

## Soliton merger rates and enhanced axion dark matter decay

Xiaolong Du<sup>1,2,\*</sup> David J. E. Marsh<sup>3,†</sup> Miguel Escudero<sup>4,‡</sup> Andrew Benson<sup>1,§</sup> Diego Blas<sup>5,6</sup>  
Charis Kaur Pooni<sup>3</sup> and Malcolm Fairbairn<sup>3</sup>

<sup>1</sup>*Carnegie Observatories, 813 Santa Barbara Street, Pasadena, California 91101, USA*

<sup>2</sup>*Department of Physics and Astronomy, University of California, Los Angeles, California 90095, USA*

<sup>3</sup>*Theoretical Particle Physics and Cosmology, King's College London,  
Strand, London, WC2R 2LS, United Kingdom*

<sup>4</sup>*Theoretical Physics Department, CERN, 1211 Geneva 23, Switzerland*

<sup>5</sup>*Grup de Física Teòrica, Departament de Física, Universitat Autònoma de Barcelona,  
08193 Bellaterra (Barcelona), Spain*

<sup>6</sup>*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology,  
Campus UAB, 08193 Bellaterra (Barcelona), Spain*



(Received 28 February 2023; accepted 28 September 2023; published 8 February 2024)

Solitons are observed to form in simulations of dark matter (DM) halos consisting of bosonic fields. We use the extended Press-Schechter formalism to compute the mass function of solitons, assuming various forms for the relationship between halo mass and soliton mass. We further provide a new calculation of the rate of soliton major mergers. Solitons composed of axion DM are unstable above a critical mass, and decay to either relativistic axions or photons, depending on the values of the coupling constants. We use the computed soliton major merger rate to predict the enhanced DM decay rate due to soliton instability. For certain values of currently allowed axion parameters, the energy injection into the intergalactic medium from soliton decays to photons is comparable to or larger than the energy injection due to core collapse supernovae at  $z > 10$ . A preceding paper explores the phenomenology of such an energy injection.

DOI: [10.1103/PhysRevD.109.043019](https://doi.org/10.1103/PhysRevD.109.043019)

### I. INTRODUCTION

Real scalar fields in general relativity have time-periodic, spatially localized, finite energy ground state solutions [1,2]. These solutions are known by various names: oscillatons, solitons, axion stars, and so forth. The inward force of gravity is balanced in these configurations by the outward pressure of scalar field gradients. In the nonrelativistic limit, these objects are truly stationary ground state soliton solutions of the Schrödinger-Poisson equations (see, e.g., Refs. [3–6]).<sup>1</sup>

If the observed [9] cosmological dark matter (DM) is composed of a real scalar field, or scalar fields, then such solitons should form in our Universe. Indeed, in numerical simulations of both axion [10] and “fuzzy” DM [11] solitons form in the centers of DM halos during the earliest stages of collapse, driven by the initial coherence of the field on scales near the de Broglie wavelength.

Furthermore, solitons can form by gravitational Bose-Einstein condensation [12,13] in any environment where the condensation timescale is shorter than the age of the Universe.

The possible existence of solitons, being very dense and coherent lumps of DM, opens a wide range of possibilities concerning their phenomenology. Soliton cores can affect stellar motions in the centers of galaxies (e.g., Refs. [14–17]). As “exotic compact objects,” gravitational waves (GWs) from soliton mergers can appear distinctly in GW observations (e.g., Ref. [18]). Soliton instabilities can also lead to new formation channels for black holes [19,20], or production of relativistic particles [21,22].

Exploiting solitons as a phenomenological window onto DM requires knowing the cosmological distribution of solitons, and their merger rates, which we dedicate this paper to studying. We study two mechanisms by which solitons lead to enhanced DM decay: by plasma blocking of parametric resonance, and by major mergers leading to formation of supercritical solitons.

In the present work, we will be concerned with soliton cores formed at very early times,  $z \gtrsim 10$ , in some of the first DM halos with masses  $M \gtrsim 10^{-5} M_{\odot}$ . The rough particle mass scale we are concerned with is  $m_a \approx 10^{-11}$  eV. For this mass scale, assuming an axionlike particle (ALP) with

\*xdu@astro.ucla.edu

†david.j.marsh@kcl.ac.uk

‡miguel.escudero@cern.ch

§abenson@carnegiescience.edu

<sup>1</sup>Solitons also form for massive spin-1 fields (see, e.g., Refs. [7,8] and references therein). In the following we simplify the discussion to scalar fields, although our methods can also be applied more generally.

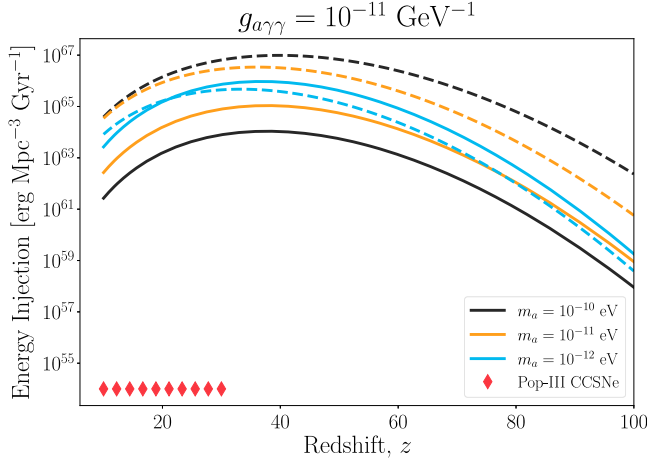


FIG. 1. Energy injection into the intergalactic medium from axion dark matter decay caused by soliton mergers. For comparison, a typical energy scale and redshift for energy injection from Pop-III core-collapse supernovae (CCSNe) is indicated [28]. For the parameters shown, axion star explosions inject more energy than supernovae. See [29] for the cosmological implications of such decays. Solid and dashed lines indicate different assumptions for the core-halo mass relation of  $\alpha = 1/3$  and  $\alpha = 3/5$ .

temperature-independent mass, halo formation is strongly suppressed for  $M < 8 \times 10^{-5} M_{\odot}$  [23]. Thus, the formation of the halos of interest is expected to be very similar to the formation of the first halos in thermal WIMP models, where the minimum halo mass is  $M \approx 10^{-6} M_{\odot}$  [24–26]. For standard  $\Lambda$ CDM cosmology the power spectrum on small scales is close to a fixed power law, which, together with the scaling symmetries of the Schrödinger-Poisson equations (and ignoring baryonic feedback), implies that the halos of interest formed within a few orders of magnitude of the cutoff scale should be morphologically similar to the  $\mathcal{O}(10^{10} M_{\odot})$  halos formed at  $z \approx 8$  and simulated directly for  $m_a \approx 10^{-22}$  eV [11]. Crucially, these first formed halos are expected to host a single soliton at their center and to obey a “core-halo mass relation” [27], which we discuss in detail below.

A key result of the present work is the computation of the energy injection into the intergalactic medium caused by the decay of axion dark matter due to soliton mergers. Therefore, note that in most of this work we use the terms “soliton” and “axion star” interchangeably. The energy injection density is shown for some representative parameters in Fig. 1. We compare this to an approximate energy injection due to core collapse supernovae from Pop-III stars [28]. The energy injection from axion dark matter soliton decay exceeds the supernova energy by many orders of magnitude, and extends to much higher redshifts, deep into the dark ages. This suggests that this phenomenon can place new constraints on axion dark matter, and open a new window onto axion observation. Details are explored in the preceding paper [29].

This paper is organized as follows. In Sec. II we compute the soliton mass function assuming various models for the core-halo mass relation between DM halos and soliton masses. We discuss the critical soliton mass in Sec. III, and compute the DM fraction that decays instantaneously at a given redshift due to plasma blocking. In Sec. IV we use the extended Press-Schechter model and stochastic merger trees to compute the soliton merger rate, and the DM decay rate from the formation of super-critical solitons by major mergers of sub-critical solitons. We conclude in Sec. V. The Appendices give details of our numerical scheme. We adopt fixed cosmological parameters  $\Omega_M = 0.3153$ ,  $\Omega_b = 0.04930$ ,  $h = 0.6736$ ,  $\sigma_8 = 0.8111$ ,  $n_s = 0.9649$  [30].

## II. SOLITON MASS FUNCTION

We consider DM composed of a real scalar field  $\phi$  minimally coupled to gravity with canonical kinetic term and potential  $V(\phi) = m_a^2 \phi^2/2 + \lambda \phi^4/4!$ , with  $m_a$  the DM particle mass (axion mass). In the nonrelativistic limit, solitons are given by the ground-state solutions of the Schrödinger-Poisson equations:

$$i\partial_t \psi = -\frac{1}{2m_a} \nabla^2 \psi + m_a \Phi \psi + \lambda |\psi|^2 \psi, \quad (1)$$

$$\nabla^2 \Phi = 4\pi G |\psi|^2, \quad (2)$$

where  $\Phi$  is the Newtonian gravitational potential,  $\lambda$  is the self-interaction coupling, and we use units  $\hbar = c = 1$ . The field  $\psi$  is related to the fundamental scalar field  $\phi$  by the WKB approximation:

$$\phi = \frac{1}{\sqrt{2}m_a} (e^{im_a t} \psi + e^{-im_a t} \psi^*). \quad (3)$$

In the relativistic limit, solitons are found as the time periodic solutions of the Einstein-Klein-Gordon equations [1,2]. The soliton solutions possess a scaling symmetry, and as such are uniquely specified by the soliton mass, or alternatively the central field value,  $\phi_0$ .

Simulations of DM structure formation with nonrelativistic scalar fields observe a scaling relation between the DM halo mass,  $M_h$  and the mass of the central soliton core,  $M_c$ , known as a “core-halo mass relation” [27]:

$$M_c = \frac{1}{4} \left[ \frac{M_h}{M_{\min}(z)} \right]^{\alpha} M_{\min}(z), \quad (4)$$

where  $z$  is cosmological redshift,  $\alpha$  is a power law exponent, and

$$M_{\min}(z) = 1.4 \times 10^{-6} \left( \frac{m_a}{10^{-13} \text{ eV}} \right)^{-3/2} \times \left[ \frac{\xi(z)}{\xi(0)} \right]^{1/4} (1+z)^{3/4} M_{\odot}, \quad (5)$$

is the minimum halo mass at  $z$ . The function  $\xi(z)$  is the virial density contrast [31]. By definition, the total mass of the central soliton/axion star  $M_S \approx 4M_c$  [27]. The core-halo mass relation can be understood by fixing the soliton radius from the de Broglie wavelength at the virial velocity of the halo with a universal coefficient [32,33].

It has been proposed that in Eq. (4)  $\alpha = 1/3$  is a universal relation in a “fully relaxed” DM halo where gravitational energy, kinetic energy, and field gradient energy are in virial equilibrium [6,16,27]. The exponent  $\alpha = 1/3$  is also the attractor solution following multiple mergers [34], and may be related to soliton condensation and growth saturation [13,35]. However, significant scatter in this relationship has also been observed [36–38], which might be explained by environmental factors, and/or merger history. Recently, it has been argued in Ref. [39] that the scatter in the halo-concentration relation could also lead to a sizable dispersion in the core-halo mass relation. The effect of scatter on merger rates is discussed in Sec. IV C.

We can relate the soliton mass function,  $F_S(M_S) = dn_S/d \ln M_S$  (where  $n_S$  is the soliton number density), to the halo mass function (HMF),  $F_h(M_h) = dn_h/d \ln M_h$ , by assuming one soliton per halo,  $n_S = n_h$ , i.e.:

$$F_S(M_S) d \ln M_S = F_h(M_h) d \ln M_h, \quad (6)$$

$$\Rightarrow F_S(M_S) = F_h(M_h(M_S)) \frac{d \ln M_h}{d \ln M_S}. \quad (7)$$

Given a core-halo mass relation such as Eq. (4) one can thus compute  $F_S$  given  $F_h$ . In the following we compute the halo mass function using the Sheth-Tormen multiplicity function [40]. This analytic approach has been shown to accurately reproduce the soliton mass function found in numerical zoom-in simulations by Ref. [33].

The primary ingredient in the HMF is the halo mass variance,  $\sigma^2(M_h)$ , which is computed from the matter power spectrum,  $P(k)$ , given a window function,  $\tilde{W}(k|M)$ . It will turn out that the most phenomenologically interesting halos have low mass,  $M \lesssim 1M_\odot$ . Accurately and quickly computing the mass variance at low mass requires special care, as described in Appendix A.

The power spectrum of scalar field DM displays a low mass cutoff due to the field gradients acting as an effective pressure and inducing a Jeans scale [41]. The linear power spectrum is strongly suppressed below the Jeans scale at radiation-matter equality [42,43]. In  $N$ -body simulations such as Refs. [23,44], suppression of the linear power spectrum leads to suppression of halo formation below the half-mode scale of the transfer function relative to CDM. The half-mode scale is redshift independent, since it is determined by initial conditions. On the other hand, the absolute minimum halo mass, Eq. (5), is determined by the redshift-dependent Jeans scale. The minimum halo mass is

in general some orders of magnitude smaller than the half-mode mass.

