderately altered lensing power spectrum.

For $n = 2$, the effects of the ULA are very similar to that of an extra radiation component [49]. CMB anisotropies are then altered for two reasons. First, at the background level,

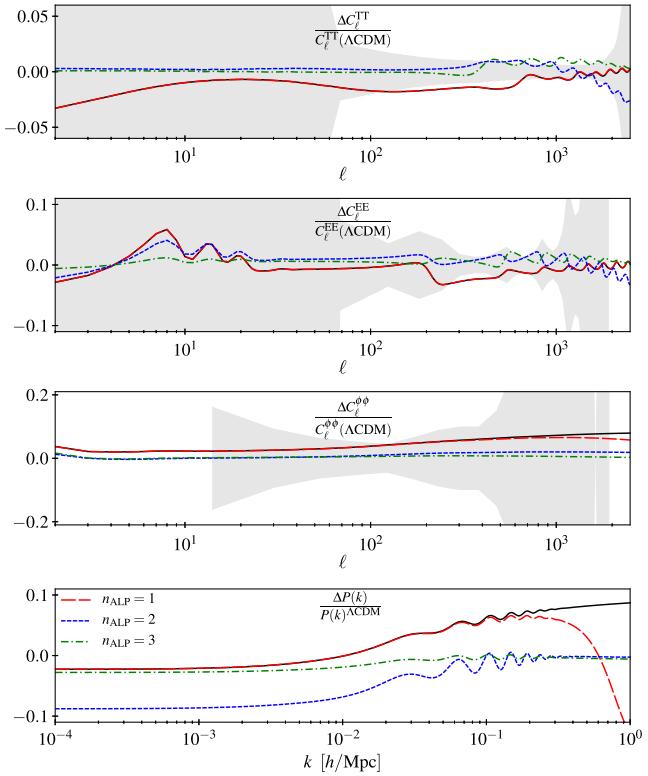

FIG. 4. Residuals of the (lensed) CMB TT power spectrum (first panel), EE power spectrum (second panel), lensing power spectrum (third) and matter power spectrum (fourth panel) computed for several values of the potential exponent $n = (1, 2, 3)$ and $1 + z_c = 10^5$. Residuals are taken with respect to the ΛCDM model, with parameters given by the best fit of *Planck* TT, TE, EE + lowP [1]. Axion densities are set at their constraints at 95% C.L. The grey bands show *Planck* 1σ sensitivity.

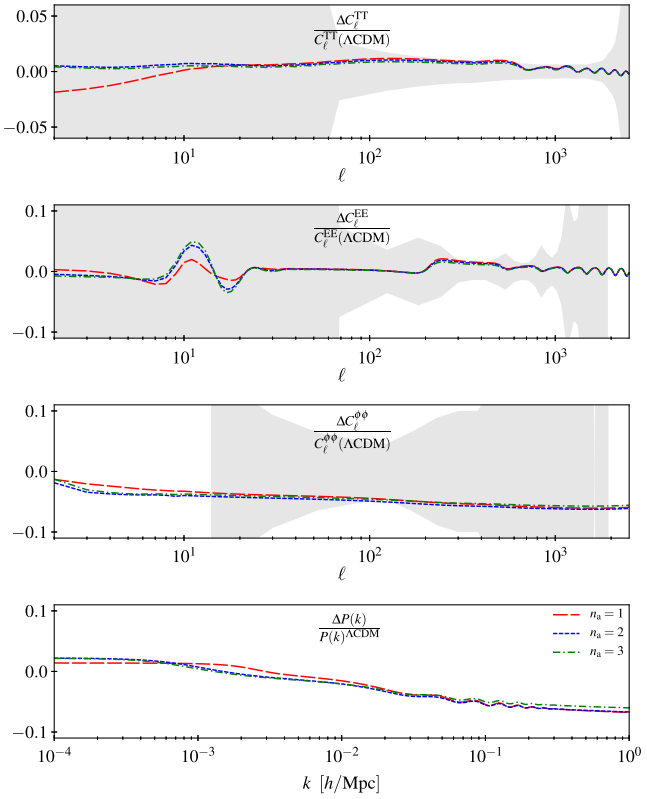

FIG. 5. Same legend as Fig. 5 for $1 + z_c = 10$.

the additional relativistic species shift matter-radiation equality, which produces modified gravitationally driven oscillations in the photon-baryon plasma and EISW. Hence, the main background effect is due to the requirement that θ_s is kept fixed and manifests as a shift in the damping scale θ_d, a very mild LISW effect and some oscillation patterns due to different reionization history in the EE spectrum. Second, at the level of perturbations, such an ULA produces a BAO phase-shift distinct from that of true free-streaming particles like neutrinos [50]. *Planck* data are not only sensitive to the background effect of neutrinos, but also to the "neutrino-drag" [51–55], and have already been used to constrain the effective sound speed c_s^2 and viscosity $c_{\rm vis}^2$ of the non-CMB radiation component and found to be consistent with that of free-streaming neutrinos [28]. These parameters are distinct in ULA models, and so we do not expect strong degeneracies between ULAs and neutrinos.

For $n = 3$, the energy density of the axion dilutes faster than any known cosmological species. This leaves less of an imprint on the CMB than the $n = 1$ or $n = 2$ cases, and most of the effects can be attributed to the EDE phase, rather than to the diluting fluid which becomes quickly invisible. Since θ_s is kept fixed, the most important effect of the extra amount of expansion is to reduce the amplitude of the damping tail. On the other hand, the nonadiabatic sound speed of the diluting ULA also leads to small peculiar phase-shift of the acoustic peaks, in a manner different from that of $n = 2$ or a free-streaming species.

If the dilution begins after recombination, the exponent n has much less impact. The EDE phase has a slight impact on the growth of metric potentials around recombination, which leads to features at high multipoles (and especially around $\ell \sim 300$). The additional residual wiggles at high-l's are mostly due to the different amount of lensing. It depends on the impact of the ULA on the matter power spectrum which we comment on below. The difference between the dynamics of the perturbations are mostly visible at small l's. Since we keep θ_s fixed, Ω_Λ is changed which in turn affects z_Λ. However, in the $n = 1$ case the additional matter component shifts z_Λ further, in turn affecting more strongly the LISW plateau, in a manner similar to massive neutrinos. Further differences can be attributed to the impact of the different $w(a)$ as the fluids dilute differently, but fall well below cosmic variance. However, we expect that experiments sensitive to late-time expansion (e.g., JLA, BAO) are sensitive to these effects.

B. The matter power spectrum in the presence of a ULA

We now turn to the matter power spectrum, which also shows interesting features strongly dependent on the EDE dilution time and potential power-law index n. In general, once A_s and n_s are fixed, the matter power spectrum depends on: (i) the sound horizon at baryon drag $r_s(z_{\rm drag})$ which dictates the phase of the BAO; (ii) the Hubble scale at matter radiation equality $k_{\rm eq} \equiv a_{\rm eq} H_{\rm eq}$ which sets the position of the peak; (iii) the ratio $\omega_b/\omega_{\rm cdm}$, which affects the power on scales $k > k_{\rm eq}$ and the contrast of the BAO; (iv) the ratio $[g(a_0, \Omega_m)/\Omega_m]^2$ which dictates $k < k_{\rm eq}$ and where $g(a, \Omega_m) = D(a)/a$ is a function expressing how much the growth rate of structures $D(a)$ is suppressed during Λ domination.

When the dilution starts before matter-radiation equality, the ULA affects $z_{\rm eq}$ especially if it dilutes like matter or radiation. The peak position $k_{\rm eq}$ therefore depends on n. If $n = 1$, the ratio $\omega_b/\omega_{\rm cdm}$ decreases which leads to a large increase for $k > k_{\rm eq}$ until the mode-dependent sound speed of the ULA kicks in. This creates a turnover at $k > k_{\rm eq}$ that is very specific to such a ULA. One can see that the only difference between a pure CDM component and a ULA is this cutoff on very small sales because of the nonzero pressure support. For $n = 2$ and $n = 3$, this branch is almost unaffected for such small values of Ω_a. The small-k branch on the other hand is affected by the increase in $\Omega_M = 1 - \Omega_\Lambda$ (decrease in Ω_Λ) that is required to keep θ_s fixed. Moreover, for all values of n the BAO is shifted because of different $r_s(z_{\rm drag})$.

When the dilution starts after matter-radiation equality, the effects are very similar to that of massive neutrinos, and manifests in two ways. First, the ratio $k_{\rm eq}/(a_0 H_0)$

governing the location of the maximum in the matter power spectrum depends on the duration of matter domination. Any modification of this ratio leads to an overall shift of the spectrum. It is affected by the presence of an EDE, but the additional matter component (for the $n = 1$ case) partially counteracts the effect of the EDE. Hence, the power spectrum in the $n = 2$ and 3 case is shifted in the same way, and slightly more than in the $n = 1$ case. Second, the additional pressure support leads to suppression of power on small scales in a manner that depends on each fluid sound speed, and thus differs for each n.

V. CURRENT CONSTRAINTS TO ULAS

Using current measurements of the CMB and other probes of large-scale structure we place constraints on the energy density of ULAs as a function of the time when they become dynamical. As mentioned before, although the CMB decouples around $z \sim 1000$, each multipole carries with it information about the evolution of the universe around the time the scales that form it entered the causal horizon. This, in principle, makes the CMB sensitive to cosmological dynamics as long ago as $z \sim 10^5$–10^6 [4,20].

To perform this analysis we consider a series of fixed values for z_c at which we constrain the energy density in the ULA. In addition to this we assume a uniform prior on the initial field value, Θ_i, which in turn implies a particular prior on the ULA's oscillation frequency today, ϖ_0 [see Eq. (20)].

