How sound are our ultralight axion approximations?

Jonathan Cookmeyer '17
Daniel Grin
Tristan L. Smith

Follow this and additional works at: https://scholarship.haverford.edu/physics_facpubs
How sound are our ultralight axion approximations?

Jonathan Cookmeyer1,2, Daniel Grin,1,* and Tristan L. Smith3

1Department of Physics and Astronomy, Haverford College, 370 Lancaster Avenue, Haverford, Pennsylvania 19041, USA
2Department of Physics, University of California, Berkeley, California 94720, USA
3Department of Physics and Astronomy, Swarthmore College, 500 College Avenue, Swarthmore, Pennsylvania 19081, USA

(Received 4 October 2019; published 2 January 2020)

Ultralight axions (ULAs) are a promising dark-matter candidate. ULAs may have implications for small-scale challenges to the ΛCDM model and arise in string scenarios. ULAs are already constrained by cosmic microwave background (CMB) experiments and large-scale structure surveys, and will be probed with much greater sensitivity by future efforts. It is challenging to compute observables in ULAs with sufficient speed and accuracy for cosmological data analysis because the ULA field oscillates rapidly. In past work, an effective fluid approximation has been used to make these computations feasible. Here this approximation is tested against an exact solution of the ULA equations, comparing the induced error of CMB observables with the sensitivity of current and future experiments. In the most constrained mass range for a ULA dark-matter component (10^{-27} eV ≤ m_{ax} ≤ 10^{-25} eV), the induced bias on the allowed ULA fraction of dark matter from Planck data is less than 1σ. In the cosmic-variance limit (including temperature and polarization data), the bias is ≤ 2σ for primary CMB anisotropies, with more severe biases (as high as ~3σ) resulting for less reliable versions of the effective fluid approximation. If all of the standard cosmological parameters are fixed by other measurements, the expected bias rises to 4 – 20σ (well beyond the validity of the Fisher approximation), though the required level of degeneracy breaking will not be achieved by any planned surveys.

DOI: 10.1103/PhysRevD.101.023501

I. INTRODUCTION

Measurements of cosmic microwave background (CMB) fluctuations [1–4], the clustering/gravitational lensing of galaxies [5–8], and the kinematics of cosmic acceleration (through Type Ia supernovae) [9,10] have ushered in the era of precision cosmology. Current data are consistent with the ΛCDM scenario, with cosmic density parameters of Ω_b h^2 = (2.22 ± 0.02) × 10^{-3} for baryons, Ω_c h^2 = 0.120 ± 0.002 for cold dark matter (CDM), and Ω_Λ = 0.685 ± 0.007 for dark energy.

The “cold” in CDM refers to the fact that observations of the cosmological large-scale structure require dark matter (DM) to be nonrelativistic when this structure forms. The standard model (SM) does not contain a DM candidate with this property and sufficiently weak couplings.

Many beyond the standard model candidates for DM have been proposed. The best motivated possibilities are weakly interacting massive particles (WIMPs) and axions [11,12]. WIMPs (e.g., neutralinos, gravitinos) arise in supersymmetric theories [11,12] as well as some other scenarios, while axions provide a solution to the strong CP (charge-parity) problem of quantum chromodynamics (QCD) [13–21]. For particle masses m_{ax} ≲ 10^{-2} eV (in units where c = 1), axions would be produced nonthermally through the oscillation of a scalar field, a distinct scenario from standard thermal production.

Direct detection experiments [22], indirect detection efforts [23,24], and Large Hadron Collider searches for evidence of supersymmetry [25] have all yielded increasingly stringent upper limits to WIMP properties [26]. The ample unexplored parameter space of QCD axion masses and couplings thus merits exploration, which is underway thanks to experimental efforts such as ADMX [27], IAXO [28], MADMAX [29], CASPer [30], and others [20], as well as astrophysical tests using stellar cooling and other effects [31,32]. Most of these efforts probe values m_{ax} ≥ 10^{-14} eV.

Scalar fields of even lower masses are an interesting possibility [33–36]. Expectations for their standard-model couplings are model dependent, and so gravitational observables are a useful complement to detection efforts. These fields are referred to fuzzy dark matter (FDM) [37], axion-like particles (ALPs), wave dark matter, or ultralight axions (ULAs). We use the latter nomenclature. ULAs could be astrophysically relevant on many scales, ranging from stellar-mass black holes to dwarf galaxies [38].

* dgrin@haverford.edu
ULAs would have unusual cosmological properties. ULAs maintain a constant energy density when $m_{\text{ax}}/\hbar < H$ (where $H$ is the cosmic expansion rate) [34,39], and then redshift with the cosmic scale factor $a$ as $\rho_{\text{ax}} \propto a^{-3}$ [39]. If $m_{\text{ax}} \lesssim 10^{-27}$ eV, ULAs would only begin to dilute after matter-radiation equality, making them unsuitable as a dark-matter candidate. In this case, ULAs would contribute to the (early or late-time, depending on the mass) dark-energy (DE) density of the universe until they begin to dilute [40,41]. If $m_{\text{ax}} \gtrsim 10^{-27}$ eV, ULAs begin to dilute before equality, allowing them to cluster as dark matter [33–37,41–43].

The existence of ULAs is predicted in string scenarios such as the “axiverse” [44], where they arise as Kaluza-Klein (KK) zero modes of antisymmetric tensors on compactified extra dimensions [45–47]. In the axiverse, there is a broad mass spectrum of many axions, which may be important during different epochs (see Ref. [43] and references therein). This motivates us to consider the observational consequences of a broad range of $m_{\text{ax}}$ values.

The large de Broglie wavelengths of the ULA scalar field cause ULAs to exhibit suppressed gravitational clustering on galactic scales [37,41,43,48–51]. For masses around $m_{\text{ax}} \sim 10^{-22}$ eV, a large DM fraction in ULAs could address small-scale challenges to the ΛCDM scenario, such as cores in some galaxy density profiles, the paucity of Milky Way (MW) satellites, and the “too big to fail” problem [43]. Other MW-scale dynamical probes also constrain ULAs [52–54].

For masses $10^{-23}$ eV $\lesssim m_{\text{ax}} \lesssim 10^{-21}$ eV, ULAs cause a suppression in neutral hydrogen (HI) density fluctuations at high redshift, suppressing the flux power spectrum of the Lyman-$\alpha$ (Lyα) forest in quasars. At lower masses still ($10^{-27}$ eV $\lesssim m_{\text{ax}} \lesssim 10^{-23}$ eV), ULAs must be subdominant, but could still have a density that is ~1%–5% of the DM density, comparable to the baryon and massive neutrino densities.

The dark sector may consist of numerous particle species (as the SM does), and in the axiverse scenario, the possibility of ULA dark matter that satisfies constraints in the range $m_{\text{ax}} \sim 10^{-22}$ eV usually coincides with the existence of ULAs with $m_{\text{ax}} \lesssim 10^{-23}$ eV [60,61]. In this lower mass range, limits of $\Omega_{\text{ax}} H^2 \lesssim 6 \times 10^{-3}$ have been obtained using observations of CMB primary temperature/polarization anisotropies [41]. Constraints of $\Omega_{\text{ax}} H^2 \lesssim 3 \times 10^{-3}$ have been obtained using CMB lensing potential reconstructions [62].

ULAs could be a spectator field during inflation and source isocurvature perturbations, which are constrained by the CMB. The existence of ULAs could yield a new probe of the inflationary energy scale [62–66], complementing constraints from B-mode polarization searches. All these probes (and future efforts) depend on reliable linear computations. Reliable simulations of nonlinear structure formation (relevant to MW-scale observations) also require reliable linear initial conditions [67–72].

Linear ULA perturbations are usually evolved using the effective fluid approximation (EFA). Stiff ordinary differential equations (ODEs) arise when ULAs are treated (exactly) as a classical scalar field (obeying the Klein-Gordon equation), due to disparate timescales ($m_{\text{ax}}/\hbar \gg H_0$). The EFA is obtained by cycle averaging out the fast timescale ($\sim m_{\text{ax}}^{-1}$) to yield fluid equations with a time-dependent equation of state $w(a)$ (which interpolates between $w = -1$ at early times and $w = 0$ when $m/\hbar \gg H$), and a scale-dependent sound speed $c_s^2$ [37,41,43,48–51,62,73–80].

This method is essentially the Wentzel-Kramers-Brillouin (WKB) approximation and is implemented in the axionCAMB [41,62] code used in a number of works. An alternative cycle-averaging formulation is proposed (and implemented) in Ref. [81], and further applied in Ref. [82]. We find that it is equivalent to the EFA (see Sec. II C), an important conclusion of our work.

The next generation of CMB experiments, e.g., CMB Stage-4 (CMB-S4) [83] could yield cosmic-variance limited measurements of CMB polarization out to $\ell < 5000$ [84], with significant signal to noise coming from lensing in the nonlinear regime [85]. Analysis of the Ly-$\alpha$ forest and nonlinear observables requires precise computations of the matter power spectrum in ULAs models.

It is timely to ask if the EFA is accurate enough: Can we trust observables predicted using the EFA? Is it sound to use a scale-dependent sound speed? How large is the bias induced by the EFA in ULAs abundance measurements? In this work, we solve the exact ODEs for ULA DM and compare results from the EFA in the range $m_{\text{ax}} \gtrsim 10^{-27}$ eV.