We compute the power spectrum as described in Appendix A, which is appropriate for an axionlike field given slow roll-initial conditions, a time-independent particle mass, adiabatic initial perturbations, and initial field value  $\phi/f_a \lesssim 1$  (where  $f_a$  is the axion decay constant in the interaction Lagrangian given below). Generalization of our results to other models requires the appropriate input  $P(k)$ , which may differ in shape (see e.g. Refs. [8,45,46]). In models with a small-scale (high- $k$ ) cut-off to  $P(k)$ , the HMF depends on the form of the window function used to compute  $\sigma$  [47]. We investigated both the sharp- $k$  space window function [47–49], and the smooth- $k$  space window function [50,51], which reproduce results of numerical simulations after so-called “spurious structure” is removed. In what follows, we use the smooth- $k$  space window function, since it gives a HMF in better agreement with numerical simulations when  $P(k)$  has a sharp cut-off at small scales.

We show the soliton mass function in Fig. 2 for various parameters. In the absence of a small-scale cutoff the soliton mass function has a universal form, with the soliton mass scaling inversely with axion mass. In the presence of a cutoff, the soliton mass function does not possess an exact scaling symmetry with  $m_a$ . We show how our results depend on the slope of the core-halo mass relation, and find that steeper slopes (larger  $\alpha$ ) lead to a higher number density of solitons, with the lowest number density for  $\alpha = 1/3$ . This is due to the mass function possessing a higher number density of low mass halos.

### III. SOLITON INSTABILITY

#### A. Critical solitons and critical halos

There are several instabilities shown in Fig. 3 that arise when solitons reach a critical mass. Correctly describing the instability requires specifying the interaction Lagrangian, and considering relativistic and environmental effects. All solitons are unstable to black hole formation above the Kaup mass [52]:

$$M_{\text{Kaup}} \approx 0.6 \times \frac{M_{\text{Pl}}^2}{m_a} \approx 8.5 \times 10^{-2} M_\odot \left( \frac{10^{-13} \text{ eV}}{m_a} \right). \quad (8)$$

This instability occurs when the soliton central field value  $\phi_0 \sim M_{\text{pl}}$  (with  $M_{\text{pl}} = 1/\sqrt{8\pi G}$  the reduced Planck mass). Soliton collapse to a BH during merger may produce distinct GW signals [53,54]. However, as we shall see, ordinary structure formation occurs in environments that are too underdense, and such mergers are expected to be very rare. Soliton collapse to BHs may occur, however, during the period of structure formation after inflation [33,55], or with enhanced primordial fluctuations [20].

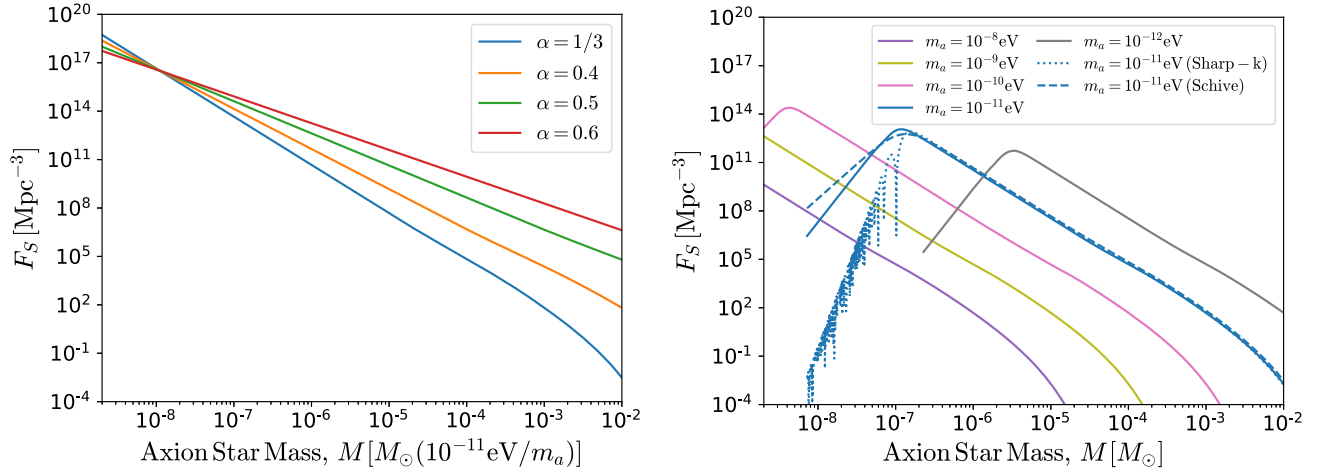


FIG. 2. The axion star mass function,  $F_S = dn_S/d \ln M_S$  defined by Eq. (7), using the Sheth-Tormen halo mass function at  $z = 10$ , assuming the core-halo mass relation, Eq. (4). Left: in the absence of the Jeans scale, the mass function has a scaling symmetry with all axion star masses proportional to  $1/m_a$ . We show various values of the exponent  $\alpha$  in the core-halo relation. Right: including the Jeans scale induces a cut-off and breaks the exact scaling symmetry with  $m_a$ . We show different models for the cut-off, and different axion masses, holding  $\alpha = 1/3$  fixed. We use the smooth- $k$  filter as our fiducial window function (solid lines). For  $m_a = 10^{-11}$  eV, we also show the axion star mass functions assuming the sharp- $k$  window function (dotted) and that computed from the halo mass function fit found by [23] (dashed).

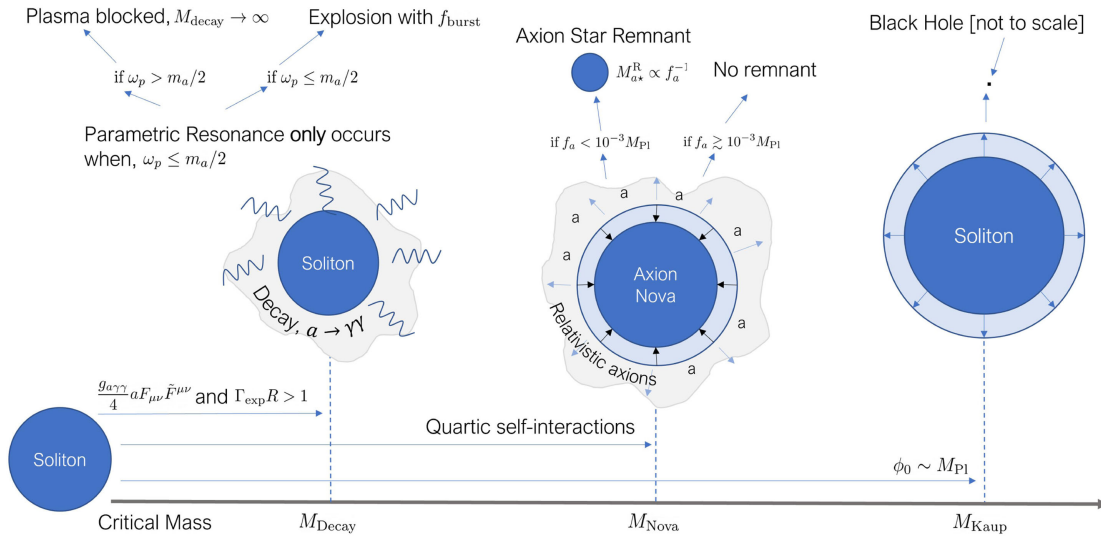


FIG. 3. Schematic of critical masses for *three* types of soliton instabilities with the smallest to largest critical masses corresponding to the Decay, Nova and Kaup instability. See Sec. III for more detail.

Solitons composed of axions, axion stars, are also unstable due to nonlinear interactions beyond  $m_a^2 \phi^2$  in the Lagrangian. The relevant terms in the interaction Lagrangian are

$$\mathcal{L}_{\text{int}} = -m_a^2 f_a^2 [1 - \cos(\phi/f_a)] - \frac{g_{a\gamma\gamma}}{4} \phi F_{\mu\nu} \tilde{F}^{\mu\nu}, \quad (9)$$

where  $F_{\mu\nu}$  is the photon field strength tensor, and  $\tilde{F}^{\mu\nu}$  is its dual.

In the first term in Eq. (9), we have introduced the axion decay constant,  $f_a$ , which controls the strength of the attractive axion quartic self-interactions,  $\lambda$ . The presence of these interactions triggers an instability of axion stars above a critical mass. The instability is known as an ‘‘axion Nova,’’ and leads to decay of the DM axion star into relativistic axions [19,21]. The critical mass for an axion Nova is

$$M_{\text{Nova}} \approx 0.1 M_{\odot} \left( \frac{f_a}{10^{14} \text{ GeV}} \right) \left( \frac{10^{-13} \text{ eV}}{m_a} \right), \quad (10)$$

The Nova leaves a remnant axion star, with the mass of the remnant depending inversely on the decay constant [21]. At large decay constant,  $f_a \gg 10^{14}$  GeV, there is no remnant and the total mass of the star is dispersed as relativistic axions. Instability to a Nova occurs at lower mass than the Kaup mass for  $f_a \lesssim 0.3M_{\text{pl}}$ , mapping out a “phase diagram” [19,56].

In the second term in Eq. (9), we have introduced the axion-photon coupling,  $g_{a\gamma\gamma}$ . This term also leads to instability of axion stars, driven by parametric resonance [22,57].<sup>2</sup> The critical mass is

$$M_{\text{Decay}} \approx 8.4 \times 10^{-5} M_{\odot} \left( \frac{10^{-11} \text{ GeV}^{-1}}{g_{a\gamma\gamma}} \right) \left( \frac{10^{-13} \text{ eV}}{m_a} \right). \quad (11)$$

The typical expectation for the axion photon coupling,  $g_{a\gamma\gamma}$ , is that it is related to the decay constant appearing in the self-interaction potential,  $f_a$ , via  $g_{a\gamma\gamma} = \alpha_{\text{EM}}/2\pi f_a$ , where  $\alpha_{\text{EM}}$  is the fine-structure constant. Taking  $f_a < M_{\text{pl}}$ , we then have  $M_{\text{Nova}} < M_{\text{Decay}} < M_{\text{Kaup}}$ , and so decay of solitons to relativistic axions via an axion Nova will be the dominant decay channel. However, there are many axion models with enhanced couplings  $g_{a\gamma\gamma}$  or suppressed potentials relative to the naive estimate such that soliton decay to photons may be the dominant channel (e.g., “aligned” models such as Ref. [59,60], and others [61]). Reference [22] further note that the initial collapse of an axion Nova can trigger the parametric resonance instability to photon production.

Taking any of the instability masses,  $M_{\text{Kaup}}$ ,  $M_{\text{Nova}}$ , of  $M_{\text{decay}}$ , rearranging Eq. (4) allows one to find which halos host unstable solitons. Thus, if there is a *universal* core halo mass relation, soliton instability occurs always in halos of a fixed mass with only mild redshift dependence. On the other hand, if there is intrinsic *diversity* in the core-halo mass relation then unstable solitons can occur in a diversity of halos. Furthermore, if  $\alpha = 1/3$ , the critical halo mass is particle mass independent and depends only on the coupling constants.

Assuming a universal core-halo mass relation, it is then useful to compare the critical halo mass hosting unstable solitons to scales in the halo mass function. The mass function has two scales: the minimum halo mass, Eq. (5), and the nonlinear mass scale, defined by  $\sigma(M_{\text{nl}}) = 1$ , above which the HMF is exponentially suppressed. Comparing these scales to the critical halo mass gives a rough estimate of when soliton decays are likely to occur in cosmic history, and is shown in Fig. 4. Larger values of  $\alpha$  and/or  $g_{a\gamma\gamma}$  lead to instability in lower mass,

<sup>2</sup>We have recently characterized this mechanism in full numerical relativity simulations [58].

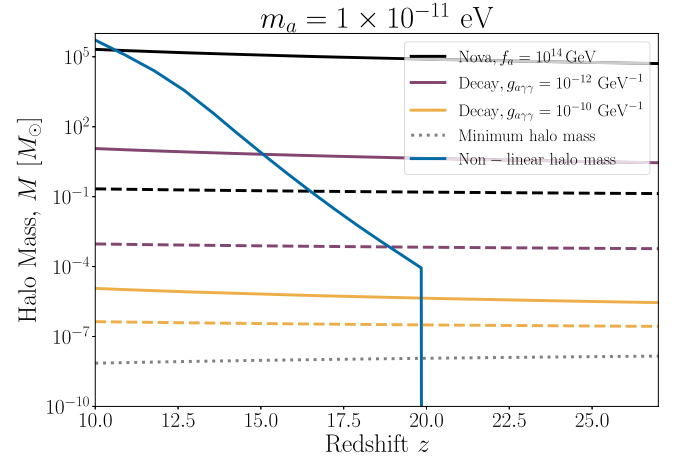


FIG. 4. Assuming an intrinsic core-halo mass relation, Eq. (4), with slope parameter  $\alpha$ , instability of a soliton to Nova [Eq. (10)] or decay to photons [Eq. (11)] occurs in halos of a fixed critical mass, as indicated. Solid lines show  $\alpha = 1/3$ , and dashed show  $\alpha = 3/5$ . For reference, we also show the nonlinear mass,  $\sigma(M_{\text{nl}}) = 1$ , and the minimum halo mass [Eq. (5)]. For the values of the axion-photon coupling and decay constant shown, decay of axion stars to photons occurs before the critical mass for an axion nova can be reached.

and thus more abundant, halos, and increases the redshift where typical decays occur.