A. Description of the data sets and analysis

We run Monte Carlo Markov chains using the public code MONTE PYTHON [39]. We perform the analysis with a Metropolis Hasting algorithm, assuming flat priors on $\{\omega_b, \theta_s, A_s, n_s, \tau_{\rm reio}, \omega_{\rm cdm}\}$ and a logarithmic prior on $\Omega_{\rm a}$. We scan over 9 points in $1 + z_c$ logarithmically distributed between 1 and 10^8. We also vary n to be equal to (1,2,3). We make use of *Planck* high-l and low-l TT, TE, EE and lensing likelihood. We include the anisotropic BAO data at $z = 0.2$–0.75 from the BOSS DR12 data release [56] and isotropic BAO data at $z = 0.105$ [57] and $z = 0.15$ [58]. We include the Joint Likelihood Analysis (JLA) of supernovae, which includes measurements of the luminosity distance of SN1a up to redshift $z \sim 1$ [34].

Although not specified here for brevity, there are many nuisance parameters that we analyze together with the cosmological ones. To this end, we make use of a Choleski decomposition which helps in handling the large number of nuisance parameters [59]. We consider chains to be converged using the Gelman-Rubin [60] criterion $R - 1 < 0.05$. The constraints on the density of ULAs today as a function of their dilution redshift $1 + z_c$ are shown in Fig. 6. These have the characteristic "belly" or U-shape first estimated in Refs. [61,62], then generated more robustly from a Boltzmann code with MCMC methods in Ref. [15], and confirmed in Ref. [45].[6]

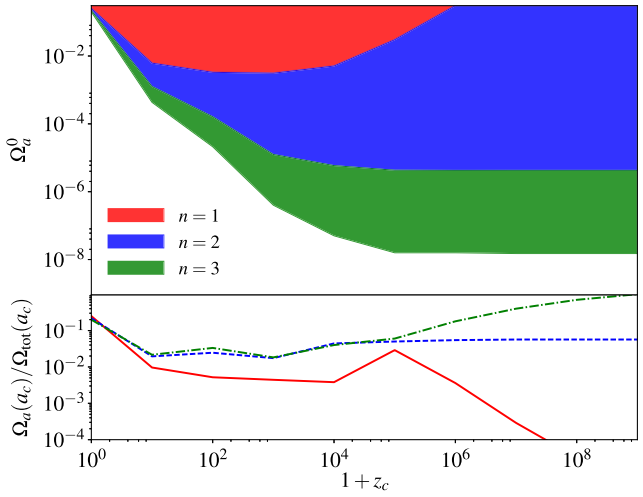

FIG. 6. Top panel—Constraints on the density of the ULA today as a function of its dilution time $1 + z_c$. Bottom panel—Constraints on the fraction of the total energy content in the form of a ULA at $a_c \equiv (1 + z_c)^{-1}$.

B. Late time constraints

Constraints on the ULA at late times are driven by measurements of the luminosity distance up to $z \simeq 1$ using the JLA data set [34] and angular diameter distance [56–58]. Note that even for $z_c = 0$ the field evolves away from $w_\phi = -1$- in particular, fitting the parametrization $w_\phi(z) = w_{a,0} + w_{a,1}[1 - 1/(1 + z)]$ to the three forms of the potential gives an axion equation of state

$$n = 1 \rightarrow w_{a,0} = -0.50, \qquad w_{a,1} = -0.79, \tag{37}$$

$$n = 2 \rightarrow w_{a,0} = -0.37, \qquad w_{a,1} = -1.18, \tag{38}$$

$$n = 3 \rightarrow w_{a,0} = -0.31, \qquad w_{a,1} = -1.36. \tag{39}$$

The values of these parameters show the behavior we expect as a function of n: as n increases the scalar field's energy density decreases more rapidly, leading to a smaller $w_{a,0}$ and $w_{a,1}$ with increasing n.

The JLA data (combined with measurements of the temperature anisotropy from *Planck*, polarization measured by WMAP and measurements of the BAO) yield a constraint of $w_0 = -0.957 \pm 0.124$ and $w_1 = -0.336 \pm 0.552$ [34], where $w(z) = w_0 + w_1[1 - 1/(1 + z)]$. If we choose

[6]This shape seems to be somewhat generic in models for which a species behaves as something other than matter up until a critical transition redshift z_c. For example, if the dark matter is generated at late times by the decay of a relativistic species, as in the late-forming dark matter model of Ref. [63], a *qualitatively* similar constraint plot results.

a small value of z_c then the ULA will behave as quintessence and contribute to driving the current epoch of accelerated expansion. Fixing the matter component at $\Omega_{m,0} = 0.3$ the equation of state of the late-time dark sector (consisting of ϕ and a cosmological constant) is given by

$$w(z) = \frac{-1 + w_a \rho_a/\rho_\Lambda}{1 + \rho_a/\rho_\Lambda}, = \frac{-1 + w_a(z)\Omega_a(z)/(0.7 - \Omega_{a,0})}{1 + \Omega_a(z)/(0.7 - \Omega_{a,0})} \quad (40)$$

Note that the cosmological constant plus ULA dark sector has $w \geq -1$.

We can then use this equation of state and fit for w_0 and w_a up to $z = 1$ to find the JLA-driven constraint on a late-time ULA. For example, if $z_c = 0$ then we find that constraints to ULAs are driven by the fact that $w_0 < -0.833$ and we find that at a 68% CL we have

$$n = 1 \rightarrow \Omega_{a,0} < 0.22, \quad (41)$$

$$n = 2 \rightarrow \Omega_{a,0} < 0.16, \quad (42)$$

$$n = 3 \rightarrow \Omega_{a,0} < 0.15. \quad (43)$$

This discussion also shows that we expect the JLA constraint to give a degeneracy between $\Omega_{a,0}$ and Ω_Λ such that $\Omega_{a,0} + \Omega_\Lambda = 1 - \Omega_m \simeq 0.7$. This simple estimate is very close, albeit slightly stronger, than what is obtained in a real analysis:

$$n = 1 \rightarrow \Omega_{a,0} < 0.25, \quad (44)$$

$$n = 2 \rightarrow \Omega_{a,0} < 0.22, \quad (45)$$

$$n = 3 \rightarrow \Omega_{a,0} < 0.20. \quad (46)$$

Note that there are some additional effects on the CMB (LISW, reduced lensing amplitude) which are well below *Planck* sensitivity but could be probed by future experiments.

C. Constraints for z_c around recombination

When the dilution begins after matter-radiation equality but before recombination, the $n = 2$ and $n = 3$ cases are basically identical; indeed, the very fast diluting fluid leaves no significant additional impact on the CMB as the universe is largely matter dominated by then. Hence, the constraints are purely driven by the EDE phase. The strongest degeneracy visible on Fig. 7—middle panel—appears to be with ω_b, which can be adjusted to counteract the effect of a faster recombination. Additional mild degeneracies appear with parameters governing the overall shape of the power spectrum $\{A_s \exp(-\tau_{\rm reio}), n_s\}$ and the amplitude of the EISW term ($\omega_{\rm cdm}$). Note that H_0 shows no degeneracy with $f_a(z_c)$. In fact $z_c = 10^3$ represents a turning point in the direction of the degeneracy; for higher value of z_c, the correlation is *positive*, and can be understood in the same manner as the degeneracy between an additional ultra-relativistic species and H_0 (e.g., Ref. [49]). For lower values of z_c however, the correlation becomes

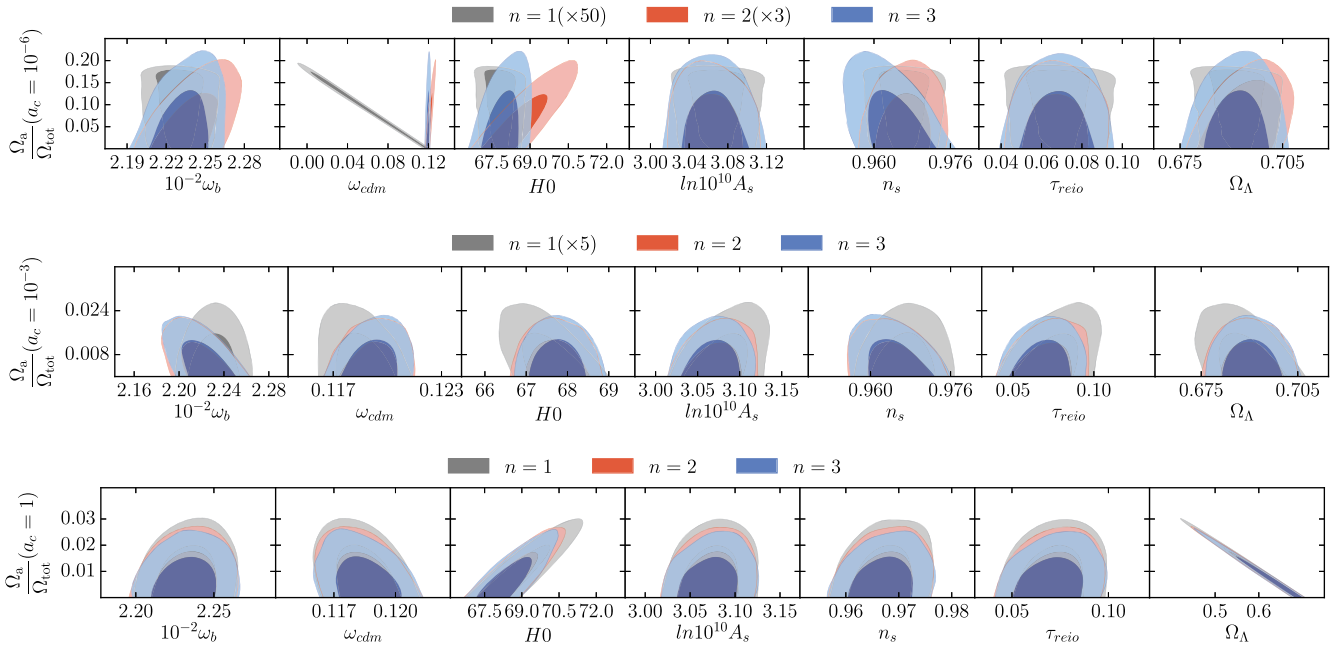

FIG. 7. Posterior distributions of the density of ULAs today vs the other ΛCDM parameters for $1 + z_c = 1$ (bottom panel), 10^3 (middle panel) and 10^6 (top panel). For some cases, the density of the ULA has been multiply by a factor written in the row's legend.

negative and is driven by the requirement of keeping the angular size of sound horizon at recombination θ_s fixed.