Mode evolution near the era of recombination is significantly altered for some of the CMB-scale $k$ values. At the most constrained ULA masses ($m_{\text{ax}} \sim 10^{-27}$ eV), we find that CMB anisotropy power spectra between the two computational approaches vary significantly, for sufficiently high ULA mass fraction $\Omega_{\text{ax}} = \Omega_{\text{ax}}/\Omega_{\text{DM}}$, where $\Omega_{\text{DM}} = \Omega_{\text{DM}} + \Omega_\text{c}$ is the total DM density parameter.

We examined three different EFA implementations, distinguished by the time at which exact equations are matched to EFA equations. For the fiducial switch at $m_{\text{ax}}/\hbar = 3H$, we evaluate the resulting bias in CMB-based determinations of $\Omega_{\text{ax}}/\Omega_{\text{DM}}$ for a cosmic-variance limited experiment (including cosmological parameter degeneracies). We find this bias to be $2\sigma$ for $m_{\text{ax}} = 3.2 \times 10^{-25}$ eV and $0.2\sigma$ for $m_{\text{ax}} = 10^{-24}$ eV, with intermediate results elsewhere in the mass range $10^{-27}$ eV $\lesssim m_{\text{ax}} \lesssim 10^{-24}$ eV. At Planck noise levels, this implementation exhibits bias $\lesssim 1\sigma$ (if parameter degeneracies are included), validating the CMB-only constraints of Ref. [41]. The other two EFA implementations have larger biases (with $m_{\text{ax}}$ dependence) by a factor of $\sim 2$. In the idealized case that external data perfectly break all CMB parameter degeneracies, larger biases of $4 - 20\sigma$ could result for the fiducial switch.

The organization of this paper is as follows. In Sec. II we review the dynamics of cosmological ULAs, introduce the
EFA, show the equivalence of an alternate cycle-averaging procedure to the EFA, and summarize the set of EFA implementations to be compared. In Sec. III we explain the details of our Boltzmann code. In Sec. IV we assess the impact of the EFA on perturbation evolution and CMB observables. In Sec. V, we estimate parameter bias induced by the use of the EFA. We conclude in Sec. VI. A short derivation of the EFA is given in Appendix A, while the numerical equivalence between the EFA and the alternate cycle-averaging approach is demonstrated in Appendix B.

The Z statistic (used to bound errors in CMB predictions) is discussed in Appendix C. Additional numerical results are discussed in Appendix D.

II. ULA PHYSICS

We briefly summarize the relevant axion physics. For a more in-depth review, see Refs. [41,43]. Depending on details of the production mechanism and cosmological bounds, the QCD axion introduced to solve the strong CP problem must have a mass $m_{\text{ax}} \gtrsim 10^{-12}$ eV [43]. ULAs arise as KK zero modes from the compactified extra dimensions predicted by string theory [44]. The masses of these particles can be extremely small ($m_{\text{ax}} \lesssim 10^{-18}$ eV), motivating the term “ultralight axions.”

A. ULA equations of motion

The equations of motion (EOMs) for ULAs are those of a scalar field in a perturbed Friedmann-Robertson-Walker metric, the homogeneous and perturbed Klein-Gordon (KG) equations. Using conformal time and synchronous gauge, these equations take the form [86] ($\hbar = c = 1$)

$$\phi_0'' = -2 \frac{a'}{a} \phi_0' - a^2 \tilde{m}^2 \phi_0,$$

$$\phi_1' = -2 \frac{a'}{a} \phi_1' - (k^2 + a^2 \tilde{m}^2) \phi_1 - \frac{H_0' \phi_0'}{2},$$

where the full field $\phi_{\hat{k}, t} = \phi_0(t) + \phi_1(\hat{k}, t)$ is expanded in terms of a background field $\phi_0(t)$ and a perturbation $\phi_1(\hat{k}, t)$. Here $a$ is the usual cosmological scale factor, $k$ is the comoving wave number of a mode, and $h_0$ is one of the degrees of freedom of the metric perturbation [86]. The derivatives in these equations are with respect to conformal time, and denoted with the $'$ symbol. The axion mass (usually given in eV) is converted to cosmological units of MeV by the conversion $\tilde{m} = m_{\text{ax}}/\hbar$, where $\hbar$ is given in eVs.

The stress-energy tensor components associated with the homogeneous field are

$$\rho_{\text{ax}} = \frac{1}{2} (\phi_0')^2 a^{-2} + \tilde{m}^2 \phi_0^2,$$

$$p_{\text{ax}} = \frac{1}{2} (\phi_0')^2 a^{-2} - \tilde{m}^2 \phi_0^2,$$

while those corresponding to perturbations are

$$\delta \rho_{\text{ax}} = \phi_0' \phi_1' a^2 - \tilde{m}^2 \phi_0 \phi_1,$$

$$\delta p_{\text{ax}} = \phi_0' \phi_1' a^2 - \tilde{m}^2 \phi_0 \phi_1,$$

$$u_{\text{ax}} = (1 + w_{\text{ax}}) v_{\text{ax}} = k \frac{\phi_0' \phi_1'}{\rho_{\text{ax}} a^2}.$$

The fluid variables are the homogeneous ULA density $\rho_{\text{ax}}$, pressure $p_{\text{ax}}$, density perturbation $\delta \rho_{\text{ax}}$, and pressure perturbation $\delta p_{\text{ax}}$. The equation of state parameter is $w_{\text{ax}} = p_{\text{ax}}/\rho_{\text{ax}}$ and $v_{\text{ax}}$ is the scalar associated with the ULA velocity. Additionally, we have the usual Friedmann equation

$$H^2 = \frac{8\pi G}{3} \sum \rho_i,$$

where the sum is over all relevant particle species. Computationally, it is often helpful to use the conformal Hubble parameter $H_0$, defined by $a' = aH_0$, and related to the standard Hubble parameter via $H_t = aH_0$

We have taken the potential to be harmonic $V = \frac{1}{2} m_{\text{ax}}^2 \phi^2$, a reasonable approximation near the minimum of the periodic instantan-generate $V \propto [1 - \cos (\phi/f_{\text{ax}})]$ potentials typical of ULAs. As noted in Refs. [41,62], most of the posterior parameter space consistent with CMB observations has $\phi \ll f_{\text{ax}}$, and so this is a sensible approximation. At higher masses, anharmonic corrections could have important implications for predictions at Lyman-\(\alpha\) forest scales or nonlinear structure formation [87]. We will explore anharmonic potentials in future work.

As Eqs. (1) and (2) are just the homogeneous and perturbed Klein-Gordon equations in an expanding universe, we expect oscillations with frequency $\tilde{m}$. Evolving these equations of motion up until the present day from the early universe to the present day is thus numerically expensive if $\tilde{m}/H \gg 1$, due to the stiff differential equations that result. Cosmological parameter constraints and tests of novel models typically require repeated calls to a Boltzmann code, as a likelihood surface is explored using Monte Carlo Markov chain techniques (MCMC) or related methods. For constraints to be obtained, the Boltzmann code should have an execution time of $\lesssim 10$ s. For ULA models, achieving this has required the use of the EFA, which essentially averages over the fast timescale of the oscillations.

B. Effective fluid approximation

In Appendix A, we briefly review a derivation of the cycle-averaged ULA background [39], which is

$$\rho_{\text{ax}} \approx \begin{cases} 
\rho_{\text{ax}}(a = 0) & \text{if } H/\tilde{m} \gg 1 \\
\rho_{\text{ax}}(a = a_{\text{osc}}) \left(\frac{\tilde{m}}{a}ight)^3 & \text{if } H/\tilde{m} \ll 1.
\end{cases}$$

023501-3
where \( H(a_{osc}) \approx \dot{m} \) (see Fig. 1). If \( H/\dot{m} \gg 1 \), then the field is overdamped and does not oscillate.

To state the EFA for perturbations, it is helpful to rewrite the ULA’s exact equations of motion in terms of synchronous gauge fluid variables to obtain [41]

\[
\rho'_{\text{ax}} = -3 \frac{d}{a} (1 + w_{\text{ax}}) \rho_{\text{ax}},
\]

\[
\delta'_{\text{ax}} = -k u_{\text{ax}} - \left(1 + w_{\text{ax}}\right) \frac{H}{2} - 3 \frac{d}{a} \left(c^2_{\text{ad}} - w_{\text{ax}}\right) \delta_{\text{ax}},
\]

\[
u'_{\text{ax}} = \frac{d}{a} (3w_{\text{ax}} - 1) u_{\text{ax}} + k c^2_{\text{ad}} \delta_{\text{ax}},
\]

\[
c^2_{\text{ad}} \delta_{\text{ax}} = \delta_{\text{ax}} + 3 \frac{d}{a} (1 - c^2_{\text{ad}}) \frac{u_{\text{ax}}}{k},
\]

with \( w_{\text{ax}} = p_{\text{ax}}/\rho_{\text{ax}}, c^2_{\text{ad}} = p'/\rho', \delta_{\text{ax}} = \delta p_{\text{ax}}/\rho_{\text{ax}}, \) and \( c^2_{s} = \delta p_{\text{ax}}/\rho_{\text{ax}} \). Note that these equations with fluid variables \( \langle \rho_{\text{ax}}, w_{\text{ax}}, \delta_{\text{ax}}, u_{\text{ax}} \rangle \) are exact and equivalent to Eqs. (1) and (2) with field variables \( \phi_0, \phi'_0, \phi_1, \phi'_1 \). This formulation is the generalized dark matter (GDM) formulation of Ref. [88].