It was shown in Ref. [22] that if an axion star close to  $M_{\text{Decay}}$  grows by slow adiabatic accretion from the background, then it loses energy by efficient photon emission, and returns to a stable condition below  $M_{\text{Decay}}$ . The same is expected to be true for axion emission close to the Nova instability. The growth rate of solitons by accretion is slow, growing as  $t^{1/2}$  at low mass [12], and even slower once virial equilibrium is reached [13] (see also [62]). Therefore, soliton mass loss by production of photons or axions from dark matter accretion in the host halo is expected to be negligible over a Hubble time.

There are two methods to obtain rapid decay of solitons:

- (i) Plasma blocking;
- (ii) Major mergers.

Plasma blocking effectively sends  $M_{\text{Decay}} \rightarrow \infty$  as long as  $\omega_p(z) > m_a/2$ , allowing supercritical solitons with masses larger than Eq. (11) to form. Once the plasma frequency drops to  $\omega_p(z) < m_a/2$  then the decay is kinematically allowed, and supercritical solitons will then decay. We discuss this case in the next subsection.

Nonadiabatic soliton growth can occur during major mergers (see, e.g., Ref. [63]). In such a case, a soliton can jump above the critical barrier and will rapidly decay losing an  $\mathcal{O}(1)$  fraction of its mass. This process is relevant for decay either to photons or to relativistic axions [i.e., either Eq. (11) or (10)]. We thus expect that the halo major merger

rate evaluated at the critical halo mass can be used to estimate the enhanced dark matter decay rate due to soliton mergers. We calculate this in Sec. IV.<sup>3</sup>

### B. Soliton decay to photons from plasma blocking

Decay of axions to two photons is blocked if the plasma frequency,  $\omega_p$ , in the environment satisfies  $\omega_p > m_a/2$ . The low mass DM halos of interest in the present work are below the baryon Jeans scale, and do not possess any cold gas. Therefore, the relevant plasma frequency is given by that of the intergalactic medium (IGM), and is determined by the evolution of the free electron density of the Universe,  $n_e(z)$ , which is well understood [64]. The Universe becomes transparent to photons at recombination,  $z \approx 1100$  when  $\omega_p \approx 10^{-9}$  eV. The plasma frequency decreases as the Universe grows and cools, reaching a minimum of  $\omega_p \approx 10^{-14}$  eV when the first stars begin to reionize the IGM, which occurs between  $z = 6$  and  $z = 10$ . Thus, post recombination plasma blocking can be neglected for  $m_a \gtrsim 10^{-9}$  eV.<sup>4</sup>

When the plasma frequency achieves  $\omega_p(z_{\text{crit}}) = m_a/2$ , decay of all super-critical solitons occurs at once, in a burst. The energy released can be expressed as a fraction of the dark matter density at  $z_{\text{crit}}$  by integrating the soliton mass function:

$$f_{\text{burst}} = \frac{1}{\rho_{\text{DM}}} \int_{M_{\text{Decay}}}^{\infty} (M_S - M_{\text{Decay}}) F_S(M_S, z_{\text{crit}}) d \ln M_S. \quad (12)$$

With Eq. (12) one can then calculate the fractional dark matter energy density that is converted into photons once  $\omega_p(z_{\text{crit}}) = m_a/2$ . However, it is important to highlight that once the energy is injected and transformed into heat and leads to reionization  $\omega_p(z_{\text{crit}})$  will in turn increase and possibly eventually drop again below  $\omega_p(z_{\text{crit}-2}) = m_a/2$  at a later time,  $z_{\text{crit}-2} < z_{\text{crit}}$ . Calculating the subsequent burst(s) energy requires solving for the free electron fraction and baryon temperature: this is addressed in the preceding paper Ref. [29].

<sup>3</sup>Reference [22] also computes a maximum relative velocity that solitons can have in order not to Doppler shift the relative frequencies too much and block parametric resonance. This relative velocity is computed for head on soliton collisions. In halo mergers, dynamical friction effects slow down the cores such that mergers occur. As such, we ignore the maximum velocity constraint in the following.

<sup>4</sup>Plasma effects on axion photon conversion in neutron star magnetospheres are discussed in, e.g., Refs. [65,66]. Pre-recombination axion-photon conversion is discussed in Ref. [67].

## IV. SOLITON MERGERS AND DARK MATTER DECAY

### A. Halo formation rate from extended Press-Schechter formalism

To compute the decay rate of axions due to soliton mergers, we need to compute the formation rate of critical axion stars, which is related to the formation rate of corresponding DM halos by the core-halo mass relation. The formation rate of DM halos have been studied extensively in the last decades using analytic formalism and/or numerical simulations. We refer the readers to Refs. [68–71] for more detailed discussions. For completeness, we briefly summarize the procedure we have taken below.

The naive halo formation rate comes from the time derivative (redshift derivative) of the mass function. This does not give the correct merger rate, because halos at a specific mass are both formed and destroyed by mergers at the same time. Thus, we can express the redshift derivative of the mass function as [70,71]:

$$\frac{d^2 n_h(M, z)}{dM dz} = \frac{d^2 n_{\text{form}}(M, z)}{dM dz} - \frac{d^2 n_{\text{dest}}(M, z)}{dM dz}. \quad (13)$$

The formation rate of halos per unit mass and per unit volume is given by

$$\frac{d^2 n_{\text{form}}(M, z)}{dM dz} = \int_0^M \frac{M'}{M} \frac{dn_h(M', z)}{dM'} \frac{d^2 f_{2 \rightarrow 1}(M, M', z)}{dM dz} dM', \quad (14)$$

with  $d^2 f_{2 \rightarrow 1}(M, M', z)/dM/dz$  the fraction of mass in a halo of mass  $M'$  that merges into halos of mass  $M$  at a later time per unit mass and per unit redshift. Here the subscripts “1” and “2” denote halos corresponding to the first ( $M$ ) and second ( $M'$ ) arguments, respectively. The arrow indicates the mass flow, e.g.,  $2 \rightarrow 1$  represents the mass in halo 2 merges into halo 1, while  $1 \leftarrow 2$  represents halo 1 obtains mass from halo 2. On the other hand, the destruction rate of halos per unit mass and per unit volume is given by

$$\frac{d^2 n_{\text{dest}}(M, z)}{dM dz} = \int_M^{+\infty} \frac{dn_h(M, z)}{dM} \frac{d^2 f_{1 \rightarrow 2}(M', M, z)}{dM' dz} dM'. \quad (15)$$

Note that in the above equation we do the integration with respect to the descendant halo mass. One can either compute the halo formation rate directly from Eq. (14) or first compute the halo destruction rate from Eq. (15) and then convert it to formation rate using Eq. (13). In this work, we choose the former approach.

The mass function and formation rate can be estimated using extended Press-Schechter (EPS) theory [48,68] as follows. Consider a random overdensity field

$\delta(\mathbf{x}) \equiv \delta\rho_m/\rho_m$  linearly extrapolated to the present time whose power spectrum is  $P(k)$ . One can smooth out overdensities below the scale  $R$  by convolving  $\delta(\mathbf{x})$  with a window function  $W(\mathbf{x}|R)$ :

$$\delta^S(\mathbf{x}) = \int \delta(\mathbf{x})W(\mathbf{x} + \mathbf{x}'|R)d^3\mathbf{x}'. \quad (16)$$

The variance of  $\delta^S(\mathbf{x})$  is given by

$$S(R) \equiv \sigma^2(R) = \int \frac{k^2}{2\pi^2}P(k)\tilde{W}(k|R)^2dk, \quad (17)$$

where  $\tilde{W}(k|R)$  is the Fourier transform of  $W(\mathbf{x}|R)$ .

For  $R \rightarrow +\infty$ ,  $S = 0$  corresponding to  $\delta^S(\mathbf{x}) = 0$  everywhere. As the smoothing scale decreases, more and more perturbations on small scales are included, so  $S$  increases. The trajectory of  $\delta^S(\mathbf{x})$  at a spatial position is a random walk if we treat  $S$  as the time variable. As  $S$  increases,  $\delta^S(\mathbf{x})$  will eventually pass a certain threshold,  $\delta_c$ , which is called the critical overdensity for collapse—the mass element at  $\mathbf{x}$  is then considered to be included in a collapsed halo with a mass  $M$  that corresponds to the smoothing scale  $R$ . The relation between  $M$  and  $R$  depends on the choice of window function (see Appendix A for details). By computing the probability that  $\delta^S(\mathbf{x})$  makes its first upcrossing of  $\delta_c$  at  $S(M)$ ,<sup>5</sup> one can estimate the fraction of mass in the Universe that is contained in halos with a mass  $M$ . Assuming the spherical collapse model [72], the critical overdensity for collapse of CDM (valid for axion DM above the Jeans scale) is mass independent and only a function of redshift. The probability that  $\delta^S(\mathbf{x})$  first crosses  $\delta_c$  upward at  $S(M)$  per unit  $S$  is given by

$$\frac{df(S, z)}{dS} = \frac{\delta_c(z)}{\sqrt{2\pi}S} \exp\left[-\frac{\delta_c^2(z)}{2S}\right]. \quad (18)$$

A fitting function for  $\delta_c$  in  $\Lambda$ CDM is given by [73]

$$\delta_c(z) = \frac{3(12\pi)^{2/3}D(0)}{20} \frac{D(0)}{D(z)} [1 + 0.0123\log_{10}\Omega_m(z)]. \quad (19)$$

Here  $\Omega_m(z)$  is the fractional matter density at  $z$ , i.e.  $\Omega_m(z) = \Omega_{m,0}(1+z)^3/[\Omega_\Lambda + \Omega_{m,0}(1+z)^3]$  and  $D(z)$  is the linear growth factor of matter perturbations [74]

$$D(z) = \frac{1}{1+z} {}_2F_1\left(\frac{1}{3}, 1, \frac{11}{6}, -\frac{\Omega_\Lambda}{\Omega_{m,0}(1+z)^3}\right). \quad (20)$$

Having  $df(S, z)/dS$ , the halo mass function can be computed as

<sup>5</sup>Since  $S$  is a monotonic function of  $R$  and  $R$  is a monotonic function of  $M$ , here we write  $S$  as a function of  $M$ .

$$\frac{dn_h(M, z)}{dM} = \frac{\rho_m}{M} \frac{df(S, z)}{dS} \left| \frac{dS}{dM} \right|. \quad (21)$$

Here  $\rho_m$  is the mean comoving matter density in the Universe.

While Eqs. (18) and (21) agree reasonably well with  $N$ -body cosmological simulations, they overpredict the abundance of large halos and underpredicts that of small halos. A more accurate model is to consider ellipsoidal collapse, which gives rise to the Sheth-Tormen fitting function [75,76] that we have used in Sec. II:

$$\frac{df^{\text{ST}}(S, z)}{dS} = A \sqrt{\frac{1}{2\pi}} \sqrt{q\nu} [1 + (\sqrt{q\nu})^{-2p}] \exp\left(-\frac{q\nu^2}{2}\right) \frac{1}{S}, \quad (22)$$

where  $\nu \equiv \frac{\delta_c(z)}{\sqrt{S}}$ ,  $A = 0.3222$ ,  $p = 0.3$ , and  $q = 0.707$ .

Now let us consider a trajectory starting from  $(\delta_c(z), S)$  and then first upcrossing the critical overdensity for collapse at an earlier redshift  $z'$  and a larger variance  $S'$ . This corresponds to a halo  $M$  collapsing at  $z$  and having a progenitor halo  $M'$  at  $z'$ . Replacing  $\delta_c$  and  $S$  in Eq. (18) by  $\delta_c(z') - \delta_c(z)$  and  $S' - S$ , we obtain the probability for such an event:

$$\begin{aligned} & \frac{df_{1\leftarrow 2}(S(M), S'(M'))}{dS'} \\ &= \frac{\delta_c(z') - \delta_c(z)}{\sqrt{2\pi}(S' - S)} \exp\left[-\frac{(\delta_c(z') - \delta_c(z))^2}{2(S' - S)}\right] \frac{1}{S' - S}. \end{aligned} \quad (23)$$

Equation (23) is related to the progenitor mass function of halos of mass  $M$ , and as we will see, can be used to compute the merger rate, which is found from the reverse conditional probability  $df_{2\rightarrow 1}(S, S')/dS$ . First, we compute the backward rate, taking a small redshift step  $\Delta z$ :

$$\begin{aligned} & \frac{d^2 f_{1\leftarrow 2}(S, S')}{dS' dz} \\ &= \frac{\delta_c(z + \Delta z) - \delta_c(z)}{\Delta z} \\ &\times \frac{1}{\sqrt{2\pi}(S' - S)} \exp\left[-\frac{(\delta_c(z + \Delta z) - \delta_c(z))^2}{2(S' - S)}\right] \frac{1}{S' - S}. \end{aligned} \quad (24)$$

When  $\Delta z \rightarrow 0$ , the exponential term in Eq. (24) approaches 1 and thus can be removed from the equation, which is commonly done in the literature (e.g., Ref. [68]), and the term involving  $\delta_c$  is simply the derivative with respect to  $z$ . However, we keep the exponential term and take a small but finite time step size when computing the halo formation rate to avoid divergent results when  $S' \rightarrow S$  upon taking integrals over  $S'$  (which we do shortly). This is because the

quantities are probability distributions, and the order of limits and integrals can be important.