The $n = 1$ case, however, represents a very distinct case: as the fluid dilutes like matter, it increases the total matter component of the universe. Hence, the constraints are driven by the additional matter component and degeneracies with ΛCDM parameters can be understood accordingly. As expected, a strong negative degeneracy appears with $\omega_{\rm cdm}$, as well as with $\Omega_\Lambda = 1 - \Omega_M$ (valid in a flat universe) because Ω_M increases. Similarly to the previous case, some mild degeneracies appear with $\{A_s \exp(-\tau_{\rm reio}), n_s\}$ as to compensate the overall shape of the spectra. Finally, a strong negative correlation appears with H_0 and is due to purely geometric effects: one needs to compensate the increase in the Hubble rate ($\propto \sqrt{\Omega_M (1+z)^3}$) by decreasing H_0 in order to keep the same angular diameter distance to recombination [15].

D. Constraints for z_c earlier than matter-radiation equality

We have described in Sec. IV the effect of an early dilution on the CMB power spectra, well before matter radiation equality. The degeneracies visible on Fig. 7—bottom panel—are straightforward to understand. First and foremost, when $z_c \gtrsim 10^5$ and $n = 1$, the ULA becomes fully degenerate with a matter component. This represents a range of mass for which the axion is a valid DM candidate, as pointed out in Refs. [3,15]. Note that the degeneracy is not perfect at $z_c = z_{\rm eq}$; this is because this requirement does not ensure that $z_{\rm eq}$ is exactly fixed, z_c represents a transition redshift and the fluid does not behave exactly like matter at that time. Moreover, the CMB is sensitive to details of the expansion history around matter-radiation equality through the EISW, which further increases the value of z_c at which the ULA is degenerate (in CMB observations) with the CDM component. Note that there are no strong degeneracies between the $n = 1$ ULA and any other cosmological parameters in this case: this is expected because $\omega_{\rm cdm}$ shows no strong degeneracy with any parameters within ΛCDM.

In the $n = 2$ case, the fluid dilutes like an extra radiation component: the constraint is therefore driven by this additional relativistic species. As explained in Sec. IV, we expect a degeneracy with $N_{\rm eff}$ to some extent. Indeed, at the background level if the dilution starts early enough ($z \gtrsim 10^6$), they have exactly the same behavior. However, we confirm that the degeneracy is far from perfect because perturbations in the ULA fluid are very different from that of a free-streaming species like neutrinos. The ULA has a scale-dependent sound speed and viscosity that differs strongly from that measured by *Planck* high-ℓ, TT, TE, EE + low-P $\{c_s^2 = 0.3240 \pm 0.0060, c_{\rm vis}^2 = 0.327 \pm 0.037\}$ [1] and therefore cannot replace the totality of the non-CMB radiation bath. We find that it can account at most for ∼20% of the total $N_{\rm eff}$. Degeneracies with other parameters can be understood in a similar way as that of an additional relativistic species (e.g., Ref. [49]), and we comment in Sec. VI on the strong correlation with H_0.

Finally, for $n = 3$ the constraints come mostly from the EDE phase and are thus very similar to that of $z_c \sim 10^3$. In particular it is straightforward to show that if CMB measurements constrain the fractional ULA contribution at $z_{\rm CMB}$—the largest redshift that the CMB is sensitive to (for *Planck* $z_{\rm CMB} \sim 10^5$–10^6)—to be less than $f_{\rm CMB}$ then as $z_c \gg z_{\rm CMB}$ the limit on $\Omega_{a,0}$ asymptotes to

$$\Omega_{a,0} < \frac{f_{\rm CMB}\Omega_{r,0}}{(1-f_{\rm CMB})\sqrt{(1+z_{\rm CMB})}}. \tag{47}$$

Taking $\Omega_{r,0} = 9.2 \times 10^{-5}$ for photons, three massless neutrinos and $h = 0.68$ and setting $z_{\rm CMB} = 10^5$ and $f_{\rm CMB} = 0.06$ we find that the asymptotic constraint to $\Omega_{a,0}$ for a ULA with $n = 3$ is approximately $\Omega_{a,0} < 2 \times 10^{-8}$ which agrees well with the constraints shown in Fig. 6. We can also translate this into a constraint on the f_{z_c}:

$$f_{z_c} < \frac{f_{\rm CMB}\sqrt{\frac{1+z_c}{1+z_{\rm CMB}}}}{1 + f_{\rm CMB}\left(\sqrt{\frac{1+z_c}{1+z_{\rm CMB}}} - 1\right)}. \tag{48}$$

This expression shows that with the current constraint $f_{\rm CMB} = 0.06$ at $z_{\rm CMB} = 10^5$ we limit f_{z_c} to be less than unity as far back as $z_c = 10^{10}$. At this time constraints on the rate of expansion of the universe during big bang nucleosynthesis from measurements of the primordial light element abundances can, in principle, be used to further restrict f_{z_c} (see, e.g., Ref. [64]).

When looking at the degeneracies between the $n = 3$ ULA and other cosmological parameters the only difference relative to the other cases is with ω_b, for which the degeneracy is flipped: in that case, the EDE does not affect the recombination physics, but it decreases the damping tail of the CMB. This effect can be partially compensated by increasing ω_b. This fact also drives the degeneracy with n_s. Note that, as z_c increases, the constraint on Ω_a today flattens: this means that the constraints on $f_a(z_c)$ *relaxes* as z_c increases. This is expected because *Planck* (limited to $\ell < 2500$) is less and less sensitive to physics above $z \sim 10^5$.

E. The role of perturbations

Finally, we comment on the extent to which the details of perturbations play a role in the constraining power. In previous discussion, we have seen that the $n = 1$ case is purely degenerate with a CDM component if $z_c \gtrsim 10^5$; naturally this degeneracy disappears if we neglect perturbations in the fluid. However, when $z_c \lesssim 10^4$, we find that neglecting perturbations leads to constraints that differ by no more than ∼20%. In the $n = 2$ and $n = 3$ case, we find a similar difference.

Note however, that in the $n = 2$ case conclusions would have changed if the perturbations of the ULA were that of a free-streaming species. In that case, we would have found a perfect degeneracy for high-enough z_c that would not have been present if perturbations are neglected. It is only because the ULA is constrained to be a subdominant fraction of the universe components (and thus never drives the expansion and evolution of perturbations), that the details of their perturbations do not matter too much.

However, in the future, next generation CMB experiments and LSS surveys are expected to improve sensitivity on ULA. Hence, any detection will require an accurate description of perturbations, potentially even beyond the fluid approximation described in this paper. In future work, we will investigate the accuracy of this approximation compared to a full solution of the KG equation, with an eye towards the sensitivity levels of future CMB experiments like CMB-S4 [65].

VI. IMPLICATIONS FOR COSMOLOGICAL TENSIONS

Although most cosmological observables are individually consistent with a ΛCDM cosmology, tensions exist between the predictions of various data sets, such as the Hubble tension [1,7]. Furthermore, the recent measurement of the sky-averaged 21-cm signal by the Experiment to Detect the Global Epoch of Reionization Signature (EDGES) is inconsistent with predictions of ΛCDM [8], although the interpretation of the signal is still being explored [9]. In this section, we examine the effect of ULAs on these two tensions.

A. The Hubble tension

One of the most prominent and persistent tensions in cosmology is the Hubble tension [66,67]. The current expansion rate of the universe as predicted by the ΛCDM model when fit to the CMB disagrees with local measurements at greater than 3σ [7]. Planck determines H_0 to be 66.93 ± 0.62 km s^{-1} Mpc^{-1}, while the SH0ES (Supernova H0 for the Equation of State Collaboration) collaboration measures a value of 73.24 ± 1.74 km s^{-1} Mpc^{-1} [7]. Numerous explanations have been proposed and studied in the literature [4,7,68–75].

In this section, we investigate whether ULAs can alleviate the tension and what regions in the $\Omega_a - z_c$ plane are best suited to do so, similar to Ref. [4]. We use the Friedmann equation to compute H_0 today, given fiducial values for the other cosmological parameters, and the indicated values for z_c and Ω_a. We keep θ_s fixed and let CLASS solve for the value of H_0. The results are shown in Fig. 8.