The oscillations at frequency \( \dot{m} \) can be removed from Eq. (7) by using the fact that the ULA fluid equations are well described by an effective fluid with cycle-averaged \( \langle w_{\text{ax}} \rangle \) and \( \langle c^2_{s} \rangle \), which are not oscillating with the ULA field. We have already seen that when \( H/\dot{m} \ll 1, \rho'_{\text{ax}} \approx -3a' \rho_{\text{ax}}/a \) implying that the cycle-averaged \( \langle w_{\text{ax}} \rangle = 0 \).

Applying the ansatz \( \phi_1 = \phi_+ \cos(\dot{m}t) + \phi_- \sin(\dot{m}t) \), the authors of Ref. [48] find the cycle-averaged sound speed to be

\[
\langle c^2_{s} \rangle = \frac{k^2 / (4 \dot{m}^2 a^2)}{1 + k^2 / (4 \dot{m}^2 a^2)},
\]

where \( \langle \delta p_{\text{ax}} \rangle \) and \( \langle \delta \rho_{\text{ax}} \rangle \) are the cycle-averaged ULA pressure and density fluctuations in the ULA’s cycle-averaged rest frame.

Using more general assumptions about equipartition between kinetic and potential terms (and without assumptions about the functional form of scalar-field oscillations), the same expression for the cycle average of \( c^2_{s} \) is derived in Ref. [51], and generalized to potentials of the form \( V \propto (1 - \cos \phi/\phi_0)^n \) near their minimum, where \( V \propto \phi^{2n} \) and the field rapidly oscillates. Scalar fields described by such potentials are one way of resolving the tension between CMB data and more local measurements of \( H_0 \), by way of an era of early dark-energy (EDE) dominance [89–92].

For \( n \geq 2 \), the hierarchy of timescales between scalar-field oscillation and the Hubble parameter is far less extreme, and the perturbed Klein-Gordon equation may be solved exactly in the course of an MCMC simulation [51,90]. The relative impact of anharmonic terms and a fluid approximation on observables and constraints in EDE models is the subject of ongoing discussion [89–92].

Substituting \( \langle w_{\text{ax}} \rangle \) and \( \langle c^2_{s} \rangle \) into Eqs. (8) and (9), we arrive at a set of effective (cycle averaged) equations of motion for the ULA fluid variables that are valid when \( H/\dot{m} \ll 1 \). These define the EFA:

\[
\rho'_{\text{ax}} = -3 \frac{d}{a} \rho_{\text{ax}},
\]

\[
\delta'_{\text{ax}} = -k u_{\text{ax}} - \left(1 + w_{\text{ax}}\right) \frac{H}{2} - 3 \frac{d}{a} \langle c^2_{s} \rangle \delta_{\text{ax}},
\]

\[
u'_{\text{ax}} = \frac{d}{a} (3w_{\text{ax}} - 1) u_{\text{ax}} + k \langle c^2_{s} \rangle \delta_{\text{ax}},
\]

The term proportional to \( \langle c^2_{s} \rangle \) in Eq. (14) is the linear-theory expansion of the quantum pressure term discussed in the fuzzy dark matter and ULA literature.

The EFA is only valid once the field (or perturbation modes) begins to coherently oscillate, and so the exact field equations must be solved at early times, with a switch implemented from exact to cycle-averaged equations when \( \dot{m}/H = N \), where \( N \) is a constant that defines a specific EFA implementation. The ULA fluid variables \( \delta_{\text{ax}}, u_{\text{ax}}, \) and \( \rho_{\text{ax}} \) are matched to ensure continuity at the switch.

C. Alternative cycle-averaging procedure

In Ref. [81], a different set of variables is used for mode evolution. Dimensionless angular coordinates \( \theta \) and \( \dot{\theta} \) are defined in lieu of \( \phi_0 \) and \( \phi_1 \), along with their difference.
Coherent oscillation (at times for which \( \tilde{m} > H \)) is explicitly built into the formalism through terms such as \( \sin(\tilde{\theta}) \) and \( \sin(\theta) \). At early times, the resulting equations are equivalent to the exact field equations [Eqs. (1) and (2)], and thus to our early-time GDM equations. At late times, these equations are stiff and an exact integration is still numerically intractable for use in fast cosmological parameter-space exploration. Some cycle-averaging procedure is thus still needed there.

To that end, an alternative cycle-averaging procedure is introduced in Ref. \[81\]. Oscillatory functions \( f(x) \) (\( x = \theta, \bar{\theta}, \) or \( \tilde{\theta} \), depending on the equation) are replaced with

\[
f_\tilde{\theta}(x) = \frac{1}{2}(1 - \tanh(x^2 - x_s^2))f(x),
\]

for \( x \leq x_s \) (where \( x_s = 100 \)). When \( x > x_s \), \( f(x) \) is numerically set to zero. For example, the replacement

\[
\cos(x) \rightarrow \begin{cases} \frac{1}{2}(1 - \tanh(x^2 - x_s^2))\cos(x) & \text{if } x \leq x_s, \\ 0 & \text{if } x > x_s \end{cases}
\]

is made.

Put simply, a smoothed switch is used to separate the fast \((m^{-1})\) and slow \((H)\) timescales for ULA evolution. The chosen smoothing width is very narrow (\( \delta t / t \approx 5 \times 10^{-5} \)), and thus qualitatively is likely to still exhibit whatever undesirable transient behavior occurs in the EFA discussed above.

Since \( \theta \approx \tilde{\theta} \approx 2m_{ax}t \), at times \( t \gg 50/m_{ax} \), the equations evolved in Ref. \[81\] are effectively

\[
\frac{d\rho_{ax}}{d\ln a} = -3\rho_{ax},
\]

\[
\frac{d\tilde{\theta}}{d\ln a} = \frac{k^2}{k_f^2} + e^{-a} \frac{dh_L}{d\ln a} \cos \tilde{\theta} - \tilde{\theta},
\]

\[
\frac{da}{d\ln a} = \frac{1}{2} e^{-a} \frac{dh_L}{d\ln a} \sin \tilde{\theta} - \tilde{\theta},
\]

where [noting that \((1 + w_{ax})v_{ax} = ku_{ax}\)]

\[
\delta_{ax} = -e^{a} \sin \frac{\tilde{\theta}}{2}; \quad ku_{ax} = -\frac{k^2}{2am_{ax}} e^{a} \cos \frac{\tilde{\theta}}{2}.
\]

Note the natural appearance of the ULA Jeans wave number \( k_f \).

Differentiating Eq. (20) using Eqs. (18) and (19), we can represent the equations of motion for the variables \( \alpha \) and \( \tilde{\theta} \) in terms of more standard fluid variables, yielding

\[
\rho'_{ax} = -\frac{3}{a} \frac{da}{d\ln a} \rho_{ax},
\]

\[
\delta'_{ax} = -ku_{ax} - \frac{1}{2} \frac{dh'_L}{d\ln a},
\]

Equation (21) is just Eq (12) in the \( \langle \nu \rangle = 0 \) limit valid at late times. Equation (22) is Eq. (13) if the term \( \propto \langle c_s^2 \rangle \) on the right-hand-side of Eq. (13) is dropped. Finally, Eq. (23) is Eq. (14) in the nonrelativistic limit of Eq. (11) that \( \langle c_s^2 \rangle \approx k^2/(4m_{ax}^2) \). We verify in Appendix B that these differences are numerically irrelevant for CMB observables, and so the method of Ref. \[81\] is equivalent to the EFA.

### III. BOLTZMANN CODE

In order to investigate the effect of the EFA, we developed a PYTHON Boltzmann code to calculate individual mode growth, CMB anisotropy power spectra, and the
manner power spectrum in the presence of ULAs. We use our own Boltzmann code (as opposed to modifying existing codes such as CLASS [93] or CAMB [94]). We do this to clearly isolate the effect of different treatments of ULA perturbations (treated with sufficient time resolution) without requiring extensive modification to these more complete Boltzmann codes to resolve such timescales. Some preliminary comparisons were made in Refs. [41,51], but without assessing implications for parameter inference.

In our code, the other components of the universe (dark matter, baryons, radiation, neutrinos, dark energy) are evolved using the equations from Ref. [95], and we compute CMB power spectra using the line-of-sight method [96]. We keep seven photon modes and 12 neutrino modes, and we truncate the hierarchy of equations as in Ref. [95]. We use the initial conditions for a universe with ULAs as derived in Refs. [41,62] (using methods also applied in Refs. [97–99]).