As in the case of the halo mass function a more accurate formula for the rate is obtained by accounting for departures from spherical collapse using an empirical modification calibrated to  $N$ -body simulations [77]:

$$\frac{d^2 f_{1\leftarrow 2}^{\text{Nbody}}(S, S')}{dS' dz} = \frac{d^2 f_{1\leftarrow 2}(S, S')}{dS' dz} G(S, S'), \quad (25)$$

where  $G$  is

$$G(S, S') = G_0 \left(\frac{S'}{S}\right)^{\gamma_1/2} \left(\frac{\delta_c(z)^2}{S}\right)^{\gamma_2/2}, \quad (26)$$

with  $G_0 = 0.57$ ,  $\gamma_1 = 0.38$  and  $\gamma_2 = -0.01$ .

We can now find the probability per unit redshift that a halo of mass  $M'$  merges into a halo of mass  $M$  at a later time [68]:

$$\frac{d^2 f_{2\rightarrow 1}^{\text{Nbody}}(S, S')}{dS dz} = \frac{d^2 f_{1\leftarrow 2}^{\text{Nbody}}(S, S')}{dS' dz} \frac{df^{\text{ST}}(S, z)/dS}{df^{\text{ST}}(S', z')/dS'}. \quad (27)$$

This reversal of conditional probabilities follows from conservation of mass, i.e. the mass that halo  $S$  gets from halo  $S'$  equals to the mass that halo  $S'$  merges into halo  $S$ .

Plugging Eqs. (21) and (27) into Eq. (14), the halo formation rate is then given by:

$$\begin{aligned} \frac{d^2 n_{\text{form}}}{dM dz} &= \int_S^{+\infty} \frac{M'}{M} \frac{\rho_m}{M'} \frac{df^{\text{ST}}(S', z')}{dS'} \frac{d^2 f_{2\rightarrow 1}^{\text{Nbody}}(S, S')}{dS dz} \Big|_{\frac{dS}{dM}} dS' \\ &= \frac{\rho_m}{M} \Big|_{\frac{dS}{dM}} \int_S^{+\infty} \frac{df^{\text{ST}}(S', z')}{dS'} \frac{d^2 f_{2\rightarrow 1}^{\text{Nbody}}(S, S')}{dS dz} dS' \\ &= \frac{\rho_m}{M} \frac{df^{\text{ST}}(S, z)}{dS} \Big|_{\frac{dS}{dM}} \int_S^{+\infty} \frac{d^2 f_{1\leftarrow 2}^{\text{Nbody}}(S, S')}{dS' dz} dS' \\ &= \frac{dn_{\text{h}}^{\text{ST}}(M, z)}{dM} \int_{S(M-M_{\text{res}})}^{S_{\text{max}}} \frac{d^2 f_{1\leftarrow 2}^{\text{Nbody}}(S, S')}{dS' dz} dS'. \quad (28) \end{aligned}$$

Here in the second and last lines we have used the definition of the halo mass function Eq. (21). In the last line, we also introduced finite  $S_{\text{max}}$  in the maximum of the integral, and finite mass resolution  $M_{\text{res}}$  in the minimum. From Eq. (28), we can see that the halo formation rate is proportional to the halo mass function.

The upper limit in the integral in Eq. (28),  $S_{\text{max}}$ , should be the largest value of  $S$  (which occurs for  $M \rightarrow 0$ ). For CDM,  $S_{\text{max}} \rightarrow \infty$ . However, the axion Jeans scale (or indeed the free streaming scale of a thermally produced WIMP) provides a natural upper limit and finite  $S_{\text{max}}$ . To avoid the divergence of the integral, we only account for halos that merged into  $M$  with a mass smaller than  $M - M_{\text{res}}$ , i.e., replacing the lower limit of the integral with  $S(M - M_{\text{res}})$ . Furthermore, as noted earlier, convergence of the integral

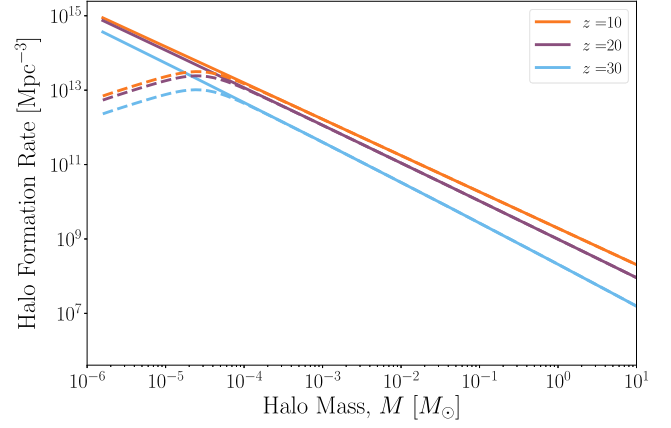


FIG. 5. Halo formation rate, per logarithmic mass bin, per redshift, with mass resolution  $M_{\text{res}} = 10^{-3}M$ . Solid lines show  $m_a = 10^{-9}$  eV, while dashed lines show  $m_a = 10^{-11}$  eV. The computation uses the  $N$ -body fit in Eq. (25) from Ref. [77], which agrees to within a factor of  $\mathcal{O}(3)$  with the EPS result for spherical collapse for  $z \gtrsim 10$ .

requires retaining finite  $\Delta z$  in Eq. (24), which retains the exponential factor, until after the integral has been performed. Retaining explicit resolution factors everywhere is consistent with the computation of these quantities in  $N$ -body simulations and in merger trees, which we use to calibrate and check the analytic results.

Figure 5 compares the halo formation rate for two axion masses,  $m_a = 10^{-9}$  eV and  $m_a = 10^{-11}$  eV. For  $m_a = 10^{-11}$  eV the effect of the Jeans scale becomes apparent for  $M \lesssim 10^{-4}M_{\odot}$ . Comparing to Fig. 4, we notice that although for  $m_a = 10^{-11}$  eV typical fluctuations only become nonlinear for  $z < 20$ , there is still significant halo formation occurring at earlier times from rarer fluctuations (consider the  $z = 30$  curve in Fig. 5). We further notice that for  $m_a = 10^{-11}$  eV the turn over in halo formation rate at low masses will suppress the formation of critical halos for  $g_{\text{eff}} \gtrsim 10^{-10}$  GeV $^{-1}$ . Bearing in mind these considerations, we now estimate the halo major merger rate, and dark matter decay rate from near-critical soliton mergers.

## B. Dark matter decay rate

As is shown in Ref. [78], during a binary merger, the mass of an axion star increases only when the mass ratio of the two axion stars  $\mu = M_{c2}/M_{c1} > 3/7$  (assuming  $M_{c2} \leq M_{c1}$ ). Therefore, we will consider only *major mergers* for which the halo mass ratio of two progenitors  $\mu_{\text{h}} > \mu_{\text{h},\text{min}} = (3/7)^{1/\alpha}$ , for  $M_c \propto M_{\text{h}}^{\alpha}$ . Then Eq. (28) becomes

$$\begin{aligned} \frac{d^2 n_{\text{form}}(M, z)}{dM dz} &= \frac{dn_{\text{h}}(M, z)}{dM} \\ &\times \int_{S(M_{\text{upper}})}^{S(M_{\text{lower}})} \frac{d^2 f_{1\leftarrow 2}^{\text{Nbody}}(S, S')}{dS' dz} dS', \quad (29) \end{aligned}$$



where

$$M_{\text{upper}} = \min \left\{ \frac{1}{1 + \mu_{h,\text{max}}} M, M_{h,\text{crit}}(z) \right\}, \quad (30)$$

$$M_{\text{lower}} = \max \left\{ \frac{\mu_{h,\text{min}}}{1 + \mu_{h,\text{min}}} M, M - M_{h,\text{crit}}(z) \right\}. \quad (31)$$

In the above equations, we have imposed another condition that the mass of two progenitor halos are both smaller than the critical halo mass.

Additionally, since the halo mass after a merger is not necessarily exactly equal to the critical halo mass, e.g., two halos with masses below the critical halo mass merge and form a bigger halo with a mass larger than the critical value, we integrate (29) over the interval  $(M_{h,\text{crit}}, 2M_{h,\text{crit}})$  to get the total number of halos that produce axion stars above the critical mass between redshifts  $z + \Delta z$  and  $z$ .

We thus arrive at a key result: the major merger rate of solitons around a given critical mass is given by

$$\begin{aligned} \frac{dn_{\text{merger}}}{dz} &= \int_{M_{\text{halo,crit}}}^{2M_{\text{halo,crit}}} \frac{dn_h(M, z)}{dM} \\ &\times \int_{S(M_{\text{upper}})}^{S(M_{\text{lower}})} \frac{d^2 f_{1 \leftarrow 2}^{\text{Nbody}}(S, S')}{dS' dz} dS' dM. \quad (32) \end{aligned}$$

An example of the soliton major merger rate is shown in Fig. 6. The rate is computed for a given soliton mass  $M_{\text{sol}}$ . The mass is chosen to correspond to the critical mass for decay to photons, Eq. (11) for various values of  $g_{\alpha\gamma\gamma}$ .

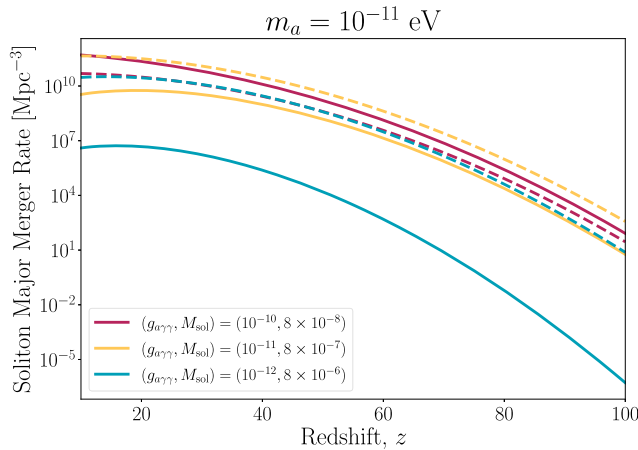


FIG. 6. Soliton major merger rate density (per unit redshift, i.e. dimensionless time) for  $m_a = 10^{-11}$  eV. The major merger rate is evaluated at a given soliton mass, measured in  $M_{\odot}$ , which corresponds to the critical soliton mass for a corresponding value of  $g_{\alpha\gamma\gamma}$  (in  $\text{GeV}^{-1}$ ). For different applications, the soliton merger rate can be computed at any desired soliton mass regardless of instability/criticality. Solid lines correspond to core-halo mass relation  $\alpha = 1/3$  and dashed lines to  $\alpha = 3/5$ .

However, the same rate calculation can be used for any soliton mass of interest. We also demonstrate dependence on the core-halo mass relation parameter,  $\alpha$ .

We can reexpress Eq. (32) in terms of the fractional decay rate of dark matter,  $df_{\text{merger}}/dz$ :

$$\frac{df_{\text{merger}}}{dz} = \frac{M_{S,\text{crit}}}{\rho_m} \frac{dn_{\text{merger}}}{dz}. \quad (33)$$

Note that here we have neglected an  $\mathcal{O}(1)$  factor and assumed that the axion star completely decays. For  $M_{\text{half}} < M_{h,\text{crit}} < 10^5 M_{\odot}$  (where  $M_{\text{half}}$  is the mass corresponding to the half-mode scale in the transfer function relative to CDM), Eq. (33) can be well approximated by

$$\frac{df_{\text{merger}}}{dz} = \chi \frac{M_{S,\text{crit}}}{\rho_m} \frac{dn_h}{d \ln M} \Big|_{M=M_{h,\text{crit}}}, \quad (34)$$

where  $\chi$  is an  $\mathcal{O}(0.1)$  coefficient

$$\chi = \frac{a(z) + b(z) \left[ \log_{10} \left( \frac{M_{h,\text{crit}}}{10^{-6} M_{\odot}} \right) \right]}{1 + c(z) \left( \frac{M_{h,\text{crit}}}{10^5 M_{\odot}} \right)^{0.39}}. \quad (35)$$

The parameters  $a$ ,  $b$  and  $c$  are redshift dependent:

$$a = a_0 + a_1 z + a_2 z^2, \quad (36)$$

$$b = b_0 + b_1 z + b_2 z^2, \quad (37)$$

$$c = c_0 + c_1 z + c_2 z^2. \quad (38)$$

The best-fit coefficients for several different values of  $\alpha$  are given in Table I.

Figure 7 shows the DM fractional decay rate,  $df_{\text{merger}}/dz$ , for a variety of model parameters. For fixed axion parameters  $m_a$  and  $g_{\alpha\gamma\gamma}$  there is a large dependence on the core-halo mass relation, expressed via  $\alpha$ . For  $\alpha = 1/3$  we notice

TABLE I. Best-fit coefficients in Eqs. (36)–(38) for different values of slope parameter  $\alpha$ .