For $n = 1$, we find that no value of $\Omega_{a,0}$ for values of $1 + z_c \in [10^0, 10^6]$ diminished the H_0 tension. With reference to Fig. 7 and Sec. V D, for $z_c \gg z_{\rm eq}$, the fluid is fully degenerate with CDM, and $\omega_{\rm cdm}$ and Ω_a are negatively correlated. That is, the CMB cannot distinguish between the fluid and CDM. An increase in the energy density of the fluid today will be accompanied by a decrease in the energy density of CDM and there is no change to the value of the Hubble parameter. At the other end, for $z_c \simeq 0$, the fluid is strongly degenerate with Λ. This degeneracy is weaker than

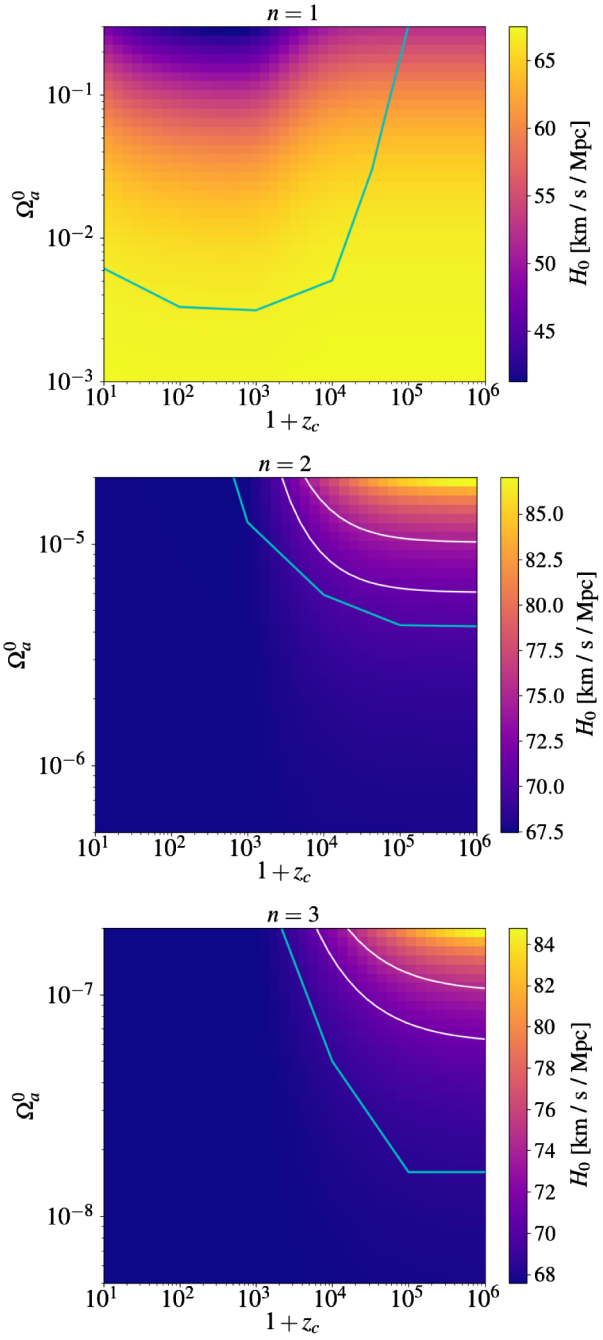

FIG. 8. Hubble parameter H_0 for various values of Ω_a^0 and z_c, for the $n = 1$ (top panel), $n = 2$ (middle panel) and $n = 3$ case (bottom panel). The cyan line represents the constraints shown in Fig. 6. The white contours show the 1σ contour on the H_0 value measured by SH0ES.

that with $\omega_{\rm cdm}$ at $z_c \gg z_{\rm eq}$, because the equation of state parameter of the fluid is not exactly -1, as discussed in Sec. V B. Again, an increase in $\Omega_{\rm a}$ is accompanied with a decrease in Ω_Λ and the value of H_0 remains unaltered. The tension is, however, somewhat alleviated as the fluid is degenerate with H_0 and leads to a larger error on H_0. For intermediate redshifts $z_c \lesssim z_{\rm eq}$, the fluid reduces the value of H_0, exacerbating the tension. *Planck* data constrain especially well the combination $\Omega_m h^3$ [1] and therefore, effectively increasing Ω_m leads to a reduction in h. Hence at best, the $n = 1$ scenario leaves H_0 unaltered, at worst, exacerbates the tension.

The $n = 2$ scenario fares better, as seen from Fig. 8. For $z_c < z_*$, it fares similarly to the $n = 1$ case. It is strongly degenerate with Ω_Λ for $z_c \simeq 0$. For $0 < z_c \ll z_*$, the fluid exacerbates the tension. Again, this is due to its effect on $D_A(z_*)$—it adds to the expansion rate at late times and H_0 must decrease to compensate and preserve $D_A(z_*)$. For $z_c \simeq 10^3$, as mentioned in Sec. V C, Ω_a and H_0 are uncorrelated. As we are already in matter domination by $z = 10^3$, a fluid that behaves like Λ before and radiation after recombination will impact expansion history only over a finite redshift range around z_c. As the angular diameter distance $D_A(z_*)$ to recombination gets most of its contribution from lower redshifts, its value and therefore H_0 remain largely unchanged. For $z_c > z_*$, the $n = 2$ scenario is degenerate with $N_{\rm eff}$, as it effectively adds more radiation to the Universe. Hence the impact of the fluid on H_0 is similar to that of $N_{\rm eff}$: it increases H_0 and diminishes the tension [7]. However, our CMB constraints disfavor this solution.

Finally, for the $n = 3$ scenario, for $z_c \lesssim z_*$, the impact of the fluid is similar to the $n = 2$ case. As mentioned before, the $n = 3$ case only impacts expansion history over a small range in redshift centered around z_c. For $z_c > z_*$ and Ω_a^0 larger than our current constraints, pre-recombination expansion rate is increased. This decreases the radius r_s of the sound horizon at recombination and H_0 increases to compensate and preserve θ_s. Hence, the fluid is capable of increasing H_0 as seen in Fig. 8, but for values of Ω_a^0 that are much larger than our constraints.

The CMB becomes insensitive to physics above $z \sim 10^6$ as noted by [4,20]. Therefore, for a given Ω_a^0, even as z_c increases above 10^6, the energy density of the fluid for $z \lesssim 10^6$ remains unchanged, as does the Hubble parameter. We hence only show the change to the Hubble parameter due to the addition of ULAs up to $1 + z_c = 10^6$.

To summarize, we find that in order for ULAs to diminish the Hubble tension, with $n = 2$ and 3, it requires $z_* < z_c \lesssim 10^6$ and Ω_a larger than our constraints. Still, in the $n = 2$ case the tension is reduced from $\sim 3.5\sigma$ to less than 2σ, which is a more significant easing of this tension than a relativistic species with arbitrary sound speed and viscosity (see e.g., [76,77]).

B. EDGES exotic 21 cm measurement

EDGES recently measured the sky-averaged 21 cm brightness temperature [8] around the redshift range $z = 15$–20 to be roughly 2.5 times smaller (3.8σ) than that predicted by ΛCDM.[7] Two main classes of solution have been suggested to explain this measurement: either the temperature of the photons against which the 21-cm temperature of the gas is measured is brighter than that of the CMB [78–80] or the baryon temperature T_b is cooler than expected based on ΛCDM [5,8,81,82]. In the latter scenario, the EDGES measurement indicates that the baryon temperature T_b at $z = 20$ is smaller than 7K at 99% C.L.

In Ref. [5], the implications of EDGES were explored for an EDE model equivalent to the limit $n \to \infty$, including only the effect of EDE on the homogeneous evolution of densities and temperatures. Here we perform a similar analysis for $n = 1$, 2, and 3, including perturbations in a ULA fluid.

In the absence of any additional sources, the baryon gas temperature is driven by the balance between Compton heating and Hubble cooling

$$\frac{dT_b}{dz} = \frac{T_b(z) - T_{\rm CMB}(z)}{(1+z)H(z)t_C(z)} + \frac{2T_b(z)}{(1+z)}. \tag{49}$$

where $t_C(z)$ is the Compton-heating timescale. The key idea used in Ref. [5] is that, if the expansion rate before $z \sim 20$ is increased, the gas temperature decouples from the CMB temperature earlier, giving the gas more time to adiabatically cool. Within ΛCDM, baryons decouple around $z \sim 150$. To reach the 99% C.L. upper limit on the level of absorption measured by EDGES at $z \sim 20$, the decoupling would need to happen around $z \sim 210$. The presence of a ULA that would dominate the expansion rate over a short period of time can potentially lead to a decoupling satisfying this condition.

We show in Fig. 9 the baryon temperature at $z = 20$ (close to the minimum of the absorption trough measured by EDGES [8]) as a function of the ULA density today Ω_a^0 and critical redshift z_c, for each value of n. To produce this figure, we fixed all ΛCDM parameters including the Hubble rate H_0 to values compatible with *Planck* 2015 data.[8] Interestingly, we confirm that there exists a region of parameter space, centered around $z_c \sim 100$ where the

[7]The proper interpretation of this measurement is still under discussion [9].

[8]In CLASS, these equations can be solved using either *Recfast* [83,84] or *HyRec* [85] and Eq. (49). Our choice of keeping H_0 fixed is motivated by the fact that adjusting θ_s requires strongly un-physical values of the Hubble rate (sometimes smaller than 0.01 km/s/Mpc) for which both *Recfast* and *HyRec* have difficulties to solve the cosmological recombination history. This also allows for a direct comparison with Ref. [5] where the same approach was used.