We ported a public C++ implementation of the RECFAST code [100–103] into PYTHON, in order to compute cosmic recombination histories with reasonable precision. It is known that a variety of additional physical effects (beyond those included in the multilevel atom computation underlying RECFAST) affects recombination at the 0.1% level (e.g., higher-\(n\) two-photon transitions, resonant scattering in the Lyman-\(\alpha\), deviations from statistical equilibrium between angular momentum substates of high-\(n\) Rydberg states). These and other recombination effects are discussed in Refs. [104–115] (and references therein).

We implemented the numerical methods used in Ref. [116] and have neglected neutrino mass. Current CMB data are consistent with this choice [117] (though neutrino experiments indicate a nonzero neutrino mass whose absolute scale could be detected by future CMB experiments [84]), allowing us to avoid the complications of following perturbation evolution in a species that transitions from relativistic to nonrelativistic on cosmological timescales [93,98,118,119]. We neglect smoothing of primary CMB anisotropies by weak gravitational lensing. We include homogeneous reionization as in Ref. [120].

ULAs are added self-consistently as described in Sec. II A. We use the SCIPY [121] ODE solver with VODE [122] integrator using a BDF (stiff) method. Rather than stepping forward using derivatives computed at present-day function values, this implicit technique solves nonlinear equations for future function values, using expressions which approximate the relevant derivatives in terms of future and present function values. Implicit techniques are well suited to stiff differential equations—the system under study here has two vastly disparate timescales \([h/(m_{\text{ax}} c^2)]\) and \([c/H]\), making it well suited to the chosen numerical method. We set a large (10^9) maximum number of steps. The most important parameters are the relative and absolute error tolerance parameters, which are set to values of 10^{-14} and 10^{-9}, respectively.

When plotting spectra, we set all the parameters except the ULA mass \(m_{\text{ax}}\) and ULA dark matter fraction \(r_{\text{ax}} = \Omega_{\text{ax}}/\Omega_{\text{DM}}\) to their best fit values from the 2015 Planck data release [123]. In Sec. V, we explore variations in all cosmological parameters when assessing the impact of the EFA on ULA parameter inferences.

Note that cold dark matter makes up a fraction of the energy density \(\Omega_{\text{CDM}} = \Omega_{\text{DM}} - \Omega_{\text{ax}}\). We consider only ULA masses \(m_{\text{ax}} \gtrsim 10^{-77} \text{ eV}\). This mass restriction guarantees that ULAs begin to dilute as matter (\(\rho_{\text{ax}} \propto a^{-3}\)) before matter-radiation equality (“equality” henceforth) [41], keeping us in the “DM-like” part of ULA parameter space. We will explore the impact of the EFA in the “DE-like” part of the mass range \((m_{\text{ax}} \lesssim 10^{-27} \text{ eV})\) in future work. As initial conditions for the homogeneous axion field \(\phi\), we use \(\phi_0 = 0\) and \(\phi_0 = C\) [43].

We use a root-finding interface with our homogeneous scalar-field evolution module to find the value of \(\phi\) which reproduces the desired value of \(\Omega_{\text{ax}}\). The initial guess is chosen to be \(\phi_0 = \sqrt{\Omega_{\text{ax}} \rho_c}/(m_{\text{ax}} a s^3)\), where \(s_i\) is defined as the scale factor when \(H/m_{\text{ax}} = 3\), with \(H\) computed using \(\Lambda CDM\) to obtain an initial guess but self-consistently including ULAs henceforth. The algorithm assumes the existence of a root in the range \(\phi_0/5 \rightarrow 5\phi_0\), which we have confirmed is sufficient for all values of the parameters explored in this work. For the perturbations, adiabatic initial conditions give \(\phi_1 = 0\) [41,43], and, because of large Hubble drag initially, \(\phi_1 = 0\) [51].

Our CMB and matter power spectrum in the absence of ULAs were compared against CAMB [94] with quantitative agreement at the 5% level, except for \(\ell < 10\), where agreement is at the \(\sim 10\%\) level for the EE angular power spectrum \(C_{\ell}^{EE}\) and 5% for the TT power 1 spectrum \(C_{\ell}^{TT}\). Although there is a difference between the code used in this work and CAMB, it is unimportant because our goal here is a comparison between the “exact” treatment and the EFA with all other assumptions and computational tools held fixed for self-consistency. Our use of a simplified, relatively accurate recombination history (and simplifying approximations for neutrinos, as described above) should be sufficient for this goal.

\section{IV. MODE GROWTH AND CMB}

We now explore the impact of the various EFA implementations on cosmological observables. Naively, one might expect that later switches (by following the exact

\footnote{While this work was nearing completion, the Planck 2018 results were released [4]. Because of the very small shifts in central values of parameters between the 2015 and 2018 [4] data releases, and very weak dependence of bias results on fiducial parameter values over the allowed range, our conclusions should be unaffected by the difference between the two iterations of Planck results.}
EOMs for longer) would always perform better, but this is not the case, as we explain below.

**A. Mode evolution**

We first study the effect of the EFA on the growth of individual modes. To understand general features, we show individual mode growth for different values of $k$ in Fig. 2. We show the evolution of $\delta_\eta$ given our obvious interest in its dynamics and of $\delta$, as it is the fluid variable most important for observed properties of the CMB. A useful reference scale is given by the comoving ULA Jeans wave number

$$k_J(a) = a \sqrt{m_H}.$$  (24)

Modes with $k > k_J$ have suppressed growth relative to $\Lambda$CDM, as the perturbation lies within the ULAs wavelength [74]. We expect that $k_J(a = a_{re})$ (where $a_{re}$ is the scale factor at recombination, with $a_{re}$ being the scale factor at recombination. For $k \ll k_J$, the approximation does very well for ULA mode evolution ($\delta_a$), but as $k_J$ is approached and exceeded, the difference between the two cases at recombination as well as asymptotically becomes significant and then grows with $k$. The y-axis scale is arbitrary.
at recombination) will be the important scale for computations of CMB anisotropies. We find a \( r_c \) using the peak of the visibility function \( g = \kappa_0 e^{\kappa} \) with \( \kappa = a n_e \sigma_T \) with \( n_e \) being the density of electrons and \( \sigma_T \) the Thomson cross section.

For \( k \ll k_J(a = a_{re}) \), the approximation and the “exact” treatment agree asymptotically, so there is only a small change in the photon overdensity, \( \delta_T \). Once \( k \gg k_J \), there are noticeable differences between the two approaches, as seen both in the ULA overdensity \( \delta_{ax} \) and in \( \delta_T \), which could have an observable impact on CMB anisotropy measurements. Although we show mode evolution for \( m_{ax} = 10^{-27} \) eV, the qualitative features hold for other values.

We compare all three EFA implementations with the “exact” calculation in Fig. 3, for the value \( m_{ax} = 10^{-24} \) eV. We see that the \( N = 100 \) implementation captures early-time homogeneous density evolution more accurately than the \( N = 3 \) implementation, as well as early-time perturbation evolution (at least for the mode shown). This is unsurprising given that this implementation follows the exact scalar-field EOMs for a much larger number of oscillation cycles. The \( N = 3 \) implementation captures the later-time evolution of both ULA and photon variables more accurately, as there is more time for numerical transients (introduced by a discontinuous swap in the second derivatives of various fluid quantities) to dissipate. Given the differences between perturbation evolution in different implementations, a more detailed comparison of CMB power spectra is necessary.

B. CMB anisotropies

We now compute CMB anisotropies for our full grid of models using the standard line-of-sight formalism as implemented in Sec. III. The ordering of the accuracy of different EFA implementations depends on the mass. The WKB approximation underlying the EFA is itself an expansion in the small (time-dependent) parameter \( \epsilon_{WKB} \equiv H/m_{ax} \sim 1/N, \) which might be taken to imply that the \( N = 100 \) implementation should always outperform the \( N = 3 \) implementation, but this is not the case. Why?

All of these implementations excite spurious transients near the time of transition from the exact KG to EFA equations. The higher-\( N \) cases also have later switches, leaving less time for transient behavior to dissipate before recombination. There is thus a delicate balance at play (between suppression of transients and formal convergence of the WKB approximation) in determining which version of the EFA is best. This balance depends on when the switch occurs compared to the evolution of the observed
CMB fluctuation modes, and is thus sensitive to both $N$ and $m_{\text{ax}}$.

For the $N = 3$ and $N = 100$ implementations, we show the primary TT, EE, and TE ($C_{\ell}^{TT}$) power spectra for $m_{\text{ax}} = 3.16 \times 10^{-27}$ eV and $r_{\text{ax}} = \Omega_{\text{ax}} / \Omega_{\text{DM}} = 1$ in Fig. 4. This value of $r_{\text{ax}}$ is not allowed by the constraints [41,124], but is chosen to visually accentuate differences between approaches. We also show relative errors between different EFA implementations. In addition to the individual $C_{\ell}$, we plot fiducial model points with error bars corresponding to cosmic variance. Following Ref. [123], the modes are binned into bins of sizes $\{1, 2, 5, 30\}$ for $\ell'$ in the range $\{2, 4, 5, 10, 11, 30, 31, 4000\}$, respectively. The value plotted is the weighted average of $C_{\ell}$ with cosmic variance error at the weighted average $\ell'$ with weight proportional to $\ell' (\ell' + 1)$, as in the Planck Collaboration pipeline [125]. For a variety of other $m_{\text{ax}}$ values, we explore the relative accuracy of different EFA implementations in Appendix D using Fig. 16.