	$\alpha = 1/3$	$\alpha = 2/5$	$\alpha = 3/5$
$a_0$	$1.07369 \times 10^{-1}$	$9.86049 \times 10^{-2}$	$8.00409 \times 10^{-2}$
$a_1$	$-7.55082 \times 10^{-5}$	$-7.07799 \times 10^{-5}$	$-6.22972 \times 10^{-5}$
$a_2$	$-9.39435 \times 10^{-7}$	$-9.54236 \times 10^{-7}$	$-9.31684 \times 10^{-7}$
$b_0$	$6.84693 \times 10^{-3}$	$6.28750 \times 10^{-3}$	$5.10077 \times 10^{-3}$
$b_1$	$-1.40186 \times 10^{-5}$	$-1.36381 \times 10^{-5}$	$-1.31868 \times 10^{-5}$
$b_2$	$-1.85082 \times 10^{-7}$	$-1.86434 \times 10^{-7}$	$-1.68590 \times 10^{-7}$
$c_0$	$2.71084 \times 10^{-1}$	$2.65829 \times 10^{-1}$	$2.57776 \times 10^{-1}$
$c_1$	$2.47512 \times 10^{-3}$	$2.67660 \times 10^{-3}$	$2.76791 \times 10^{-3}$
$c_2$	$6.89772 \times 10^{-5}$	$7.77110 \times 10^{-5}$	$1.02358 \times 10^{-4}$

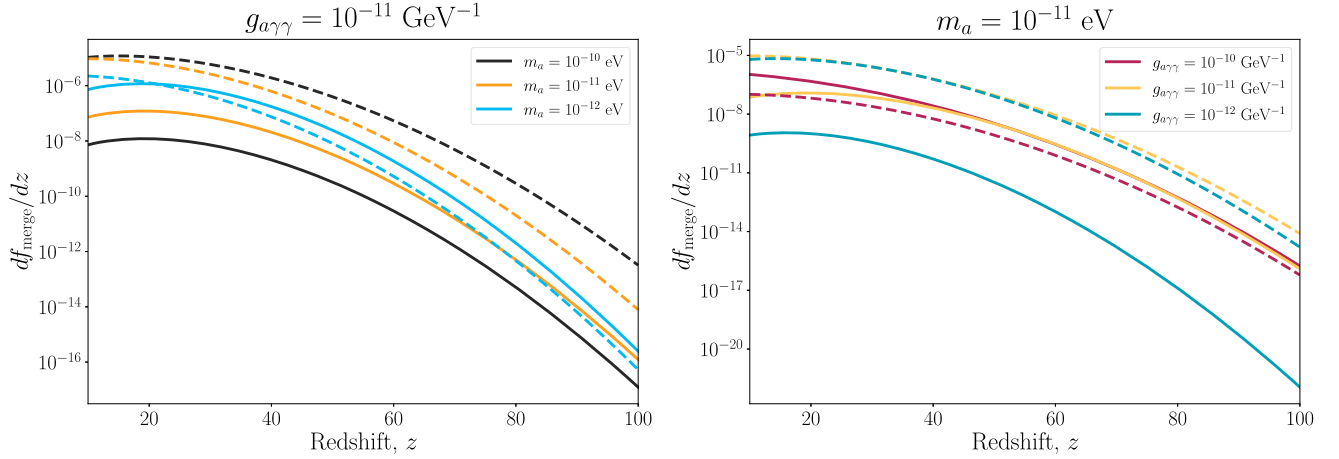


FIG. 7. Axion dark matter fractional decay rate to photons due to soliton major mergers and parametric resonance. Solid lines assume a core-halo mass relation with slope  $\alpha = 1/3$ , while dashed lines show  $\alpha = 3/5$ . Left: dependence on axion mass at fixed coupling. Right: dependence on coupling at fixed axion mass.

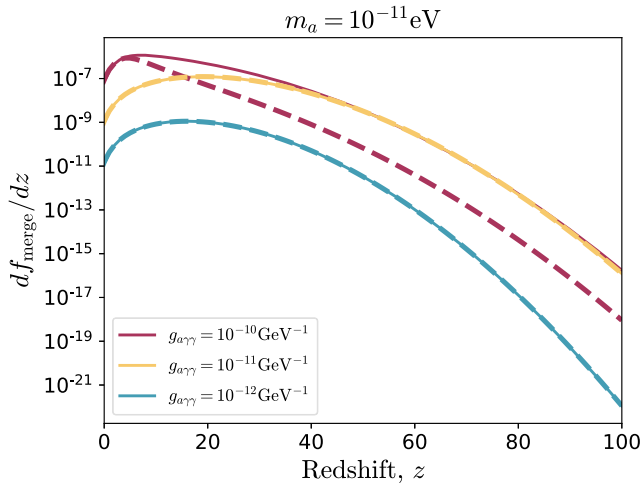


FIG. 8. Comparison of fractional dark matter decay rate due to soliton mergers comparing the full EPS calculation, Eq. (33) (solid lines), to the fitting formula Eq. (34) (bold dashed lines). We fix the axion mass  $m_a = 10^{-11}$  eV and core-halo mass relation  $\alpha = 1/3$ .

almost constant power law scaling of the result with  $m_a$  and  $g_{a\gamma\gamma}$ .

Figure 8 compares the DM fractional decay rate from the full EPS calculation to the fit given in Eq. (34). The fit is in general very good, except that the fit breaks down at high values of  $g_{a\gamma\gamma} \gtrsim 10^{-10}$  GeV $^{-1}$  when the critical halo mass drops below  $M_{\text{half}}$  where the halo mass function shape changes abruptly. Such high values of  $g_{a\gamma\gamma}$ , however, are robustly excluded by nonobservation of solar axions by the Cern Axion Solar Telescope [79]. We conclude that the fit in Eq. (34) can be used quite generally.

The decay rate, Eq. (33) can be converted into energy per unit volume per unit time with appropriate factors of the

Hubble parameter,  $H(z)$ , and dark matter density  $\Omega_{\text{DM}}h^2 = 0.12$ . The energy injection is shown in Fig. 1, compared to a typical energy injection from supernovae. For the supernova model we use the results of Ref. [28] for core-collapse Pop-III supernovae. We approximate the star formation rate density as a constant between redshifts 10 and 30, and assume one core collapse supernova per 100 Solar masses of star formation. We observe that for the axion parameters considered the energy injection into the intergalactic medium caused by axion star explosions is significantly larger than the energy injection due to supernovae, and with energy injection extending significantly into the dark ages at  $z \gg 20$ , suggesting that this is a phenomenon with observable consequences.

The python code we used to do the calculations in this subsection is publicly available at [https://github.com/Xiaolong-Du/Merger\\_Rate\\_of\\_Axion\\_Stars](https://github.com/Xiaolong-Du/Merger_Rate_of_Axion_Stars).

### C. Merger trees

In the previous subsection, we showed the calculations of DM fractional decay rate for a given power-law core-halo mass relation using the EPS formalism. In reality, there can be a large dispersion in the core-halo mass relation [37] and a universal core-halo mass relation may not exist [38]. Thus for fixed axion parameters,  $m_a$  and  $g_{a\gamma\gamma}$ , the critical halo mass also has a scatter making it difficult to compute the DM fractional decay rate using Eqs. (32) and (33). The lower and upper limits of the integral are not determined by a unique core-halo mass relation, e.g. the halo and its progenitors may have different critical values. One way to include possible variation in the core-halo mass relation is to build a large number of “merger trees” (Monte Carlo realizations of halo formation based on EPS) [68,77] which record the merger history of halos. Then each halo can be assigned a core mass from a distribution function or

following the core mass growth model proposed by one of us in Ref. [34].

Being Monte Carlo models, merger trees also allow for additional physics to be included within individual halos. In our case, we allow that the axion stars are removed from halos after they cross criticality, which circumvents a possible double counting in our calculation of the soliton merger rate from the halo merger rate using EPS. For example, when computing the integral Eq. (29), we consider only the cases with progenitor halo masses smaller than  $M_{h,\text{crit}}(z)$  [see Eqs. (30) and (31)]. However,  $M_{h,\text{crit}}(z)$  decreases with increasing redshift (increases with time), so there might be cases where the progenitor halo mass is above the critical value at an earlier time. Those cases should not be included in the integral Eq. (29).

Figure 9 shows the schematic diagram of a merger tree. Two halos,  $M_1$  and  $M_2$ , below  $M_{h,\text{crit}}$  merge and form a larger halo  $M_3 = M_1 + M_2 + \Delta M_{\text{acc}}$  whose mass is larger than  $M_{h,\text{crit}}$ . Here  $\Delta M_{\text{acc}}$  is the mass accreted from sub-resolution halos and the mass smoothly accreted from the environment (see Refs. [49,80] for more details). After the merger, we assume the central axion star in  $M_3$  completely decays and remove it from the halo. In reality, when two halos merge, it will take some time for their central axion stars to merge and form a new axion star, which may lead to delayed axion decays. The timescale for the merging process is discussed in Appendix C. We leave more detailed study of such a delay effect to further work. We also do not consider the reformation of axion star after the first

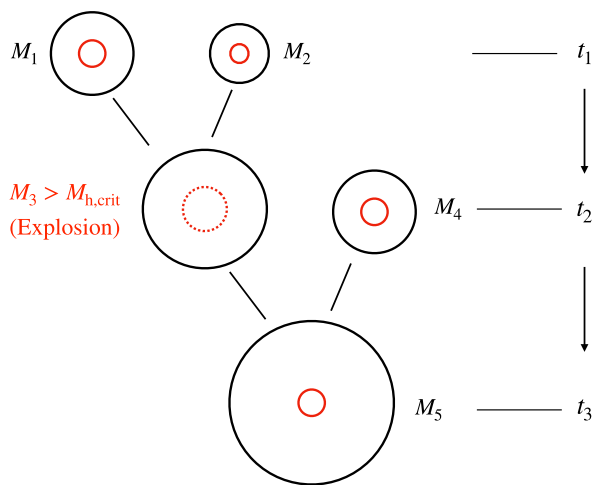


FIG. 9. A sketch of merger tree. At time  $t_1$ , two halos (black circles)  $M_1$  and  $M_2$  are below the critical halo mass. Each halo has an axion star in its center (red circles). At time  $t_2$ ,  $M_1$  and  $M_2$  merge and form a larger halo  $M_3 > M_{h,\text{crit}}$ . The axion star in the center of  $M_3$  (dashed red circle) becomes unstable and explodes. Thus we remove the central axion star from  $M_3$ , i.e. resetting the core mass to 0. At a later time  $t_3$ , halo  $M_3$  merges with another halo  $M_4$  (below the critical halo mass). If the central axion star in  $M_4$  is not disrupted during the merger, the final halo  $M_5$  will have a new core (central axion star).

explosion through gravitational relaxation since this process is slow [12,13] compared to mergers, but we do allow the halo to accrete a new axion star at a later time through major merger in some of the models we consider below. This essentially changes the core-halo mass relation above the critical halo mass.

We use the semianalytic code GALACTICUS [81] to generate realizations of merger trees. More details about the merger tree setup can be found in Appendix B. To cross-check with the calculations in the previous subsection, we first look at a fixed core-halo mass relation without any scatter. As mentioned previously, the core-halo mass relation is changed due to the decay of supercritical axion stars. So the core-halo mass relation is not applied to those halos that were ever above the critical halo mass in the past. We make a naive assumption that their core masses remain 0 since the first explosion and leave the study of other details such as the formation of a new core through dynamical relaxation to future work. The results of this calculation are shown in Fig. 10. We see good agreement between the analytic EPS model and the merger tree. This indicates that the calibration of the halo formation and destruction rates was performed correctly, and that there was no double counting of solitons caused by their non-removal after explosion in the EPS result.

To introduce scatter to the core-halo mass relation, we consider two models: (1) determining the core mass based on the merger histories as in Ref. [34]; (2) sampling the slope of the core-halo mass relation from a Gaussian distribution such that the core masses have a large scatter as found by Ref. [37]. In the first model, the core mass grows as halo mass increases. However, this is not guaranteed in the second model. So in the second model after we draw a core mass for the halo in the merger tree, we checked whether it is smaller than the core mass of the

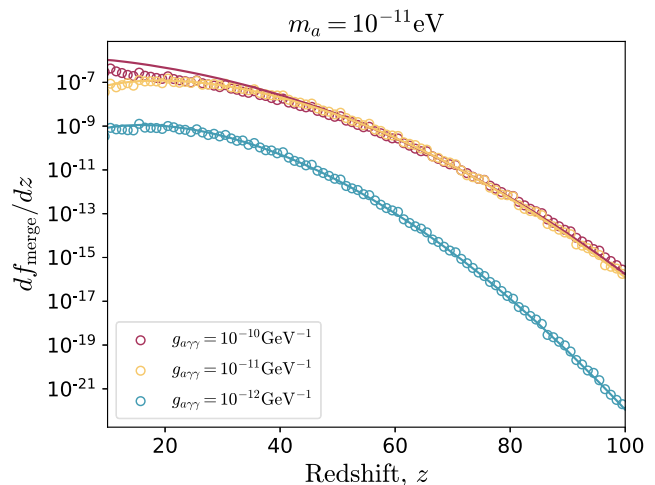


FIG. 10. Comparison of mean DM fractional decay rate from merger trees (empty circles) with that from EPS formalism (solid lines).

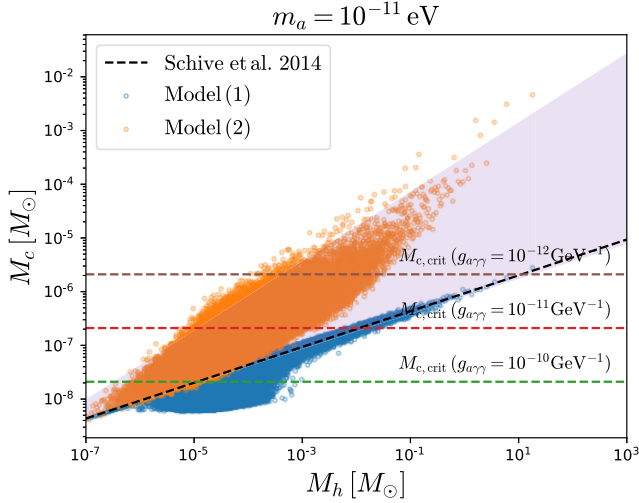


FIG. 11. Core mass versus halo mass at  $z = 10$  for  $m_a = 10^{-11}$  eV from different models: (1) Du *et al.* 2017 [34] (blue circles); (2) random sampling (orange circles). The shaded region shows the range of core masses found by [37]. The horizontal lines mark the critical core mass  $M_{c,crit} = M_{\text{decay}}/4$ .

halo's main progenitor. If it is, the halo is assigned the same core mass as its main progenitor. This additional constraint tends to make the slope of the core-halo mass relation steeper as mergers happen. To recover the core mass range found by Ref. [37], we set the mean and standard deviation of the slope parameter as  $\alpha_{\text{mean}} = 0.326^6$  and  $\sigma_\alpha = 0.1$ .