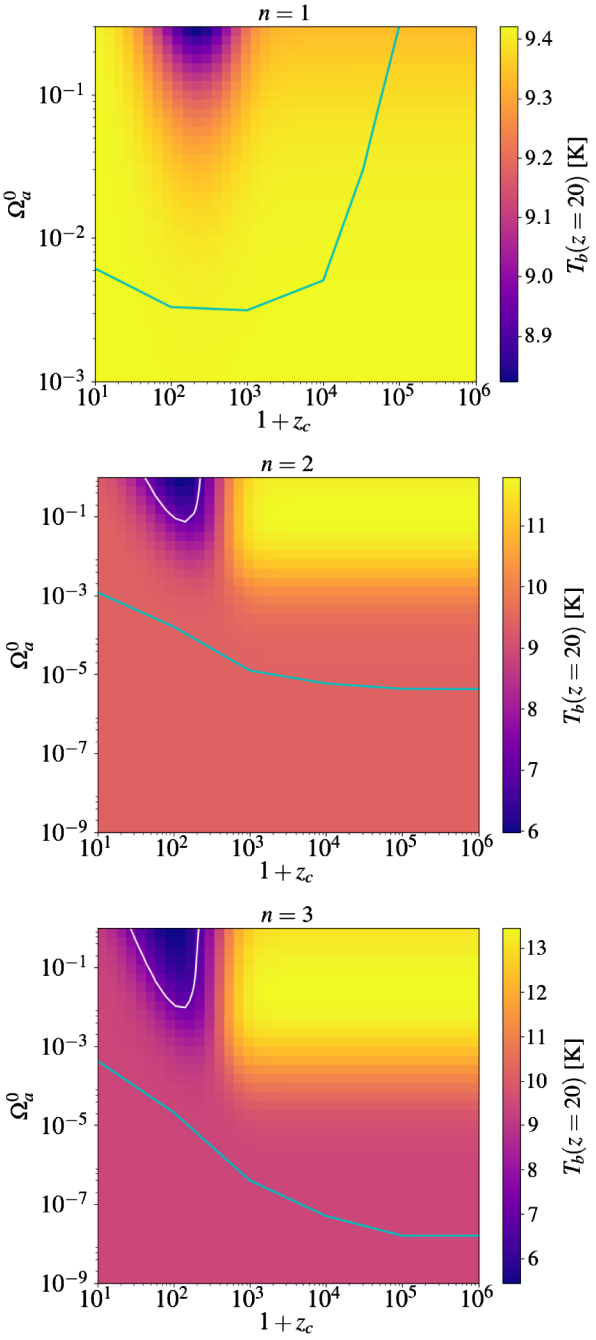

FIG. 9. Baryon temperature at $z = 20$ (close to the minimum of the absorption trough measured by EDGES [8]) as a function of the ULA density today Ω_a^0 and critical redshift z_c. The top panel presents the $n = 1$ case, the middle panel the $n = 2$ case and the bottom panel the $n = 3$ case. The white line shows $T_b = 7$ K, i.e., the 99% upper-limit on the temperature measured by EDGES. The cyan line shows the *Planck* 95% C.L. limit derived in this work. All models of interest are excluded by our analysis.

EDGES signal can be explained, in the $n = 2$ and $n = 3$ case. Our constraints on the ULA density from *Planck* data however strongly exclude all of these models, in agreement with Ref. [5].

VII. CONCLUSIONS

In this paper, we have studied the impact of ULAs on cosmological observations as they become dynamical at different times. We have considered potentials of the form $V_n(\phi) \propto (1-\cos\phi)^n$, which show a wide variety of phenomenological consequences. At early times, each field is frozen in its potential due to Hubble friction such that their equations of state are dark-energy like, i.e., $w_a \simeq -1$. Once Hubble friction becomes weak enough, the field becomes dynamical and eventually starts to oscillate at the bottom of its potential. Once averaged over the oscillation period, the potential leads to an equation of state equal to $w_a \simeq (n-1)/(n+1)$.

Such fields had been previously invoked in several contexts. First, ULAs with $n = 1$ and becoming dynamical at early times ($z \gtrsim 10^5$) are known to be a viable DM candidate. On the other hand, ULAs still frozen today are a viable dark energy candidate [3,15]. Second, a statistical ensemble of such fields may alleviate the coincidence problem today [4,19,21,22]. This general scenario may also provide a way to connect the physics of cosmic inflation to our current period of accelerated expansion [21]. Third, the presence of an EDE can possibly reduce the Hubble constant tension [4] and explain the anomalously low baryon temperature inferred by the EDGES experiment [5]. More generally, the presence of a collection of scalar fields (beyond the sole DM or DE) is a generic feature of the axiverse which gives us additional handle on this scenario and should thus be looked for.

We have extended these previous studies in several significant ways. First, we have presented a parametrization of the ULA dynamics in terms of the redshift when the field becomes dynamical, z_c, and the fractional energy density in the axion field at z_c, f_{z_c}. Second of all, we have extended the effective fluid formalism for ULAs to anharmonic ULA potentials. These perturbations can be approximately described by a time-averaged fluid component with a time and scale dependent effective sound speed [3,15,23–27] within the "generalized dark matter" parametrization [28]. Recently, alternative formalisms have been proposed to deal with the numerical challenges that arise when modeling perturbations to a coherently oscillating scalar field in the expanding universe [86]. Under an appropriate change of variables, these are equivalent to the effective fluid formalism used here, modulo subtleties to be explored in future work.

Moreover, we derived a mapping between this parametrization and the ULA mass, decay constant and initial field value and attested of the accuracy of our fluid approximation by direct comparison with the exact KG solution in Appendix B. We have also shown that this WKB approximation is strictly only valid for potentials with $n \leq 3$, otherwise the period of oscillation is shorter than a Hubble time, violating the WKB assumptions.

Second, equipped with this fluid formalism we have compared the phenomenological consequences of

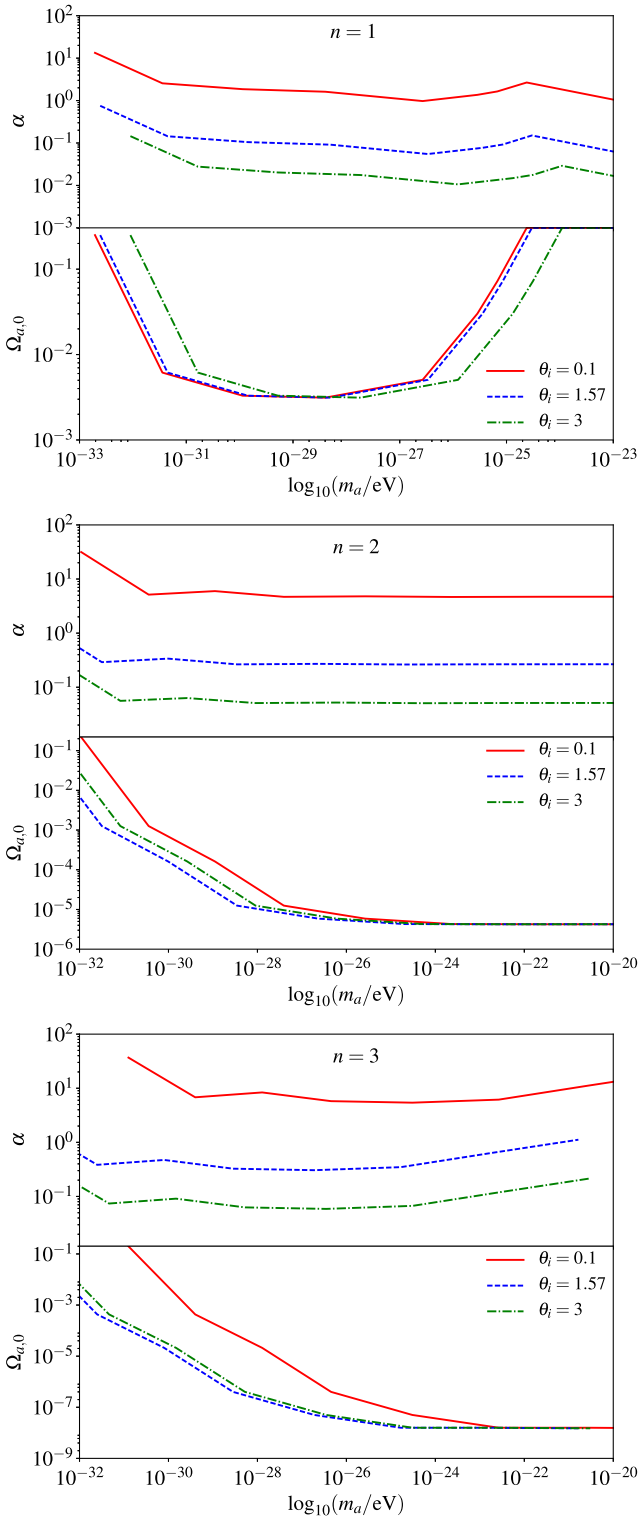

FIG. 10. Constraints on the axion model parameters (μ, α). We illustrate the impact of the initial field values by choosing $\theta_i = 0.1, \pi/2, 3$.

axionlike potentials with $n = 1$ which dilutes as cold dark matter (CDM), $n = 2$ which dilutes as radiation, and $n = 3$ which dilutes faster than radiation. We were thus able to explore any degeneracy the ULAs may have with known cosmological components, in particular CDM and neutrinos, and quantify the sensitivity of the data to a ULA component that decays even faster than radiation. We have constrained the abundance of ULAs as a function of z_c using current cosmological data sets with a MCMC analysis, in order to fully explore degeneracies between the ULA parameters and the standard cosmological parameters. Remarkably, the details of the ULA effective sound speed could distinguish the effects of a ULA from other cosmological components, even if the ULA time-averaged equation of state is equal to zero (CDM-like) or $1/3$ (radiationlike). Moreover, we have found that the CMB is sensitive to the field becoming dynamical as early as $z_c \sim 10^{10}$. Interestingly, such an EDE could change the frequency spectrum of gravitational waves generated by phase transitions in the early universe, and could have implications for experimental efforts like LIGO and LISA [87], a possibility to be explored in future work.