For $m_{\text{ax}} = 3.16 \times 10^{-27}$ eV, the $N = 3$ implementation is less accurate than the $N = 100$ implementation for TT and EE spectra, and comparable for TE. For $m_{\text{ax}} = 3.16 \times 10^{-26}$ eV, the $N = 3$ and $N = 100$ implementations are of comparable accuracy (compared with the exact KG equation) at most $\ell'$ values. For $m_{\text{ax}} = 3.16 \times 10^{-25}$ eV, the $N = 3$ implementation is more accurate than the $N = 100$ implementation.

Similar comparisons can be made for the $N = 1.6$ implementation, whose performance relative to other choices is a function of $m_{\text{ax}}$. At higher $m_{\text{ax}}$ values, the residuals are in between those of the $N = 3$ and $N = 100$ implementations, while at lower $m_{\text{ax}}$ values, the residuals are worse than any of the other implementations. In the interests of simplifying the discussion, we have omitted these results from figures showing power-spectra comparisons but include it in our final computations of bias in ULA density constraints.

Depending on the implementation and $\ell'$ value, the fractional error $\Delta C_{\ell} / C_{\ell}$ can be several orders of magnitude in excess of the approximate Planck noise level ($\Delta C_{\ell} / C_{\ell} \sim 10^{-3}$) when $r_{\text{ax}} = 1$. Forthcoming experiments (e.g., CMB-S4 [83]) will measure nearly $N_{\text{modes}} \sim 10^7$ perturbation modes. Roughly speaking [127], the fractional effect of any systematic error in the computation of CMB fluctuations...
anisotropies grows as $\sqrt{N_{\text{modes}}} \Delta C_{\ell}/C_{\ell}$, where $N_{\text{modes}} \propto \ell^2$ is the total number of multipole values with $\ell' \leq \ell$. Small biases thus require that $\Delta C_{\ell}/C_{\ell} \lesssim 1/\ell$, with detailed numerical considerations yielding the useful rule of thumb that bias-free parameter inference requires $\Delta C_{\ell}/C_{\ell} \lesssim 3/\ell$ [126–128]. We see in Figs. 4 and 16 that this condition is violated by several orders of magnitude at high $\ell \gtrsim 10^2$ when $\Delta \alpha \approx 1$.

As $\Delta C_{\ell}/C_{\ell} \propto \Delta \alpha$ and $\Delta \alpha \sim O(0.1)$, it stands to reason that the fractional error between all three EFA implementations and the “exact” calculation exceeds the needed accuracy for unbiased parameter inference, depending on the detailed full structure of the likelihood function (e.g., parameter degeneracies). To determine if the errors induced by the EFA significantly affect cosmological constraints to (or measurements of) ULA DM, we must compare the scale-dependent error with the information content of the CMB, as represented by the Fisher matrix [129–131]. We now estimate the EFA-induced parameter bias.

V. Z STATISTIC AND BIAS

We wish to estimate the systematic error in CMB measurements of (or limits on) the axion relic density $\Omega_{\alpha x}$ from the CMB that results from the use of the EFA. Although a full analysis would require mock datasets and MCMC analysis, reasonable estimates may be obtained using standard Fisher-analysis techniques [129–131].

The elements of the Fisher matrix for the CMB anisotropy power spectrum are given by [127,129,132–133]

$$F_{ij} = \sum_{A,A' \in \{TT,EE,TE\}} \sum_{\ell} f_{\text{sky}} \frac{\partial C_{\ell}^A}{\partial \lambda_i} \frac{\partial C_{\ell}^{A'}}{\partial \lambda_j} \left( \mathcal{H}^{-1}_{\ell A A'} \right),$$

$$\mathcal{H}_{\ell A A'} = \left\langle \left( \hat{C}_{\ell}^A - \bar{C}_{\ell}^A \right) \left( \hat{C}_{\ell}^{A'} - \bar{C}_{\ell}^{A'} \right) \right\rangle,$$

where $\lambda = (h, \Omega_b h^2, \Omega_{DM} h^2, z_{re}, n_s, \Omega_{\alpha x})$ is a choice of $\Lambda$CDM parameters, along with the ULA density of $\Omega_{\alpha x}$ today, and $f_{\text{sky}}$ is the sky fraction covered by the CMB experiment of interest. Here the data covariance matrix is $\mathcal{H}_{\ell A A'}$. The brackets $\langle \rangle$ denote an ensemble average, and $\bar{C}_{\ell}^A = \sum_{\ell'=\ell}^{\ell} a_{\ell'}^A a_{\ell'}^A/(2\ell+1)$ is the usual optimal estimator of angular power spectra using the multipole moments as data, where the observable $A = \{X, Y\}$ consists of the pair $X, Y \in \{T, E\}$. We neglect $B$-mode (curl) anisotropies here, as our analysis neglects weak gravitational lensing of the CMB and primordial tensor modes.

The ULA mass $m_{\alpha x}$ is varied and the full Fisher matrix $\mathcal{F}$ with elements $F_{ij}$ is recomputed at each value, to see how biases and parameter errors depend on ULA mass. The derivative $\partial C_{\ell}^A/\partial \lambda_i$ quantifies the response of the observables to the $\Lambda$CDM and ULA parameters of interest; typically this derivative must be obtained numerically using a Boltzmann code, though for $A_s$, the derivative may be obtained analytically ($\partial C_{\ell}^A/\partial A_s = C_{\ell}^A/A_s$). If we compute the Fisher matrix for a theoretical scenario and some experiment, we may forecast the error on the best fit parameters via $\sigma_{\lambda} = \sqrt{(\mathcal{F}^{-1})_{\lambda \lambda}}$, assuming that the parameter likelihood (given the data) distribution is Gaussian.

The elements of the data covariance matrix are [132,133]

$$\mathcal{H}_{\ell A A'} = \frac{2}{2\ell+1} \left( C_{\ell}^A + N_{\ell A}^A \right)^2; \quad A \in \{TT, EE\},$$

$$\mathcal{H}_{\ell A A'} = \frac{1}{2\ell+1} \left( C_{\ell}^{TE} \right)^2,$$

$$\mathcal{H}_{\ell A A'} = \frac{1}{2\ell+1} \left[ (C_{\ell}^{TT})^2 + (C_{\ell}^{TT} + N_{\ell TT}^T)(C_{\ell}^{EE} + N_{\ell EE}^T) \right],$$

$$\mathcal{H}_{\ell A A'} = \frac{2}{2\ell+1} C_{\ell}^{TE} (C_{\ell}^{TT} + N_{\ell TT}^T),$$

$$\mathcal{H}_{\ell A A'} = \frac{2}{2\ell+1} C_{\ell}^{TE} (C_{\ell}^{EE} + N_{\ell EE}^T),$$

$$N_{\ell A}^A = \delta_{AA} \sigma_{\ell}^2, \quad \sigma_{\ell}^2 = \frac{\delta_{\ell}^2 (\ell+1) \theta_{\text{FWHM}}^2}{16 \pi^2},$$

where $N_{\ell A}^A$ is the noise power spectrum for the observable $AA^\prime$. The remaining elements follow trivially since $\mathcal{H}_{\ell AA} = \mathcal{H}_{\ell A A'}$. The above covariance matrix takes into account both the cosmic variance, and the noise of the detector [132]. We use the usual approximations of Ref. [133], with an overall amplitude noise amplitude $\frac{\delta_{\ell}^2}{\ell}$ in $\mu K^2$. The instrument beam is assumed to be Gaussian with a full-width half-max angular size of $\theta_{\text{FWHM}}^2$ in radians.

The errors $\Delta C_{\ell}^{XY}$ in the theoretical calculation of $C_{\ell}^{XY}$ drive the peak of the likelihood to a different set of parameter values than the true model parameters, resulting in parameter biases. Under a Gaussian likelihood approximation, we follow Refs. [127,128,134–137] to compute this bias, which is [137]

$$\delta_i = -\frac{1}{\ell} \sum_j (\mathcal{F}^{-1})_{ij} V_j,$$

where

$$V_j = \sum_{A,A' \in \{TT,EE,TE\}} \sum_{\ell} f_{\text{sky}} \Delta C_{\ell}^A \frac{\partial C_{\ell}^{A'}}{\partial \lambda_i} \left( \mathcal{H}^{-1}_{\ell A A'} \right),$$

and $\Delta C_{\ell}^A = C_{\ell}^{A,\text{cut}} - C_{\ell}^{A,\text{exact}}$ is the shift in computed theoretical power spectra between the exact treatment of ULAs and the EFA. This expression takes into full account degeneracies between different cosmological parameters. If the bias in $\Omega_{\alpha x}$ resulting from the use of the EFA can be absorbed by adjusting other cosmological parameters, Eq. (28) will indicate a negligible shift.