Figure 11 shows the core mass versus halo mass from the two models above. In this test, to compare with previous findings on the core-halo mass relation, we have not included the decay of super-critical axion stars. In model (1), the core mass grows only in a major merger event, so for small halos which have experienced only a few mergers the core masses have a large scatter, i.e. a halo growing through major mergers will have a large core mass than the one growing through minor mergers. As more mergers happen, the core-halo mass relation approaches a  $1/3$ -power law, which is consistent with that found by Ref. [27] in cosmological simulations, but the scatter in core masses is much smaller than the other model. As we expected, model (2) reproduces the dispersion in the core-halo mass relation reported in Ref. [37] (shaded region).

Then we rerun the above two models and allow the axion star to decay when it becomes supercritical. For model (2), as in the case with a fixed core-halo mass relation, we apply Eq. (4) only to halos that have never been above the critical halo mass. For model (1), we allow the halo to accrete a new core from major merger, but assume there is no adiabatic core growth.

<sup>6</sup>Note that the mean value required for the dynamical merger tree model to reproduce the scatter is smaller than the measured mean value at fixed  $z$ , 0.515, reported in Ref. [37].

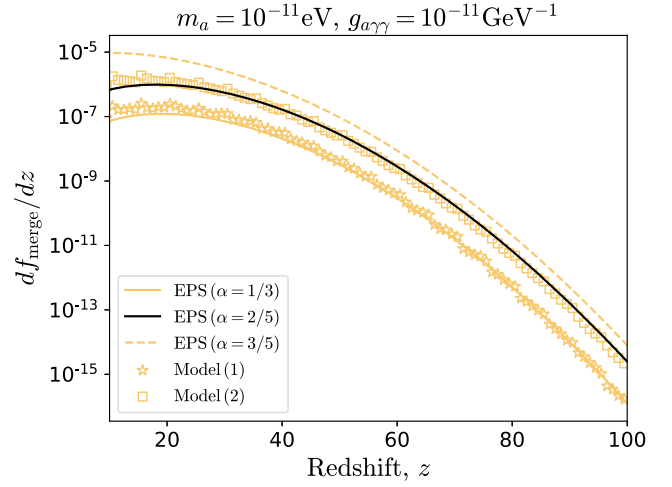


FIG. 12. DM fractional decay rate from merger trees assuming different core-halo mass relation models comparing with the EPS models assuming a fixed core-halo mass relation with different values of  $\alpha$ .

In Fig. 12, we compare the DM fractional decay rate predicted by the models above with those assuming a fixed core-halo mass relation. We found that the results from model (1) agree well with a fixed core-halo mass relation with  $\alpha = 1/3$ , while the results from model (2) lies between two limiting cases with  $\alpha = 1/3$  and  $\alpha = 3/5$ , corresponding to an effective average value  $\alpha \approx 2/5$ .

## V. SUMMARY AND CONCLUSION

For DM composed of a bosonic field, numerical work in the last decade has shown that a general prediction of such a model is the formation of a soliton in every dark matter halo. Numerical and semianalytic models can be used to compute the soliton mass function and merger rate. In the present work, we used the extended Press-Schechter formalism to first write down the soliton mass function assuming a core-halo mass relation with the host DM halos. Figure 2 shows the resulting number density of solitons predicted in a standard cosmology.

Using this result as a baseline, we then presented a new calculation of the soliton formation rate and merger rate, culminating in the double-differential formation rate, Eq. (28). Solitons form and grow primarily due to mergers, and the formation rate can be used to calculate the rate of *major mergers* of solitons of any given mass, Eq. (32), and Fig. 6.

Due to nonlinear interactions, and including relativistic corrections, bosonic DM solitons are unstable above a critical mass. For gravity, this instability leads to BH formation, but is predicted to occur only in very massive DM halos. For quartic self-interactions, and for axionlike interactions with photons, the instability leads to soliton explosions, and decay of DM into relativistic degrees of freedom. Soliton explosions caused by major mergers

producing heavy and unstable solitons thus leads to an enhanced DM decay rate.

We write the decay rate as  $df/dz$ , where  $f$  is the fraction of the total DM density, and  $z$  is redshift. Assuming a given core-halo mass relation, our computation of this rate can be approximated from the halo mass function and is given in Eq. (34), which applies for any instability to decay at a mass  $M_{\text{decay}}$ , regardless of the specific mechanism. Taking the axion-like instability to production of radio photons [22], we used our computation of the major merger rate to compute the DM decay rate for different values of the axion mass,  $m_a$ , and coupling constant,  $g_{\text{arr}}$ , as shown in Fig. 7.

The analytical EPS model can only be used to compute the DM decay rate if there exists an exact analytical core-halo mass relation. Given that this is thought to be a statistical phenomenon, with scatter due to halo formation histories and departures from equilibrium, we extended our computation to semianalytical Monte Carlo methods using the GALACTICUS code [81]. Firstly, we used our Monte Carlo model to validate the EPS calculation of the merger rate for a fixed core-halo mass relation. Next, we introduced scatter to the core halo mass relation tuned to match results of numerical studies [37], as shown in Fig. 11. We found that the results of such a model can be matched by the EPS model using a mean value of the core-halo mass relation slope parameter  $\alpha$ , as shown in Fig. 12.

The enhanced decay of axion DM induced by soliton mergers may have phenomenological consequences that either offer new windows to axion indirect detection, or place stronger constraints on the axion parameter space. Soliton decay injects energy into the intergalactic medium, which, for the reference parameters shown in Fig. 1, is significantly larger than the energy injection due to core collapse supernovae, and extends to much higher redshifts. We explore phenomenology of this energy injection in detail in a preceding paper [29].

## ACKNOWLEDGMENTS

D. J. E. M. is supported by an Ernest Rutherford Fellowship from the Science and Technologies Facilities Council (NO. ST/T004037/1). D. J. E. M. thanks Tilman Hartwig and Mattis Magg for useful correspondence on supernova energy injection. D. B. is supported by a ‘‘Ayuda Beatriz Galindo Senior’’ from the Spanish ‘‘Ministerio de Universidades,’’ Grant No. BG20/00228. M. F. and C. K. P. are funded by the Science and Technology Facilities Council (UK). The research leading to these results has received funding from the Spanish Ministry of Science and Innovation (No. PID2020–115845 GB-I00/AEI/10.13039/501100011033). I. F. A. E. is partially funded by the Centres de Recerca de Catalunya (CERCA) program of the Generalitat de Catalunya. D. B. acknowledges the support from the Departament de Recerca i Universitats de la Generalitat de Catalunya al Grup de Recerca i

Universitat de Catalunya to the Grup de Recerca 00649 (Codi: 2021 SGR 00649). Computing resources used in this work were made available by a generous grant from the Ahmanson Foundation.

## APPENDIX A: DETAILS OF THE CALCULATION

Our calculation of the soliton merger rate at very high redshift and at very low halo mass requires the computation of many nested integrals covering a wide range of scales. We describe briefly here how this calculation is done in a numerically efficient and physically accurate manner.

We used CLASS [82] to compute the matter power spectrum,  $P(k)$ , in a standard CDM cosmology. We then applied the analytic approximation for the effect of the Jeans scale cut-off given by Ref. [42]. We chose this methodology, rather than direct computation of the Jeans scale using AXIONCAMB [83] because: (a) we use CLASS in Ref. [29] to compute the reionization effect of soliton decays, and (b) the analytic approximation is simpler to implement at larger particle mass  $m_a$  and wave number  $k$  of interest at present.<sup>7</sup>

We pre-compute  $\sigma(M)$  at  $z = 0$  at fixed particle masses,  $m_a$  and fixed cosmological parameters,  $\Omega_m = 0.3153$ ,  $\Omega_b = 0.04930$ ,  $\Omega_\Lambda = 0.6847$ ,  $\sigma_8 = 0.8111$ ,  $n_s = 0.9649$ ,  $h = 0.6736$  [30]. We first compute  $\sigma(R)$  using three different window functions,  $\tilde{W}(k|R)$  discussed in the text: (a) real space top-hat, (b) sharp- $k$  space, (c) smooth- $k$  space window function [50,51]. These are given by:

$$\tilde{W}^{\text{top-hat}}(k|R) = \frac{3[\sin(kR) - kR \cos(kR)]}{(kR)^3}, \quad (\text{A1})$$

$$\tilde{W}^{\text{sharp-k}}(k|R) = \theta(1 - kR), \quad (\text{A2})$$

$$\tilde{W}^{\text{smooth-k}}(k|R) = \frac{1}{1 + (kR)^\beta}. \quad (\text{A3})$$

Here  $\theta$  is the Heaviside step function. The variance is defined by:

$$\sigma^2(R) = \int_0^\infty \frac{k^2}{2\pi^2} P(k) \tilde{W}(k|R)^2 dk. \quad (\text{A4})$$

In the calculation of the soliton merger rate using the EPS formalism in Sec. IV A, we adopt the smooth- $k$  space window function, Eq. (A3) as a reference, with mass assignment Eq. (A7). This is because this mass variance reproduces the HMF cutoff seen in simulations very well directly from  $\sigma(M)$ , without the need for additional fitting functions. This  $\sigma(M)$  can then be used as direct input in the merger rate calculation.

<sup>7</sup>See Ref. [84] for discussion of the accuracy of these various methods for computing the Jeans scale cutoff.

To find the mass variance, we need to assign mass to a scale ( $R$ ). This is trivial for the real-space top hat, but requires calibration for both sharp- $k$  and smooth- $k$  space window functions. We use the following mass assignments:

$$M^{\text{top-hat}}(R) = \frac{4}{3}\pi R^3 \rho_m, \quad (\text{A5})$$

$$M^{\text{sharp-k}}(R) = \frac{4}{3}\pi (a_W R)^3 \rho_m, \quad (\text{A6})$$

$$M^{\text{smooth-k}}(R) = \frac{4}{3}\pi (c_W R)^3 \rho_m. \quad (\text{A7})$$

For the sharp- $k$  filter we take  $a_W = 2.5$  as in [49]. For the smooth- $k$  filter, we take  $c_W = 2.15940$  and  $\beta = 9.10049$  which were found by fitting large  $N$ -body simulations [85].

A converged computation of  $\sigma(M)$  at the required low values of  $M$  requires computing  $P(k)$  to large  $k$ . In CLASS we set  $k_{\text{max}} = 2.5 \times 10^4 \text{ Mpc}^{-1}$ , and extrapolate  $P(k)$  at  $k > k_{\text{max}}$  using a power law. The large value of  $k_{\text{max}}$  required might suggest that an analytic approximation to  $P(k)$  would be useful. However, the best such approximation, Ref. [86] is not accurate on  $k$  larger than the baryon Jeans scale, and gives  $P(k)$  which is too large compared to the one computed accurately by CLASS direct solution of the Boltzmann equation.

For cosmological parameters adopted in this paper, we found the following fitting functions for  $\sigma(M)$  for both CDM and axion DM models:

$$\sigma^{\text{CDM}}(M) = a \left( \frac{M}{10^3 M_\odot} \right)^{-b} \left[ 1 - c \ln \left( 1 + d \sqrt{\frac{M}{10^3 M_\odot}} \right) \right], \quad (\text{A8})$$

$$\sigma^{\text{axion}}(M) = \left[ 1 + \left( \frac{M}{e M_{\text{half}}} \right)^f \right]^{b/f} \sigma^{\text{CDM}}(M), \quad (\text{A9})$$

where  $M_{\text{half}}$  is the half-mode mass

$$M_{\text{half}} = \frac{4}{3}\pi \left( \frac{\pi}{k_{\text{half}}} \right)^3 \rho_m, \quad (\text{A10})$$

$$k_{\text{half}} = 4.986 \times 10^4 \left( \frac{m_a}{10^{-13} \text{ eV}} \right)^{4/9} \text{ Mpc}^{-1}, \quad (\text{A11})$$

and the best-fit parameters take

$$a = 11.2934, \quad (\text{A12})$$

$$b = 0.0231895, \quad (\text{A13})$$

$$c = 0.0800510, \quad (\text{A14})$$

$$d = 0.125902, \quad (\text{A15})$$

$$e = 0.329159, \quad (\text{A16})$$

$$f = -2.41133. \quad (\text{A17})$$

Note that the fitting functions above are only fitted to halo masses smaller than  $10^{12} M_\odot$  and axion masses larger than  $10^{-17} \text{ eV}$ . For those cases, they provide an accuracy better than 1%. To compute the halo mass function and halo formation rate, one also need to compute  $\alpha \equiv d \ln \sigma / d \ln M$ .<sup>8</sup> This can easily be done using the fitting functions Eqs. (A8) and (A9). But we found that directly computing  $\alpha$  for axion DM models using Eq. (A9) leads to large errors for  $M \ll M_{\text{half}}$ . So we provide an additional fitting function for  $\alpha^{\text{axion}}$  below.

$$\alpha^{\text{axion}}(M) = \left[ 1 + \left( \frac{M}{e M_{\text{half}}} \right)^f \right]^{g/f} \alpha^{\text{CDM}}(M). \quad (\text{A18})$$

Here  $\alpha^{\text{CDM}}(M) \equiv d \ln \sigma^{\text{CDM}} / d \ln M$  is computed using Eq. (A8) and  $g = 2.86571$ .