We illustrate in Fig. 10 how the constraints derived in this work in the (z_c, Ω_a^0)—plane translate onto constraints on the axion parameters, i.e., (μ, α). We use the relations introduced in Sec. II D to map the ULA parameters to our fluid formalism. As an example, we choose three different initial field values, namely $\theta_i = 0.1, \pi/2, 3$. The smallest value of θ_i allows direct comparison with results from Refs. [15,45] (derived in the quadratic approximation), while the two others show how these constraints vary with the initial field value.

Finally, we have studied the implications of our constraints for cosmological tensions. We have shown that fields with $n = 2$ and $n = 3$ can significantly ease the tension, as previously found for $n \to \infty$. However, our results put this scenario under strong pressure. On the one hand, the explanation of the EDGES signal is excluded by more than three orders of magnitude. On the other hand, we find that ULAs could at best ease the H_0 tension from $\sim 3.4\sigma$ to $\sim 2\sigma$ given the level of our constraints. Contrary to expectation, the $n = 2$ scenario is favored over $n = 3$ even if the latter dilutes faster. This scenario also does slightly better than a relativistic species with arbitrary sound speed and viscosity, which can only relax the tension at the 3σ level [76,77].

Our formalism represents a state-of-the-art treatment of the effect of ULAs on cosmological observables and can be used safely to analyze *Planck* data. In the future, CMB and LSS experiments with yet unreached precision will be built. It is still to be established whether the fluid approximation will be accurate enough to describe the impact of ULAs without introducing strong bias in the reconstruction of cosmological parameters. However, this formalism is essential in order to perform extensive MCMC scan given the difficulty of solving the full KG equations. It will therefore be necessary to further develop methods to calculate the cosmological implications of ULAs as observations become more precise.

ACKNOWLEDGMENTS

This research project was conducted using computational resources at the Maryland Advanced Research Computing Center (MARCC). Part of this work has been done thanks to the facilities offered by the Université Savoie Mont-Blanc MUST computing center. V. P. and T. K. were supported at Johns Hopkins by NSF Grant No. 0244990, NASA NNX17AK38G, and the Simons Foundation. T. L. S. and D. G. acknowledge support in part by NASA ATP Grant No. 17-ATP17-0162. T. L. S. acknowledges support from the Provost's office at Swarthmore College. D. G. acknowledges support from the Provost's office at Haverford College. This work was supported in part by the National Science Foundation under Grant No. NSF PHY-1125915 at the Kavli Institute for Theoretical Physics (KITP) at UC Santa Barbara. We thank D. J. E. Marsh for a thorough and insightful reading of the manuscript. We thank Charlotte Owen for her help in computing the exact ULA dynamics, as well as her very careful reading of the manuscript. D. G. thanks KITP for its hospitality during the completion of this work.

APPENDIX A: DERIVATION OF THE TIME-AVERAGED EFFECTIVE SOUND SPEED FOR A GENERIC OSCILLATING POTENTIAL

Here we derive the effective sound speed, following the covariant perturbation theory notation used in Refs. [28,88]. We can write the linearly perturbed Friedman-Robertson-Walker (FRW) metric as

$$g^{00} = -a^{-2}(1-2A), \tag{A1}$$

$$g^{0i} = -a^{-2}B^i, \tag{A2}$$

$$g^{ij} = a^{-2}(\gamma^{ij} - 2H_L\gamma^{ij} - 2H_T^{ij}), \tag{A3}$$

where $\gamma_{ij}dx^i dx^j = d\chi^2 + \chi^2 d\Omega$ and χ is the comoving distance. Using conformal time, the equation of motion for the linear perturbation of the axion field is given by

$$\ddot{\phi}_1 + 3H\dot{\phi}_1 + \left(\frac{k^2}{a^2} + V''\right)\phi_1 = (\dot{A} + 3\dot{H}_L - k/aB)\dot{\phi}_0 - 2AV', \tag{A4}$$

where B is the longitudinal part of B^i. In synchronous gauge we have $A = B = 0, \eta \equiv -1/3H_T - H_L, h \equiv 6H_L$, where η and h are the metric variables used in Ref. [43].

We can write the density, pressure, and velocity perturbations in the scalar field stress energy as

$$\delta\rho_a = (\dot{\phi}_0\dot{\phi}_1 - \dot{\phi}_0^2 A) + V'\phi_1, \tag{A5}$$

$$\delta P_a = (\dot{\phi}_0\dot{\phi}_1 - \dot{\phi}_0^2 A) - V'\phi_1, \tag{A6}$$

$$T_i^0 = \nabla_i Q_\phi = \nabla_i\dot{\phi}_0\phi_1, \tag{A7}$$

where $Q_\phi \equiv (\rho_a + p_a)(v_a - B)$. We suppose that the field is oscillating about the minimum of its potential with a frequency $\varpi \gg H$ and we want to find the sound speed in the axion's average "rest frame." In this rest frame when averaging over the fast oscillations we have

$$\langle T_i^0\rangle = 0 \rightarrow \langle\dot{\phi}_0\phi_1\rangle = 0, \tag{A8}$$

which fixes the gauge condition for the metric perturbation B. We also require that in the axion rest frame the time-averaged axion heat-flux is locally conserved:

$$\left\langle\left[\frac{\partial}{\partial\eta} + 4\mathcal{H}\right]Q_a\right\rangle = 0 \tag{A9}$$

which, through the Euler equation for the axion stress energy, implies our second gauge condition [24,88]

$$\langle\rho_a + P_a\rangle A = -\langle\delta P_a\rangle. \tag{A10}$$

We can write the linearly perturbed axion energy density as

$$\delta\rho_a = \dot{\phi}_0\dot{\phi}_1 - (\rho_a + P_a)A + V'\phi_1, \tag{A11}$$

$$\delta P_a = \dot{\phi}_0\dot{\phi}_1 - (\rho_a + P_a)A - V'\phi_1, \tag{A12}$$

which along with our gauge condition in Eq. (A10) gives

$$\varpi^2\langle\phi_0\phi_1\rangle = \langle V'\phi_1\rangle, \tag{A13}$$

$$\langle\delta\rho_a\rangle = \langle\delta P_a\rangle + 2\varpi^2\langle\phi_0\phi_1\rangle. \tag{A14}$$

Keeping only the terms which vary on the (short) oscillation timescale, the perturbed Klein-Gordon equation is

$$\ddot{\phi}_1 + \left(\frac{k^2}{a^2} + V''\right)\phi_1 \simeq -2AV'. \tag{A15}$$

Multiplying this equation by ϕ_0 and averaging over the short period we have

$$-\varpi^2\langle\phi_1\phi_0\rangle + \frac{k^2}{a^2}\langle\phi_1\phi_0\rangle + (2n-1)\langle V'\phi_1\rangle \simeq -4An\langle V\rangle \tag{A16}$$

Finally, the virial theorem allows us to write

$$\langle\rho_a\rangle = (n+1)\langle V\rangle, \tag{A17}$$

$$\langle P_a\rangle = (n-1)\langle V\rangle, \tag{A18}$$

so that

$$\langle \rho_a + P_a \rangle = 2n\langle V \rangle, \tag{A19}$$

and the Klein-Gordon equation can be written

$$\left(\frac{k^2}{a^2} + 2(n-1)\varpi^2\right)\langle \phi_1 \phi_0 \rangle \simeq 2\langle \delta P_a \rangle. \tag{A20}$$

This allows us to write

$$c_s^2 \equiv \frac{\langle \delta P_a \rangle}{\langle \delta \rho_a \rangle} = \frac{2a^2(n-1)\varpi^2 + k^2}{2a^2(n+1)\varpi^2 + k^2}. \tag{A21}$$

The effective sound speed is computed in a gauge where

$$B = \langle v_a \rangle, \tag{A22}$$

$$A = -\frac{\langle \delta P_a \rangle}{\langle \rho_a + P_a \rangle}, \tag{A23}$$

but we are doing our calculations in synchronous gauge where $A = B = 0$. Next we will show that by transforming to synchronous gauge the effective sound speed enters into the fluid dynamics as dictated by the GDM equations of motion [28].

A general gauge transformation takes the form

$$\eta = \tilde{\eta} + T, \tag{A24}$$

$$x^i = \tilde{x}^i + L^i, \tag{A25}$$

which leads to a transformation of the scalar metric potentials

$$A = \tilde{A} - \dot{T} - \mathcal{H}T, \tag{A26}$$

$$B = \tilde{B} + \dot{L} + kT, \tag{A27}$$

$$H_L = \tilde{H}_L - \frac{k}{3}L - \mathcal{H}T, \tag{A28}$$

$$H_T = \tilde{H}_T + kL, \tag{A29}$$

and transformation of the components of the stress-energy tensor

$$\delta\rho_a = \delta\tilde{\rho}_a - \dot{\rho}_a T, \tag{A30}$$

$$\delta P_a = \delta\tilde{P}_a - \dot{P}_a T, \tag{A31}$$

$$v_a = \tilde{v}_a + \dot{L}. \tag{A32}$$

This tells us that to transform from our comoving gauge to synchronous gauge where $B = 0$ we must have

$$\dot{L} + kT = -\langle v_a \rangle, \tag{A33}$$

which in turn, using the transformation for the velocity, implies

$$T = -v_a/k, \tag{A34}$$

where v_ϕ is the axion velocity perturbation in synchronous gauge.