---

3We follow Refs. [127,128,134–136] and drop higher-order terms of order $\Delta C_{\ell}$ in the Fisher matrix itself, as these lead to corrections of the form $O(C_{\ell}^2)$ to the bias. These terms are subdominant unless there is a ULA detection of high significance.
In the future, if other measurements (e.g., large-scale structure surveys or the CMB weak lensing trispectrum) are combined with CMB power spectrum measurements (dominated by primary anisotropies) to break parameter degeneracies, a larger bias may result. The bias and error in $\Omega_{ax}$ with all other cosmological parameters held fixed are given by

$$\frac{\delta \Omega_{ax}}{\sigma_{\Omega_{ax}}} = -\sigma_{\Omega_{ax}} \left[ \sum_{A,A' \in \{TT,TE,EE\}} \sum_{\ell} f_{\text{sky}} \Delta C^A_{\ell} \frac{\partial C^{A'}_{\ell}}{\partial \Omega_{ax}} (\Xi^{-1})_{AA'} \right].$$  \tag{29}$$

$$\sigma_{\Omega_{ax}} = \sqrt{\frac{1}{F_{\Omega_{ax} \Omega_{ax}}}}. \tag{30}$$

Before computing an extensive set of numerical derivatives, it would be useful to establish the maximum bias on a single parameter, normalized to its Fisher-level error. Using Eq. (29) and the Cauchy-Schwarz inequality (see Appendix C for a derivation), it can be shown that

$$\frac{\delta \Omega_{ax}}{\sigma_{\Omega_{ax}}} \leq Z = \sqrt{\sum_{A,A' \in \{TT,TE,EE\}} \sum_{\ell = \ell_{\text{min}}}^{\ell_{\text{max}}} f_{\text{sky}} \Delta C^A_{\ell} \Delta C^{A'}_{\ell} (\Xi^{-1})_{AA'}.} \tag{31}$$

This $Z$ statistic [138] has been used to estimate the impact of improvements to cosmic recombination history computations on CMB parameter estimation [106,108,110]. If $Z \ll 1$, we can safely conclude that the systematic errors in $C^A_{\ell}$ induced by use of the EFA make a negligible impact on the estimation of both $\Omega_{ax}$ and standard cosmological parameters. Conversely, if $Z \gg 1$, there could be large shifts in central values, and a bias estimate computed using Eq. (28) or (29) is needed. \footnote{Z is just the square root of the predicted change in the $\chi^2$ statistic induced by the use of the EFA. The interpretation of $Z$ as the maximum fractional bias in a single parameter (neglecting degeneracies) requires the added assumption of a Gaussian likelihood for the parameters.}

### A. Z statistic for current and future CMB experiments

We compute the $Z$ statistic using Eqs. (26) and (31), assuming noise properties for various CMB instruments (past, present, anticipated) given in Table II. Of course, even an ideal zero-noise experiment is affected by cosmic variance and is shown as the cosmic-variance limited (CVL) case. In this case, $Z$ quantifies the maximum impact of EFA-induced errors on CMB constraints to $\Omega_{ax}$.

We create a $7 \times 7$ logarithmically spaced grid with $10^{-27} \text{ eV} \leq m_{ax} \leq 10^{-24} \text{ eV}$ and $10^{-2} \leq \Omega_{ax}/\Omega_{DM} \leq 1$, and evaluate the $Z$ statistic for the three different versions of the EFA. For temperature, we restrict ourselves to $\ell \leq \ell_{\text{max}} = 2200$, as secondary temperature anisotropies dominate the primordial CMB at smaller scales (although there are futuristic proposals to go well beyond this limit, e.g., Refs. [85,145]). Since polarization foregrounds are expected to be less severe (than ones in the temperature) at small angular scales [84], we assume $\ell \leq 4000$ for polarization data and use the similar (polarization-only) expression

$$Z_{>\ell_{\text{max}}} = \sqrt{\sum_{\ell = \ell_{\text{max}}}^{\ell_{\text{max}}} f_{\text{sky}} \Delta C^E_{\ell} \Delta C^E_{\ell} (\Xi_{EE,EE}^{-1})_{\ell} \tag{32}$$

with $Z_{\text{tot}} = Z(\ell_{\text{max}} = \ell_{\text{max}}) + Z_{>\ell_{\text{max}}}$. As an example, in Fig. 5, we show the $Z$ statistic as a function of $\Omega_{ax}/\Omega_{DM}$ for $m_{ax} = 10^{-27} \text{ eV}$, where the constraints of Ref. [41] are most stringent, and where the effects of the EFA are most severe. If the true ULA abundance saturates current constraints (from the CMB power spectra without lensing, as in Ref. [41]), we see that (depending on the experiment), the $Z$ bound indicates potential biases as large as $\sim 4 \rightarrow 30 \sigma$. The effect is significantly less pronounced if only temperature data are used, highlighting the importance of CMB polarization. If the actual ULA density is far lower than present-day constraints, $Z$ could be much smaller, allowing the EFA-induced error to be neglected.

To succinctly capture the potential impact of the EFA (for the $N = 3$ implementation) at all masses, we plot $Z$ for $\Omega_{ax}/\Omega_{DM}$ values saturating the constraints of Ref. [41] in Fig. 6, including temperature and polarization anisotropies. We see that for $m_{ax} \lesssim 3.2 \times 10^{-26} \text{ eV}$, $-3 \sigma$ and greater biases could occur for all three experimental scenarios considered there. For CMB-S4 or the CVL case, significant ($\gtrsim 1 \sigma$) $Z$ values occur for values as high as $m_{ax} = 3.2 \times 10^{-25} \text{ eV}$. We also compute $Z$ for the $N = 1.6$
and \( \mathcal{N} = 100 \) implementations. These results are of the same order of magnitude and are thus omitted for brevity.

To check that the constraints of Ref. \[41\] are robust and to assess the impact of the EFA on future CMB tests of ULA physics, we must thus evaluate the more complete Fisher-matrix based estimates [Eqs. (28) and (29)]. These are needed to properly include the detailed response of \( C_\ell \) to \( \Omega_{\text{ax}} \) variations in all implementations considered, and to properly assess the impact of parameter degeneracies.

### B. Bias

Since the \( Z \) statistic was \( O(1) \) for a significant part of parameter space, we directly compute the bias in the value of \( \Omega_{\text{ax}} \) induced by the use of the EFA. We compute the parameter bias for our full \( 7 \times 7 \) grid of models using Eq. (28) and forecast the error on the best fit parameters via \( \sigma_i = \sqrt{\left( F^{-1}\right)_{ii}} \). As noted in Sec. VA, for \( \ell' > 2200 \) we expect temperature measurements to be dominated by secondary anisotropies, and we restrict all Fisher sums to EE-only portions of the data covariance matrix.

We calculate the relevant numerical derivatives with respect to \( \Lambda CDM \) parameters by modifying the Planck values at the percent level as in Ref. \[131\] and using the two-point symmetric finite difference method. We also compute derivatives with respect to \( \Omega_{\text{ax}} \). These bias/error estimates are thus a first-order approximation, given the use of \( \Omega_{\text{ax}} = 0 \) for derivatives with respect to \( \Lambda CDM \) parameters.

For the derivative with respect to \( \Omega_{\text{ax}} \neq 0 \), we check convergence by comparing the two-point symmetric, left, and right finite-difference methods and verifying that the bias is converged at the \( \sim 5\% - 10\% \) level for \( m_{\text{ax}} < 10^{-24} \) eV, where ULA sensitivity is very poor and the overall bias \( \delta_{\Omega_{\text{ax}}} \ll \sigma_{\Omega_{\text{ax}}} \). Below, we discuss how large the bias would be for \( \Omega_{\text{ax}} \) values saturating current constraints, but note that the bias will be less severe for \( \Omega_{\text{ax}} \) values well below the current upper limits. We confirm for the null (no ULA) hypothesis that Planck parameter error levels are reproduced. For the \( \mathcal{N} = 3 \) implementation, we also confirm that if ULAs are included that the marginalized \( \sigma_{\Omega_{\text{ax}}} \) curve (as a function of \( m_{\text{ax}} \)) is consistent (at order-of-magnitude level) with the constraints of Ref. \[41\], which were obtained using the same EFA implementation.

We begin by focusing our attention on the fiducial \( \mathcal{N} = 3 \) case. As an example, in Fig. 7, we show the bias for \( m_{\text{ax}} = 3.16 \times 10^{-27} \) eV using hypothetical temperature and polarization data, as a function of \( \Omega_{\text{ax}} / \Omega_{\text{DM}} \). We see that for this \( m_{\text{ax}} \) value, the bias is negligible for \( \Omega_{\text{ax}} / \Omega_{\text{DM}} \) values satisfying the constraints of Ref. \[41\]. As beam, noise, and \( \Omega_{\text{ax}} / \Omega_{\text{DM}} \) are varied, different scales are emphasized in the calculation of bias, and so it is natural that there is some nonmonotonicity with respect to experimental ordering and \( \Omega_{\text{ax}} / \Omega_{\text{DM}} \). Additionally, the bias can be positive or negative depending on the subtle interplay of some of these parameters.