With the above fitting functions, the halo mass function can be computed as

$$\frac{dn_h}{dM} = 2 \frac{\rho_m}{M} \frac{df^{\text{ST}}(S(M), z)}{dS} \frac{\sigma(M)^2}{M} |\alpha(M)|, \quad (\text{A19})$$

where  $df^{\text{ST}}(S, z)/dS$  is given by Eq. (22).

## APPENDIX B: GENERATING MERGER TREES

The merger trees are generated using the semi-analytic code GALACTICUS [81] which employs the EPS Monte Carlo algorithm proposed in Ref. [77] (see also Ref. [87]). To compute the formation rate of axion stars with critical masses at different redshifts, we start building the merger from a final redshift of  $z_f$ . As we go backward in time along the merger trees, halos fragment into smaller and smaller progenitor halos. After significant fragmentation has occurred the number of halos with masses in the range we are interested in becomes too small, so we only use the data in the redshift range  $[z_f, z_f + 5]$ . At this point a new set of trees is built starting from from  $z_f + 5$ . For each case, we generate 400,000 trees with root halo masses sampled from the halo mass function in the mass range  $[10^{-7}, 10^3] M_\odot$ . The mass resolution of the tree is set to  $10^{-7} M_\odot$  to make sure that for the cases we shown in Sec. IV C, the critical halo mass is always resolved.

<sup>8</sup>It should not be confused with the slope parameter in the core-halo mass relation Eq. (4).

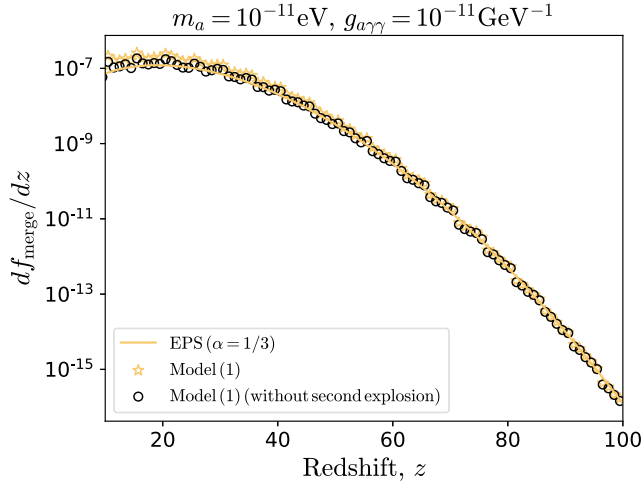


FIG. 13. DM fractional decay rate from merger trees assuming core growth model from Ref. [34]. If a second explosion is allowed, the result is slightly higher at lower redshifts.

For  $m_a = 10^{-11}$  eV,  $g_{a\gamma\gamma} = 10^{-12}$  GeV $^{-1}$ , and the core-halo mass relation from Ref. [11], the critical halo mass  $M_{h,\text{crit}}$  at  $z = 10$  ( $z = 100$ ) is  $11.4M_\odot$  ( $0.4M_\odot$ ). If we sample the root halo masses from the halo mass function, we will have too few halos close to  $M_{h,\text{crit}}$  making the calculation of formation rate inaccurate. So for this case, we sample the root halo masses from a loguniform distribution in the mass range  $[10^{-2}, 10^3]M_\odot$  instead. To get the correct formation rate, each tree is assigned a weight

$$w_i = \int_{M_{i-1/2}}^{M_{i+1/2}} \frac{dn}{dM} dM, \quad (\text{B1})$$

where  $\{M_i\}$  is in ascending order and  $M_{i-1/2} = \sqrt{M_i M_{i-1}}$ .

Having the merger trees, we can assign each halo in the tree a core mass  $M_c$  (the axion star mass  $M_S \approx 4M_c$ ) using either a chosen core-halo mass relation or the model proposed in Ref. [34] (see Sec. IV C). For the latter case, we include some additional physics: (1) when the axion star in the center of a halo is above the critical star mass  $M_{S,\text{crit}}$ , we set  $M_c = 0$ ; (2) if a halo, with mass above  $M_{h,\text{crit}}$ , and whose central axion star has completely decayed, has another major merger with a halo below  $M_{h,\text{crit}}$ , the axion star from the second progenitor will remain in the new halo, i.e. it is not disrupted (see Fig. 9). The major merger mass ratio is defined as 1:13.<sup>9</sup> With the second assumption, a halo can have a second explosion of the central axion star, which leads to slightly higher dark matter decay rate at lower redshifts, see Fig. 13.

<sup>9</sup>The result is only slightly affected by the specific choice of this ratio.

## APPENDIX C: TIMESCALE OF MERGING PROCESS

In Sec. IV, when we compute the decay rate of axion stars due to major mergers, we have assumed that the axion stars in the center of halos merge at the same time when two halos merge. In reality, when two halos encounter each other, the smaller one will become the subhalo of the larger one (the host). The orbit of the subhalo decays due to dynamical friction [88]. As its orbit decays with time, the subhalo gradually sinks into the host's center. The central axion star of the subhalo finally merges with the one residing in the host's center if it is not disrupted by tidal forces. The dynamical friction is stronger for more massive subhalos, thus the timescale of the merging process,  $\tau_{\text{merging}}$ , is shorter for larger mass ratios  $\mu_h = M_{\text{sub}}/M_{\text{host}}$ .

To obtain the timescale  $\tau_{\text{merging}}$ , we simulate the merger of two halos with different mass ratios using the  $N$ -body code GADGET4 [89]. In this work, we are only concerned with major mergers, i.e.,  $\mu_h \geq (3/7)^{1/\alpha}$ , which lead to the growth of axion stars (for more details, see the discussions in Sec. IV B).

As an example, we consider a host mass of  $10^{-3}M_\odot$  which corresponds to a halo containing an axion star close to the critical mass at  $z = 20$  for  $m_a = 10^{-11}$  eV,  $g_{a\gamma\gamma} = 10^{-11}$  GeV $^{-1}$ , and  $\alpha = 1/3$  (assuming a core-halo mass relation found in Ref. [27]). The subhalo first enters the host's virial radius,  $R_{\text{vir,host}}$ , at redshift  $z_{\text{infall}} = 20$ , at which redshift both halos are assumed to have Navarro-Frenk-White (NFW) profiles [90]. The concentration parameters of the NFW profiles are computed using the model proposed in Ref. [91].<sup>10</sup> We start the simulation at  $z_{\text{infall}}$  and consider two types of initial orbits: (1) the subhalo is initially on a circular orbit with a zero radial velocity and a tangential velocity computed as  $V_t = \sqrt{GM_{\text{host}}/R_{\text{vir,host}}(1 + \mu_h)}$ ; (2) the subhalo has typical radial and tangential velocities found in cosmological simulations [93].

Figure 14 shows the distance between the subhalo and the host as a function of time since infall obtained from our simulations. The results for type (1) and type (2) initial orbits are shown in the left and right panels, respectively. Compared to type (1) orbits, subhalos with type (2) orbits can reach further inside the host where the host density is higher, thus are subjected to stronger dynamical friction and have smaller  $\tau_{\text{merging}}$ . For  $\mu_h = 0.1$ , the orbit of the subhalo decays rapidly at early times, but then the subhalo remains on a nearly stable orbit when the dynamical friction becomes negligible. Instead of sinking into the host's center, the subhalo gradually loses its mass and may

<sup>10</sup>The model in Ref. [91] is shown to work reasonably well for the halo mass we consider here. At  $z = 0$  the concentration predicted by this model for a halo of mass  $10^{-3}M_\odot$  is only slightly higher than that found in numerical simulations [92].

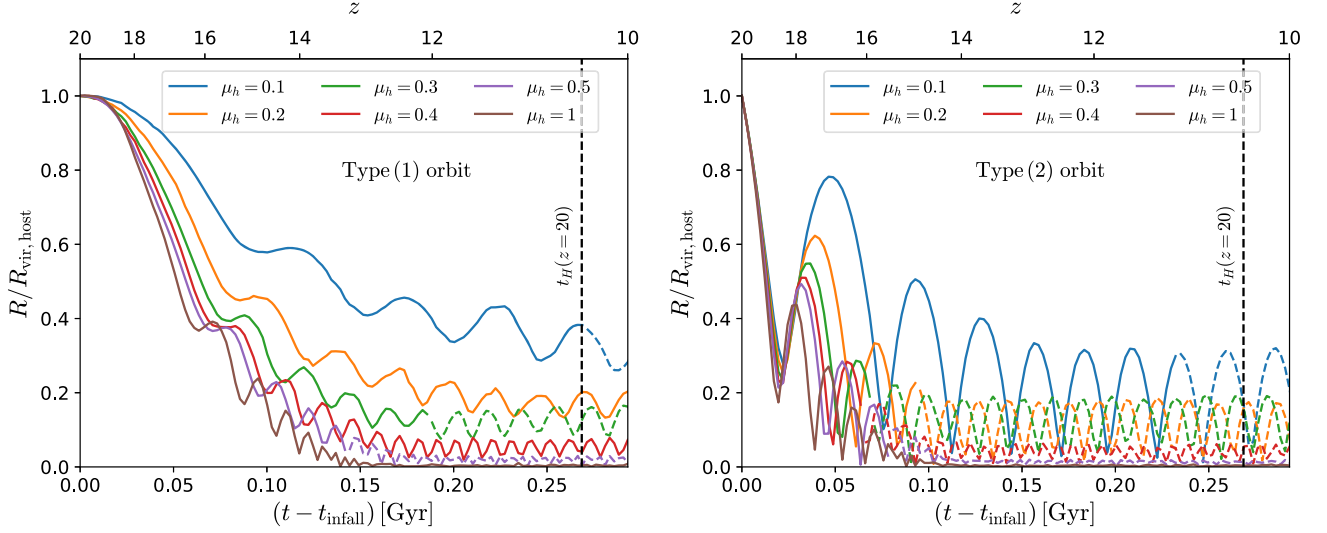


FIG. 14. Orbital decays caused by dynamical friction for halos mergers with different mass ratios. The host has a mass of  $10^{-3}M_{\odot}$ . The subhalo falls into the host's virial radius,  $R_{\text{vir,host}}$  at  $z_{\text{infall}} = 20$ . Left: type (1) orbit, the subhalo is initially on a circular orbit. Right: type (2) orbit, the subhalo is initially on an elliptical orbit with the radial and tangential velocity taken for typical infalling subhalos [93]. When the subhalo contains less than 20 bound particles, its position tracked by the most-bound particle is show as dashed curve.

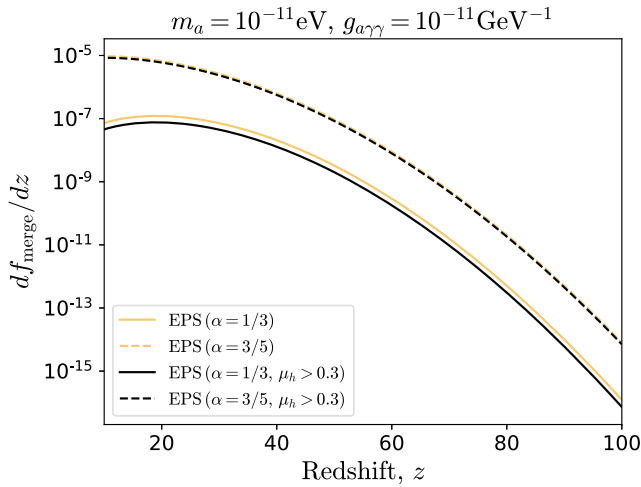


FIG. 15. DM fractional decay rate due to soliton mergers with  $\mu_h > 0.3$  (black curves) comparing with our fiducial EPS models (colored curves).

finally be destroyed by the tidal forces from the host. When the number of particles gravitationally bound to the subhalo drops below 20, we track the position of the subhalo by the most-bound particle, the particle that has the most negative potential energy (identified in the last snapshot at which the subhalo contained more than 20 bound particles). The orbits of the most-bound particle are shown as dashed curves. For  $\mu_h \gtrsim 0.3$ , the subhalo is able to reach close enough to the host's center so that the axion star in the subhalo can merge with that in the host.

The  $N$ -body simulations we perform cannot describe the axion stars in the halos and their mergers. Thus we only obtain a rough estimate of  $\tau_{\text{merging}}$  from Fig. 14. While type (2) orbits are better representations of realistic subhalo orbits, we take  $\tau_{\text{merging}}$  from the type (1) simulations as a conservative estimate. For  $\mu_h \gtrsim 0.3$ , we find  $\tau_{\text{merging}} \sim 0.16$  Gyr. After the merger of two axion stars, if the new axion star has a mass above the critical value given by Eq. (11), the axion particles in the axion star will decay to photons exponentially within a timescale negligible compared to  $\tau_{\text{merging}}$ . So we expect the axion decays are delayed by  $\tau_{\text{merging}}$  compared to the time that we calculate in Sec. IV. Note that  $\tau_{\text{merging}}$  is smaller than the Hubble time at  $z_{\text{infall}}$  (vertical dashed curve in Fig. 14), so the effect of delay is expected to be minor. Simulations that are capable of correctly including the quantum pressure of the scalar field are needed to study the merging process more accurately. We leave such a detailed and statistical analysis to further work (although see Ref. [94]).