In order to determine how c_s^2 affects the evolution of the averaged field in synchronous gauge we now compute the synchronous gauge entropy perturbation, $P_a\Gamma_a \equiv \delta P_a - c_{\rm a}^2\delta\rho_a$, where $c_{\rm a}^2 = \dot{P}_a/\dot{\rho}_a$, in terms of the averaged field variable in the comoving gauge. We start with an expression for the pressure perturbation in synchronous gauge:

$$\delta P_a = \langle \delta P_a \rangle + \dot{P}_a v_a/k, \tag{A35}$$

$$= c_s^2\langle \delta\rho_a \rangle + \dot{P}_a v_a/k, \tag{A36}$$

where in the second line we have used the effective sound speed. Next we write the comoving density perturbation in terms of the synchronous density perturbation and use the homogeneous continuity equation:

$$\delta P_a = c_s^2(\delta\rho_a - \dot{\rho}_a v_\phi/k) + \dot{P}_a v_a/k, \tag{A37}$$

$$= c_s^2\delta\rho_a + 3\mathcal{H}(1+w_a)\rho_a(c_s^2 - c_{\rm ad}^2)v_a/k. \tag{A38}$$

This leads to

$$P_a\Gamma_a = (c_s^2 - c_{\rm a}^2)[\delta\rho_a + 3\mathcal{H}(1+w_a)\rho_a v_a/k]. \tag{A39}$$

This implies that we can use the GDM equations of motion to approximate the evolution of the perturbations in the axion field with an effective sound-speed which transitions from $c_s^2 = 1$ for $z > z_c$ to Eq. (A21) for $z < z_c$.

APPENDIX B: APPROXIMATE VS. EXACT DYNAMICS

To check the validity of our fluid approach, we compare it to the solution of the full KG equations for specific (arbitrary) values $(\mu, \alpha) = (10^5, 0.05)$. We choose an initial field value $\Theta_i = 1.5, 2.5, 3$ for $n = 1, 2, 3$ respectively. We use the relations introduced in Sec. II D to map the ULA parameters to our fluid formalism. From a given mu, α and Θ_i, we can easily calculate $\Omega_a(z_c)$ and x_c. We then make use of a shooting method in order to achieve Eq. (29) (that cannot be solved analytically except if we assume that a single species dominates the universe energy content). We have checked that changing these parameters do not affect our conclusions. We plot the evolution of density perturbations with wave number $k = 1, 10^{-3}$ Mpc^{-1} in Fig. 11. The impact of our approximation on the CMB and power

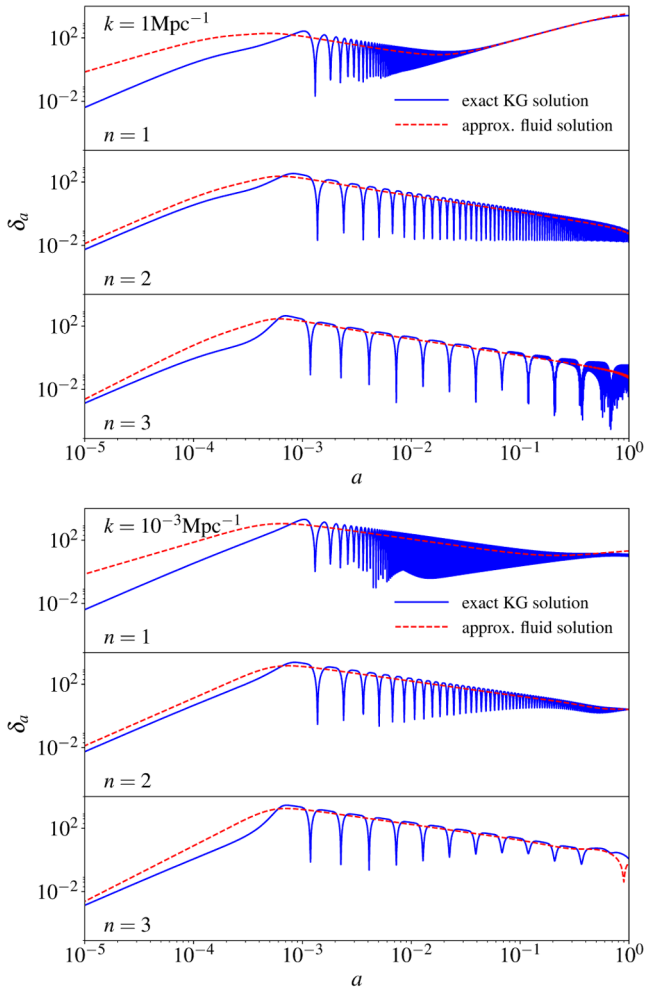

FIG. 11. The exact and approximate evolution of density perturbations with wave number $k = 1$ Mpc^{-1} (top panel) and $k = 10^{-3}$ Mpc^{-1} (bottom panel) for $n = 1$, 2, 3 and $(\mu, \alpha) = (10^5, 0.05)$. The initial field values Θ_i were set to 1.5,2.5,3, respectively.

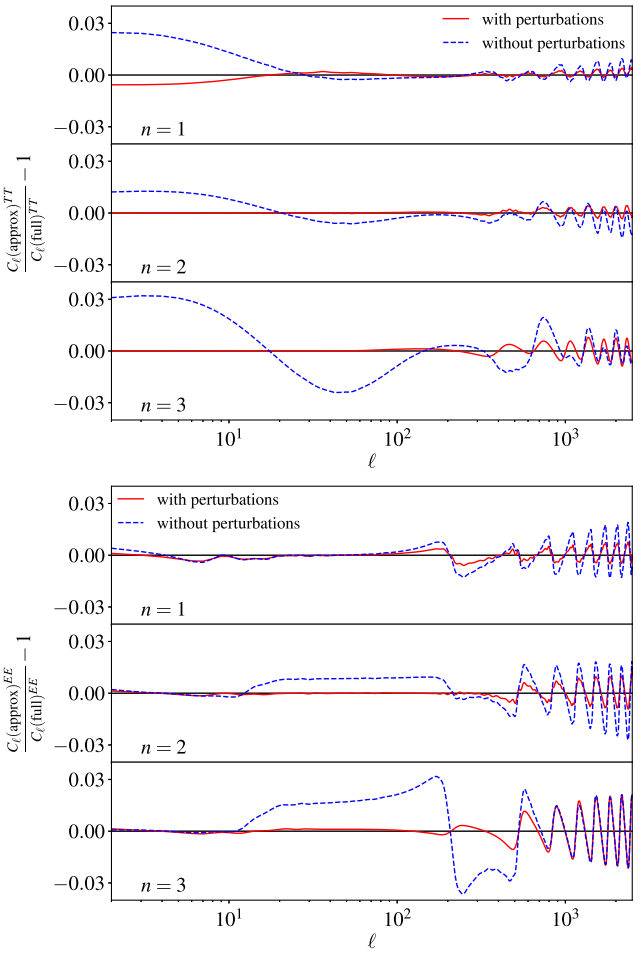

FIG. 12. Residuals of the CMB TT (top panel) and EE (bottom panel) power spectra calculated in the fluid approximation with respect to solving exactly the KG equations for $n = 1$, 2, 3 and $(\mu, \alpha) = (10^5, 0.05)$. The initial field values Θ_i were set to 1.5,2.5,3. We show the case of neglecting perturbations of the ULA for comparison.

spectra is shown in Fig. 12. We also show the case of neglecting perturbations of the ULA for comparison.

By looking at Fig. 11, one can see that our parametrization captures well the overall behavior of the density perturbations. While it fails at following all of the oscillations, the envelope (i.e., the amplitude) of these is well reproduced. The agreement improves when the ULA starts oscillating, since our parametrization is designed for that regime.

In Fig. 12, one can see that the CMB TT and EE power spectra are calculated at a few percent accuracy. The agreement is better for $n = 1$ (it is always below a percent point) and degrades when going to higher power of n. This is expected as the WKB approximation, valid when the field oscillations are much more rapid than the Hubble time, breaks-down for $n \geq 3$. One can also gauge the impact of including perturbations: it is particularly important to avoid creating large deviations at multipoles $\ell \lesssim 100$. Remarkably, below multipoles of a few hundred the agreement is always well below a percent when including perturbations. Perturbations also have an impact at high multipoles, especially in the $n = 1$ and 2 case, but does not improve the agreement very significantly in the $n = 3$ case. From this quick comparison, we conclude that it is safe to use our parametrization given the precision of *Planck* data and the fact that we merely derive constraints on the ULAs abundances. However, we note that given the accuracy of next generation CMB experiments at high multipoles, searches for ULA in future cosmological data might require the evolution of the full KG equations (especially in the $n > 1$ cases).