To better capture the information content of all these bias figures, we tabulate the dimensionless bias as a function of \( m_{\text{ax}} \), with \( \Omega_{\text{ax}} / \Omega_{\text{DM}} \) fixed at the current 3\( \sigma \) constraint level
FIG. 7. The points show the dimensionless bias in \( \Omega_{\text{ax}} \) for the \( \mathcal{N} = 3 \) implementation, calculated using Eq. (28) (thus including parameter degeneracies), plotted as a function of the ULA dark matter fraction, \( \Omega_{\text{ax}}/\Omega_{\text{DM}} \). The lines show a simple point-to-point linear interpolation on a log-log plot. The vertical black line shows the 3\( \sigma \) upper limit to \( \Omega_{\text{ax}}/\Omega_{\text{DM}} \) from Ref. [41] for the \( m_{\text{ax}} \) values shown. Color code as in Fig. 5. Dashed lines indicate a positive bias, while solid lines indicate a negative bias.

FIG. 8. Dimensionless bias in \( \Omega_{\text{ax}} \) as a function of \( m_{\text{ax}} \) for the \( \mathcal{N} = 3 \) implementation, calculated using Eq. (28) (thus including parameter degeneracies), with \( \Omega_{\text{ax}}/\Omega_{\text{DM}} \) set equal to the current (Planck) 3\( \sigma \) upper limits.
regime, causing the amplitude of the fractional bias to retain its sign but increase in amplitude.

In Fig. 10, we show the dimensionless bias (now neglecting degeneracies), as a function of $m_{\text{ax}}$, with $\Omega_{\text{ax}}/\Omega_{\text{DM}}$ fixed exactly at the current $3\sigma$ constraint level. As in some of the cases above, the sharp dip in the CVL bias at $m_{\text{ax}} = 3.16 \times 10^{-25}$ eV (from which the bias curve returns to a more standard ordering) is driven by physical sign changes in the summand of Eq. (29). We see that at some of the most constrained masses ($3.16 \times 10^{-27}$ eV $\leq m_{\text{ax}} \leq 3.16 \times 10^{-26}$ eV), large biases ($|\delta \Omega_{\text{ax}}| > 2\sigma_{\Omega_{\text{ax}}}$) in ULA densities occur at Planck noise levels.

For more futuristic noise levels (as shown by the CMB-S4 and CVL cases), similar and even larger biases result if $m_{\text{ax}} \leq 10^{-25}$ eV. We note Fisher-matrix calculations that indicate biases greater than several $\sigma_{\Omega_{\text{ax}}}$ should just be taken as an indication that the actual bias is severe and not a precise result, given the breakdown of the Gaussian likelihood approximation for large deviations from central values.

We see that biases are large when parameters are fixed, but not when they are marginalized over; this implies that the systematic errors in $C_{\ell}$ values induced by the EFA are significantly degenerate with shifts in other cosmological parameters. It is natural to wonder then, if future more precise measurements of standard cosmological parameters will push bias toward the larger values shown in Fig. 10.

We then used Eq. (27) to compute the bias to $\Omega_{\text{ax}}$ resulting from the use of the EFA in the presence of priors and found in all cases that the results reproduced those of Fig. 8 at the $\sim 10\%-20\%$ level and were not comparable to the large bias in Fig. 10. In other words, the primary CMB Fisher matrix exhibits sufficient parameter degeneracy for the impact of the EFA on ULA parameters to be absorbed by variations in other parameters, even in the presence of sensible priors from upcoming experiments. Future work to extend this analysis should include CMB lensing and the full off-diagonal Fisher matrices from galaxy surveys (and other efforts), including the impact of ULAs on galaxy and cosmic shear clustering power spectra.

We also evaluated bias with no priors for the $N = 1.6$ and $N = 100$ implementations, and we show the results in Figs. 11–14. At Planck noise levels for the $N = 1.6$ implementation, $|\delta \Omega_{\text{ax}}| \lesssim \sigma_{\Omega_{\text{ax}}}$ for all $m_{\text{ax}}$ values if we marginalize over cosmological parameters (see Fig. 11).

We explored this issue further using forecasts for matter power spectrum measurements by the Large Synoptic Survey Telescope (LSST) [146], Dark Energy Spectroscopic Instrument (DESI) [147], Euclid satellite [147,148], and Square Kilometer Area (SKA) [149], as well as more direct distance ladder-based/gravitational-wave inferred measurements of the Hubble constant $H_0$ [150]. We imposed priors to cosmological parameters via a diagonal modification to the Fisher matrix, $F_{ij} \rightarrow F_{ij} + \delta_{ij}/\sigma_i^2$, where $\sigma_i$ is the forecasted error to a ΛCDM parameter from one of these efforts (priors were applied to $A_s$, $n_s$, $H_0$, and $\Omega_c$ as appropriate given Refs. [146–150]).
For the CVL case here, $2\sigma_{\Omega_{\text{ax}}} \lesssim |\delta_{\Omega_{\text{ax}}}| \lesssim 4\sigma_{\Omega_{\text{ax}}}$. If cosmological parameters are fixed, the bias in the CVL case is large: $3\sigma_{\Omega_{\text{ax}}} \lesssim |\delta_{\Omega_{\text{ax}}}| \lesssim 20\sigma_{\Omega_{\text{ax}}}$ (see Fig. 12).

For the $N = 100$ implementation at Planck noise levels including cosmological parameter marginalization, $|\delta_{\Omega_{\text{ax}}}| \lesssim \sigma_{\Omega_{\text{ax}}}$, while $|\delta_{\Omega_{\text{ax}}}| \lesssim 2\sigma_{\Omega_{\text{ax}}}$ if cosmological parameters are assumed fixed, at the same noise level (see Fig. 13). For the CVL case with cosmological parameter marginalization $\sigma_{\Omega_{\text{ax}}} \lesssim |\delta_{\Omega_{\text{ax}}}| \lesssim 4\sigma_{\Omega_{\text{ax}}}$ if $m_{\text{ax}} \lesssim 10^{-25}$ eV and $|\delta_{\Omega_{\text{ax}}}| \ll \sigma_{\Omega_{\text{ax}}}$ otherwise. For the $N = 100$ implementation with cosmological parameters held fixed in the CVL case, $\sigma_{\Omega_{\text{ax}}} \lesssim |\delta_{\Omega_{\text{ax}}}| \lesssim 7\sigma_{\Omega_{\text{ax}}}$ if $m_{\text{ax}} \gtrsim 10^{-26}$ eV and $|\delta_{\Omega_{\text{ax}}}| \ll \sigma_{\Omega_{\text{ax}}}$ otherwise.

We note that the $N = 100$ implementation exhibits an unusual property in the nonmarginalized bias computation [see Eq. (29) and Fig. 14]. Unlike the $N = 3$ and $N = 1.6$ implementations (see Figs. 10 and 12), the low $m_{\text{ax}}$ single-parameter bias computed in the $N = 100$ implementation is far smaller than the bound implied by the order of magnitude of $Z$ (see Fig. 6), i.e., $|\delta_{\Omega_{\text{ax}}}/\sigma_{\Omega_{\text{ax}}}| \ll Z$. The dimensionless bias in this case [Eq. (29)] is essentially a dot product between the vectors $dC_{l}/d\Omega_{\text{ax}}$ and $\Delta C_{l}$, under a metric set by the data covariance.

As discussed in Appendix C, the Cauchy-Schwarz inequality thus imposes the bound $|\delta_{\Omega_{\text{ax}}}/\sigma_{\Omega_{\text{ax}}}| \leq Z$, but does not require the bias to saturate this bound. Examining the summand of this dot product, we note that the $N = 100$ implementation yields $\Delta C_{l}$ values that oscillate about zero with alternating sign, compared with the $N = 3$ and $N = 1.6$ implementations, which typically show a negative semidefinite offset ($\Delta C_{l} \leq 0$) at $\ell$ values that contribute significantly to the sum. This sign structure appears to be responsible for the single-parameter $|\delta_{\Omega_{\text{ax}}}/\sigma_{\Omega_{\text{ax}}}| \ll Z$ behavior of the $N = 100$ implementation.

VI. CONCLUSIONS

In this work, we have quantitatively compared predictions for cosmological observables in ultra-light axion (ULA) models obtained from the effective fluid approximation (EFA) with the results of a full Klein-Gordon (KG) treatment of the dynamics. Along the way, we found that an alternative treatment of ULA perturbations [81] is, in fact, equivalent to the EFA, when written in terms of appropriate variables; in future work, it would be valuable to undertake a more systematic comparison of our methods.

We forecasted the resulting bias to ULA parameters from primary (unlensed) CMB anisotropy power spectrum measurements, using well-established Fisher-matrix techniques to estimate how the numerical discrepancies between EFA and KG predictions will offset the centroids of the parameter likelihood, assuming that a ULA detection is hiding just underneath the sensitivity level of completed experimental analyses.

For the $N = 3$ version of the EFA, we find that if the full field dynamics are neglected, primary CMB anisotropy constraints to ULAs from Planck are robust, while constraints to measurements of $\Omega_{\text{ax}}$ from future CMB experiments (e.g., CMB-S4 [83]) will be moderately biased at the $\delta/\sigma \gtrsim 2$ level, if marginalization over the usual cosmological parameters is assumed. If an idealistic external dataset can break all the relevant degeneracies, the bias could be as high as $\delta/\sigma \sim \mathcal{O}(10)$. We find that no immediately planned effort will approach this bias level.