[1] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **594**, A13 (2016).
[2] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, and J. March-Russell, Phys. Rev. D **81**, 123530 (2010).
[3] D. J. E. Marsh, Phys. Rep. **643**, 1 (2016).
[4] T. Karwal and M. Kamionkowski, Phys. Rev. D **94**, 103523 (2016).
[5] J. C. Hill and E. J. Baxter, J. Cosmol. Astropart. Phys. 08 (2018) 037.
[6] S. M. Feeney, D. J. Mortlock, and N. Dalmasso, Mon. Not. R. Astron. Soc. **476**, 3861 (2018).
[7] A. G. Riess *et al.*, Astrophys. J. **826**, 56 (2016).
[8] J. D. Bowman, A. E. E. Rogers, R. A. Monsalve, T. J. Mozdzen, and N. Mahesh, Nature (London) **555**, 67 (2018).
[9] R. Hills, G. Kulkarni, P. D. Meerburg, and E. Puchwein, arXiv:1805.01421.
[10] B. Bozek, D. J. E. Marsh, J. Silk, and R. F. G. Wyse, Mon. Not. R. Astron. Soc. **450**, 209 (2015).
[11] A. Lidz and L. Hui, Phys. Rev. D **98**, 023011 (2018).
[12] P. Svrcek and E. Witten, J. High Energy Phys. 06 (2006) 051.
[13] M. J. Stott, D. J. E. Marsh, C. Pongkitivanichkul, L. C. Price, and B. S. Acharya, Phys. Rev. D **96**, 083510 (2017).
[14] M. Cicoli, M. Goodsell, and A. Ringwald, J. High Energy Phys. 10 (2012) 146.
[15] R. Hlozek, D. Grin, D. J. E. Marsh, and P. G. Ferreira, Phys. Rev. D **91**, 103512 (2015).
[16] S. Mukherjee, R. Khatri, and B. D. Wandelt, J. Cosmol. Astropart. Phys. 04 (2018) 045.
[17] R. Kappl, H. P. Nilles, and M. W. Winkler, Phys. Lett. B **753**, 653 (2016).
[18] M. S. Turner, Phys. Rev. D **28**, 1243 (1983).
[19] K. Griest, Phys. Rev. D **66**, 123501 (2002).
[20] E. V. Linder and T. L. Smith, J. Cosmol. Astropart. Phys. 04 (2011) 001.
[21] M. Kamionkowski, J. Pradler, and D. G. E. Walker, Phys. Rev. Lett. **113**, 251302 (2014).
[22] R. Emami, D. Grin, J. Pradler, A. Raccanelli, and M. Kamionkowski, Phys. Rev. D **93**, 123005 (2016).
[23] W. Hu, R. Barkana, and A. Gruzinov, Phys. Rev. Lett. **85**, 1158 (2000).
[24] J.-c. Hwang and H. Noh, Phys. Lett. B **680**, 1 (2009).
[25] D. J. E. Marsh and P. G. Ferreira, Phys. Rev. D **82**, 103528 (2010).
[26] C.-G. Park, J.-c. Hwang, and H. Noh, Phys. Rev. D **86**, 083535 (2012).
[27] H. Noh, J.-c. Hwang, and C.-G. Park, Astrophys. J. **846**, 1 (2017).
[28] W. Hu, Astrophys. J. **506**, 485 (1998).
[29] D. H. Lyth, Phys. Rev. D **45**, 3394 (1992).
[30] J. A. R. Cembranos, A. L. Maroto, and S. J. Núñez Jareño, J. High Energy Phys. 03 (2016) 013.
[31] F. X. L. Cedeño, A. X. González-Morales, and L. A. Ureña López, Phys. Rev. D **96**, 061301 (2017).
[32] V. Desjacques, A. Kehagias, and A. Riotto, Phys. Rev. D **97**, 023529 (2018).
[33] J. Fan, Phys. Dark Universe **14**, 84 (2016).
[34] M. Betoule *et al.* (SDSS Collaboration), Astron. Astrophys. **568**, A22 (2014).
[35] J. Lesgourgues, arXiv:1104.2932.
[36] D. Blas, J. Lesgourgues, and T. Tram, J. Cosmol. Astropart. Phys. 07 (2011) 034.
[37] J. Lesgourgues, arXiv:1104.2934.
[38] J. Lesgourgues and T. Tram, J. Cosmol. Astropart. Phys. 09 (2011) 032.
[39] B. Audren, J. Lesgourgues, K. Benabed, and S. Prunet, J. Cosmol. Astropart. Phys. 02 (2013) 001.
[40] B. Li, T. Rindler-Daller, and P. R. Shapiro, Phys. Rev. D **89**, 083536 (2014).
[41] B. Li, P. R. Shapiro, and T. Rindler-Daller, Phys. Rev. D **96**, 063505 (2017).
[42] M. C. Johnson and M. Kamionkowski, Phys. Rev. D **78**, 063010 (2008).
[43] C.-P. Ma and E. Bertschinger, Astrophys. J. **455**, 7 (1995).
[44] G. Ballesteros and J. Lesgourgues, J. Cosmol. Astropart. Phys. 10 (2010) 014.
[45] R. Hlozek, D. J. E. Marsh, and D. Grin, Mon. Not. R. Astron. Soc. **476**, 3063 (2018).
[46] L. Visinelli, Phys. Rev. D **96**, 023013 (2017).
[47] W. Hu, N. Sugiyama, and J. Silk, Nature (London) **386**, 37 (1997).
[48] L. Kofman and A. A. Starobinsky, Pis'ma Astron. Zh. **11**, 643 (1985) [Sov. Astron. Lett. **11**, 271 (1985)].
[49] Z. Hou, R. Keisler, L. Knox, M. Millea, and C. Reichardt, Phys. Rev. D **87**, 083008 (2013).
[50] D. Baumann, D. Green, J. Meyers, and B. Wallisch, J. Cosmol. Astropart. Phys. 01 (2016) 007.
[51] T. L. Smith, S. Das, and O. Zahn, Phys. Rev. D **85**, 023001 (2012).
[52] I. Sendra and T. L. Smith, Phys. Rev. D **85**, 123002 (2012).
[53] B. Audren *et al.*, J. Cosmol. Astropart. Phys. 03 (2015) 036.
[54] B. Follin, L. Knox, M. Millea, and Z. Pan, Phys. Rev. Lett. **115**, 091301 (2015).
[55] E. Sellentin and R. Durrer, Phys. Rev. D **92**, 063012 (2015).
[56] S. Alam *et al.* (BOSS Collaboration), Mon. Not. R. Astron. Soc. **470**, 2617 (2017).
[57] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, Mon. Not. R. Astron. Soc. **416**, 3017 (2011).
[58] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera, Mon. Not. R. Astron. Soc. **449**, 835 (2015).
[59] A. Lewis, Phys. Rev. D **87**, 103529 (2013).
[60] A. Gelman and D. B. Rubin, Stat. Sci. **7**, 457 (1992).
[61] J. A. Frieman, C. T. Hill, A. Stebbins, and I. Waga, Phys. Rev. Lett. **75**, 2077 (1995).
[62] L. Amendola and R. Barbieri, Phys. Lett. B **642**, 192 (2006).
[63] A. Sarkar, S. Das, and S. K. Sethi, J. Cosmol. Astropart. Phys. 03 (2015) 004.
[64] S. M. Carroll and M. Kaplinghat, Phys. Rev. D **65**, 063507 (2002).
[65] K. N. Abazajian *et al.* (CMB-S4 Collaboration), arXiv:1610.02743.
[66] W. L. Freedman, Nat. Astron. **1**, 0121 (2017).
[67] J. L. Bernal, L. Verde, and A. G. Riess, J. Cosmol. Astropart. Phys. 10 (2016) 019.
[68] E. Aubourg *et al.*, Phys. Rev. D **92**, 123516 (2015).
[69] C. Dvorkin, M. Wyman, D. H. Rudd, and W. Hu, Phys. Rev. D **90**, 083503 (2014).

[70] M. Wyman, D. H. Rudd, R. A. Vanderveld, and W. Hu, Phys. Rev. Lett. **112**, 051302 (2014).
[71] B. Leistedt, H. V. Peiris, and L. Verde, Phys. Rev. Lett. **113**, 041301 (2014).
[72] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **594**, A14 (2016).
[73] E. Di Valentino, A. Melchiorri, and J. Silk, Phys. Lett. B **761**, 242 (2016).
[74] E. Di Valentino, E. V. Linder, and A. Melchiorri, Phys. Rev. D **97**, 043528 (2018).
[75] E. Di Valentino, A. Melchiorri, and O. Mena, Phys. Rev. D **96**, 043503 (2017).
[76] V. Poulin, K. K. Boddy, S. Bird, and M. Kamionkowski, Phys. Rev. D **97**, 123504 (2018).
[77] N. Aghanim *et al.* (Planck Collaboration), arXiv:1807.06209.
[78] A. Ewall-Wice, T.-C. Chang, J. Lazio, O. Dore, M. Seiffert, and R. A. Monsalve, arXiv:1803.01815.
[79] J. Mirocha and S. R. Furlanetto, arXiv:1803.03272.
[80] M. Pospelov, J. Pradler, J. T. Ruderman, and A. Urbano, Phys. Rev. Lett. **121**, 031103 (2018).
[81] R. Barkana, Nature (London) **555**, 71 (2018).
[82] A. A. Costa, R. C. G. Landim, B. Wang, and E. Abdalla, Eur. Phys. J. C **78**, 746 (2018).
[83] S. Seager, D. D. Sasselov, and D. Scott, Astrophys. J. Suppl. Ser. **128**, 407 (2000).
[84] S. Seager, D. D. Sasselov, and D. Scott, Astrophys. J. **523**, L1 (1999).
[85] Y. Ali-Haimoud and C. M. Hirata, Phys. Rev. D **83**, 043513 (2011).
[86] L. A. Ureña-López and A. X. Gonzalez-Morales, J. Cosmol. Astropart. Phys. 07 (2016) 048.
[87] Y. Cui, M. Lewicki, D. E. Morrissey, and J. D. Wells, Phys. Rev. D **97**, 123505 (2018).
[88] W. Hu, Astroparticle Physics and Cosmology, in *Proceedings of Summer School, Trieste, Italy, 2002* (2004), http://publications.ictp.it/lns/vol14.html.