These conclusions could be altered when the full covariance matrix of future large-scale structure efforts (including degeneracies among standard cosmological parameters and their covariance with ULA parameters) is used, and we will address this issue in future work. For the $N = 1.6$ and $N = 100$ implementations in the marginalized case, we find comparable (but slightly worse, by a factor of $\sim 2$) bias levels when cosmological parameters are marginalized over.
It is interesting to note that the fractional differences between the different choices of $N$—which generate residuals well in excess of the $3/\epsilon$ threshold—lead to relatively small biases at least, in part, because the residuals have support over a large range of $\epsilon$. On the other hand, it stands to reason that if the residuals were more localized in $\epsilon$ we might have found a much larger bias (as recently pointed out in a different context in Ref. [92]).

Measurements of CMB weak lensing are a key scientific driver for upcoming experiments like CMB-S4 [83] or more futuristic ideas like CMB-HD [85]. These efforts have the promise of sensitivity to ULAs in the window $m_{\text{UL}} \sim 10^{-22}$ eV, where they might comprise all the DM. In future work, we will thus extend our work to include measurements of the lensing potential power spectrum and higher $m_{\text{UL}}$ values. The EFA generally works best at late times (where the lensing kernel peaks [151]), but it is also possible that the large information content of the full CMB 4-point correlation function (which drives constraints to the lensing potential power spectrum $C_{\ell}^{\gamma \gamma}$ [151]) induces larger biases.

The tools used in this work to compute observables for the “exact” computational benchmark only solve the Klein-Gordon equation until $m_{\text{UL}} = 10^4 H_0$, which is more than sufficient for the time and length scales probed by primary CMB anisotropies (as discussed in Sec. V.B). Future work will require us to extend the “exact” case to later times, in order to conduct definitive comparisons with the EFA in the case of CMB lensing, and other observables sensitive to the matter power spectrum (e.g., the galaxy correlation and shear power spectra to be probed with exquisite accuracy by LSST [152] and other comparable efforts).

In this work, we have assumed purely adiabatic initial conditions. If the relevant $U(1)$ global symmetry is broken before the end of the inflationary era, the ULA field (as a nearly massless scalar spectator) will carry quantum fluctuations with $\phi_1 \sim H_1/(2\pi)$, where $H_1$ is the Hubble expansion rate during inflation. These fluctuations will in turn source primordial DM isocurvature (entropy) fluctuations. (See Refs. [62] and references therein.) These are observationally known to be subdominant [153] (while still allowing large $\Omega_{\text{UL}}$ values), but could be detected by ongoing/upcoming CMB experimental efforts [83].

The CMB will thus provide interesting constraints to $H_1$ and $\Omega_{\text{UL}}$, complementing the impact of CMB polarization-based tests of the inflationary tensor-to-scalar ratio $r$. Entropy fluctuations are more directly sensitive to dark-sector fluctuations than the adiabatic CMB power spectrum, and it is thus important to fully evaluate the EFA-induced bias in ULA isocurvature scenarios.

**ACKNOWLEDGMENTS**

T. L. S. and D. G. acknowledge support in part by NASA Grant No. 80NSSC18K0728. D. G. and J. C. acknowledge support from the Provost’s office at Haverford College. J. C. thanks Haverford College for hospitality during the completion of this work. T. L. S. acknowledges support from the Provost’s office at Swarthmore 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 A. Aoki, R. Hložek, W. Hu, L. Hui, M. Kamionkowski, D. J. E. Marsh, C. Owen, V. Poulin, and M. Raveri for useful conversations. D. G. thanks KITP for its hospitality during the completion of this work.

**APPENDIX A: WKB**

Recall that for the ULA background, we want to solve

$$\phi''_0 + 2\frac{a'}{a} \phi'_0 + a^2 m^2 \phi_0 = 0. \tag{A1}$$

Substituting $\dot{\phi} = a\phi$, we find a differential equation solvable with the WKB method (see, for example, Ref. [154]). It has the solution,

$$\phi_0 = \frac{1}{aQ^{1/4}} \left( C_1 \exp \left( i \int_{\eta}^{\eta_0} \sqrt{Q} d\eta' + \cdots \right) + C_2 \exp \left( -i \int_{\eta}^{\eta_0} \sqrt{Q} d\eta' + \cdots \right) \right), \tag{A2}$$

with $Q = a^2 \tilde{m}^2 - a'' / a$. Self-consistently, if $H / \tilde{m} \ll 1$, then we assume $\rho_{\text{UL}} \sim 1 / a^3$ which implies $a'' / a^3 \lesssim H^2$, so we can assert $Q \approx a^2 \tilde{m}^2$. Then, the usual ansatz for the ULA field is recovered $\phi_0 = A \cos(\tilde{m} \theta + \theta) / a^{3/2}$ for some constant $A$ and $\theta$. It then immediately follows that $\rho_{\text{UL}} = A^2 \tilde{m}^2 / (2a^3)$ to leading order in $H / \tilde{m}$.

**APPENDIX B: ALTERNATIVE CYCLE AVERAGING PROCEDURES**

We note that our implementation of the $N = 100$ implementation [which applies a $2 \tilde{m} t = 100$ switch with equations of motion (EOMs) given by Eqs. (7)–(9), a $\langle w_{\text{UL}} \rangle = 0$ approximation, and the cycle-averaged sound speed, Eq. (11)] has additional gauge terms coming from the transformation from the fluid rest frame to the synchronous gauge. These terms are not present in Eqs. (21)–(23), which represent the formalism of Ref. [81] in terms of fluid variables.

There are also differences between our expression for the cycle-averaged sound speed [Eq. (11)], and the simpler expression $c_s^2 = k^2 / (4 m^2 a^2)$ obtained in Ref. [81].

As a result of the scaling of various terms with $k$, the first set of differences can only be large for $k \ll H$ (at super-horizon scales), while the difference between the two different cycle-averaged sound speeds only grows large when $k \gg \tilde{m} a$, deep in the Jeans-suppressed regime of perturbation evolution. To be sure that these implementation differences do not affect our conclusions, we wrote a modified code that exactly implemented Eqs. (21)–(23).

The resulting differences in observables are shown in Fig. 15. Fractional differences in power spectra between the different versions of the $N = 100$ implementation are
several orders of magnitude smaller than differences between it and other implementations at scales contributing to the sums in expressions for $Z$ and $\delta \Omega_{\text{ax}}$. Our conclusions about the relative accuracies of different EFA implementations are thus robust.

**APPENDIX C: Z-STATISTIC DERIVATION**

We define the following inner product (which clearly satisfies the usual axioms),

$$\langle \vec{a}, \vec{b} \rangle = \sum_{X,Y \in \{TT, TE, EE\}} \sum_{\ell = \ell_{\min}}^{\ell_{\max}} f_{\text{sky}} a^X_{\ell} b^Y_{\ell} (\Xi^X_{\ell} \Xi^Y_{\ell})_{\ell},$$

where the components of $\vec{a}$ are $a^X_{\ell}$ for $\ell = \ell_{\min}, \ell_{\min} + 1, \ldots, \ell_{\max}$ and $X \in \{TT, TE, EE\}$ in any arbitrary order. The Cauchy-Schwarz inequality holds for $\langle \vec{a}, \vec{b} \rangle$, as it is an inner-product. If we consider a one parameter model (so that $\sigma_\lambda = 1/\sqrt{F}$ for $F$ the Fisher matrix), we see

$$\frac{\delta_\lambda}{\sigma_\lambda} = \left| \langle \Delta C_\ell, \frac{\partial C_\ell}{\partial \lambda} \rangle \right| \leq \sqrt{\left| \langle \frac{\partial C_\ell}{\partial \lambda} \cdot \langle \Delta C_\ell, \Delta C_\ell \rangle \right|} \left| \langle \Delta C_\ell, \Delta C_\ell \rangle \right| = \frac{Z}{\sigma_\lambda}. \quad (C2)$$

FIG. 15. Plots of CMB power spectra comparing different implementations of the $N = 100$ implementation, as explained in Appendix B.
where we used Eqs. (28) and (31). Since $Z, \sigma_j \geq 0$, we see $Z \geq |\delta_j/\sigma_j|$. If there are multiple parameters being varied for the calculation of $\sigma_j = \sqrt{\langle F^{-1} \rangle_{ij}}$, the explicit bound may not hold, but the $Z$ statistic can still give an estimate of the dimensionless bias.

**APPENDIX D: ADDITIONAL CMB POWER SPECTRA**

**FIG. 16.** Relative error of CMB anisotropy power spectra computed with different EFA implementations ($N = 3$ and $N = 100$), computed in comparison to the exact solution. The constant $N$ defines the moment at which the exact equations are switched to the EFA, using the criterion $m_{\text{st}}/h > N H$ (in units where $c = 1$). The black curve $(3/\ell)$ is a rough precision threshold beyond which parameter biases may be significant [126]. If this curve is exceeded at many $\ell$ values by the actual EFA relative errors, an explicit computation of bias is needed to assess the full implications of these errors for cosmological parameter inference and ULA constraints.
How sound are our ultralight axion ...


[138] C. Hirata (private communication).


[143] T. Crawford (private communication).


[152] P. A. Abell et al. (LSST Science Collaboration and LSST Project), arXiv:0912.0201.