As shown above, not all major mergers as we defined in Sec. IV, i.e.,  $\mu_h \geq (3/7)^{1/\alpha}$ , result in the merger of the central axion stars in halos. To check how this may affect the axion decay rate, we impose an additional condition,  $\mu_h \geq 0.3$ , when computing the integral Eq. (29). Figure 15 shows the DM fractional decay rate due to mergers if we impose this additional condition comparing with that from our fiducial EPS models. For  $\alpha = 1/3$  ( $\alpha = 3/5$ ), the DM fractional decay rate is reduced by 40% (10%), a relatively small effect on the scales shown that does not affect our main conclusions.



- [1] E. Seidel and W. M. Suen, *Phys. Rev. Lett.* **66**, 1659 (1991).
- [2] E. Seidel and W.-M. Suen, *Phys. Rev. Lett.* **72**, 2516 (1994).
- [3] F. S. Guzman and L. A. Urena-Lopez, *Astrophys. J.* **645**, 814 (2006).
- [4] P.-H. Chavanis, *Phys. Rev. D* **84**, 043531 (2011).
- [5] D. J. E. Marsh, *Phys. Rep.* **643**, 1 (2016).
- [6] L. Hui, J. P. Ostriker, S. Tremaine, and E. Witten, *Phys. Rev. D* **95**, 043541 (2017).
- [7] M. Jain and M. A. Amin, *Phys. Rev. D* **105**, 056019 (2022).
- [8] M. Gorghetto, E. Hardy, J. March-Russell, N. Song, and S. M. West, *J. Cosmol. Astropart. Phys.* 08 (2022) 018.
- [9] N. Aghanim *et al.* (Planck Collaboration), *Astron. Astrophys.* **641**, A6 (2020); *Astron. Astrophys.* **652**, C4(E) (2021).
- [10] B. Eggemeier and J. C. Niemeyer, *Phys. Rev. D* **100**, 063528 (2019).
- [11] H.-Y. Schive, T. Chiueh, and T. Broadhurst, *Nat. Phys.* **10**, 496 (2014).
- [12] D. Levkov, A. Panin, and I. Tkachev, *Phys. Rev. Lett.* **121**, 151301 (2018).
- [13] J. Chen, X. Du, E. W. Lentz, D. J. E. Marsh, and J. C. Niemeyer, *Phys. Rev. D* **104**, 083022 (2021).
- [14] D. J. E. Marsh and A.-R. Pop, *Mon. Not. R. Astron. Soc.* **451**, 2479 (2015).
- [15] A. X. González-Morales, D. J. E. Marsh, J. Peñarrubia, and L. A. Ureña López, *Mon. Not. R. Astron. Soc.* **472**, 1346 (2017).
- [16] N. Bar, D. Blas, K. Blum, and S. Sibiryakov, *Phys. Rev. D* **98**, 083027 (2018).
- [17] D. Dutta Chowdhury, F. C. van den Bosch, V. H. Robles, P. van Dokkum, H.-Y. Schive, T. Chiueh, and T. Broadhurst, *Astrophys. J.* **916**, 27 (2021).
- [18] G. F. Giudice, M. McCullough, and A. Urbano, *J. Cosmol. Astropart. Phys.* 10 (2016) 001.
- [19] T. Helfer, D. J. E. Marsh, K. Clough, M. Fairbairn, E. A. Lim, and R. Becerril, *J. Cosmol. Astropart. Phys.* 03 (2017) 055.
- [20] J. Y. Widdicombe, T. Helfer, D. J. E. Marsh, and E. A. Lim, *J. Cosmol. Astropart. Phys.* 10 (2018) 005.
- [21] D. G. Levkov, A. G. Panin, and I. I. Tkachev, *Phys. Rev. Lett.* **118**, 011301 (2017).
- [22] D. G. Levkov, A. G. Panin, and I. I. Tkachev, *Phys. Rev. D* **102**, 023501 (2020).
- [23] H.-Y. Schive, T. Chiueh, T. Broadhurst, and K.-W. Huang, *Astrophys. J.* **818**, 89 (2016).
- [24] A. M. Green, S. Hofmann, and D. J. Schwarz, *J. Cosmol. Astropart. Phys.* 08 (2005) 003.
- [25] A. Loeb and M. Zaldarriaga, *Phys. Rev. D* **71**, 103520 (2005).
- [26] R. E. Angulo, O. Hahn, A. Ludlow, and S. Bonoli, *Mon. Not. R. Astron. Soc.* **471**, 4687 (2017).
- [27] H.-Y. Schive, M.-H. Liao, T.-P. Woo, S.-K. Wong, T. Chiueh, T. Broadhurst, and W.-Y. P. Hwang, *Phys. Rev. Lett.* **113**, 261302 (2014).
- [28] T. Hartwig *et al.*, *Astrophys. J.* **936**, 45 (2022).
- [29] M. Escudero, C. K. Pooni, M. Fairbairn, D. Blas, X. Du, and D. J. E. Marsh, preceding paper, *Phys. Rev. D* **109**, 043018 (2024).
- [30] Y. Akrami *et al.* (Planck Collaboration), *Astron. Astrophys.* **641**, A10 (2020).
- [31] G. L. Bryan and M. L. Norman, *Astrophys. J.* **495**, 80 (1998).
- [32] J. C. Niemeyer, *Prog. Part. Nucl. Phys.* **113**, 103787 (2020).
- [33] B. Eggemeier, B. Schwabe, J. C. Niemeyer, and R. Easther, *Phys. Rev. D* **105**, 023516 (2022).
- [34] X. Du, C. Behrens, J. C. Niemeyer, and B. Schwabe, *Phys. Rev. D* **95**, 043519 (2017).
- [35] J. H.-H. Chan, S. Sibiryakov, and W. Xue, arXiv:2207.04057.
- [36] M. Nori and M. Baldi, *Mon. Not. R. Astron. Soc.* **501**, 1539 (2021).
- [37] H. Y. J. Chan, E. G. M. Ferreira, S. May, K. Hayashi, and M. Chiba, *Mon. Not. R. Astron. Soc.* **511**, 943 (2022).
- [38] J. L. Zagorac, E. Kendall, N. Padmanabhan, and R. Easther, *Phys. Rev. D* **107**, 083513 (2023).
- [39] A. Taruya and S. Saga, *Phys. Rev. D* **106**, 103532 (2022).
- [40] R. K. Sheth and G. Tormen, *Mon. Not. R. Astron. Soc.* **329**, 61 (2002).
- [41] M. Khlopov, B. A. Malomed, and I. B. Zeldovich, *Mon. Not. R. Astron. Soc.* **215**, 575 (1985).
- [42] W. Hu, R. Barkana, and A. Gruzinov, *Phys. Rev. Lett.* **85**, 1158 (2000).
- [43] D. J. E. Marsh and J. Silk, *Mon. Not. R. Astron. Soc.* **437**, 2652 (2014).
- [44] P. S. Corasaniti, S. Agarwal, D. J. E. Marsh, and S. Das, *Phys. Rev. D* **95**, 083512 (2017).
- [45] D. Ellis, D. J. E. Marsh, B. Eggemeier, J. Niemeyer, J. Redondo, and K. Dolag, *Phys. Rev. D* **106**, 103514 (2022).
- [46] K.-H. Leong, H.-Y. Schive, U.-H. Zhang, and T. Chiueh, *Mon. Not. R. Astron. Soc.* **484**, 4273 (2019).
- [47] A. Schneider, R. E. Smith, and D. Reed, *Mon. Not. R. Astron. Soc.* **433**, 1573 (2013).
- [48] J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser, *Astrophys. J.* **379**, 440 (1991).
- [49] A. J. Benson, A. Farahi, S. Cole, L. A. Moustakas, A. Jenkins, M. Lovell, R. Kennedy, J. Helly, and C. Frenk, *Mon. Not. R. Astron. Soc.* **428**, 1774 (2013).
- [50] M. Leo, C. M. Baugh, B. Li, and S. Pascoli, *J. Cosmol. Astropart. Phys.* 04 (2018) 010.
- [51] S. Bohr, J. Zavala, F.-Y. Cyr-Racine, and M. Vogelsberger, *Mon. Not. R. Astron. Soc.* **506**, 128 (2021).
- [52] D. J. Kaup, *Phys. Rev.* **172**, 1331 (1968).
- [53] T. Helfer, E. A. Lim, M. A. G. Garcia, and M. A. Amin, *Phys. Rev. D* **99**, 044046 (2019).
- [54] J. Y. Widdicombe, T. Helfer, and E. A. Lim, *J. Cosmol. Astropart. Phys.* 01 (2020) 027.
- [55] V. De Luca, G. Franciolini, A. Kehagias, P. Pani, and A. Riotto, *Phys. Lett. B* **832**, 137265 (2022).
- [56] F. Michel and I. G. Moss, *Phys. Lett. B* **785**, 9 (2018).
- [57] M. P. Hertzberg and E. D. Schiappacasse, *J. Cosmol. Astropart. Phys.* 11 (2018) 004.
- [58] L. M. Chung-Jukko, E. A. Lim, D. J. E. Marsh, J. C. Aurrekoetxea, E. de Jong, and B.-X. Ge, *Phys. Rev. D* **108**, L061302 (2023).
- [59] R. Daido, T. Kobayashi, and F. Takahashi, *Phys. Lett. B* **765**, 293 (2017).
- [60] R. Daido, F. Takahashi, and N. Yokozaki, *Phys. Lett. B* **780**, 538 (2018).
- [61] A. V. Sokolov and A. Ringwald, arXiv:2205.02605.

- [62] A. S. Dmitriev, D. G. Levkov, A. G. Panin, and I. I. Tkachev, [arXiv:2305.01005](https://arxiv.org/abs/2305.01005).
- [63] M. P. Hertzberg, Y. Li, and E. D. Schiappacasse, *J. Cosmol. Astropart. Phys.* **07** (2020) 067.
- [64] R. Adam *et al.* (Planck Collaboration), *Astron. Astrophys.* **596**, A108 (2016).
- [65] A. Hook, Y. Kahn, B. R. Safdi, and Z. Sun, *Phys. Rev. Lett.* **121**, 241102 (2018).
- [66] S. J. Witte, D. Noordhuis, T. D. P. Edwards, and C. Weniger, *Phys. Rev. D* **104**, 103030 (2021).
- [67] A. Mirizzi, J. Redondo, and G. Sigl, *J. Cosmol. Astropart. Phys.* **03** (2009) 026.
- [68] C. G. Lacey and S. Cole, *Mon. Not. R. Astron. Soc.* **262**, 627 (1993).
- [69] C. G. Lacey and S. Cole, *Mon. Not. R. Astron. Soc.* **271**, 676 (1994).
- [70] S. Sasaki, *Publ. Astron. Soc. Jpn.* **46**, 427 (1994), <https://ui.adsabs.harvard.edu/abs/1994PASJ...46..427S>.
- [71] S. Mitra, G. Kulkarni, J. S. Bagla, and J. K. Yadav, *Bull. Astr. Soc. India* **39**, 563 (2011), <https://ui.adsabs.harvard.edu/abs/2011BASI...39..563M/abstract>.
- [72] J. E. Gunn and J. R. Gott, *Astrophys. J.* **176**, 1 (1972).
- [73] T. Kitayama and Y. Suto, *Astrophys. J.* **469**, 480 (1996).
- [74] W. J. Percival, *Astron. Astrophys.* **443**, 819 (2005).
- [75] R. K. Sheth and G. Tormen, *Mon. Not. R. Astron. Soc.* **308**, 119 (1999).
- [76] R. K. Sheth, H. J. Mo, and G. Tormen, *Mon. Not. R. Astron. Soc.* **323**, 1 (2001).
- [77] H. Parkinson, S. Cole, and J. Helly, *Mon. Not. R. Astron. Soc.* **383**, 557 (2008).
- [78] B. Schwabe, J. C. Niemeyer, and J. F. Engels, *Phys. Rev. D* **94**, 043513 (2016).
- [79] V. Anastassopoulos *et al.* (CAST Collaboration), *Nat. Phys.* **13**, 584 (2017).
- [80] X. Du, C. Behrens, and J. C. Niemeyer, *Mon. Not. R. Astron. Soc.* **465**, 941 (2017).
- [81] A. J. Benson, *New Astron.* **17**, 175 (2012).
- [82] J. Lesgourgues, [arXiv:1104.2932](https://arxiv.org/abs/1104.2932).
- [83] R. Hlozek, D. Grin, D. J. E. Marsh, and P. G. Ferreira, *Phys. Rev. D* **91**, 103512 (2015).
- [84] S. Passaglia and W. Hu, *Phys. Rev. D* **105**, 123529 (2022).
- [85] A. Benson *et al.* (to be published).
- [86] D. J. Eisenstein and W. Hu, *Astrophys. J.* **496**, 605 (1998).
- [87] S. Cole, C. G. Lacey, C. M. Baugh, and C. S. Frenk, *Mon. Not. R. Astron. Soc.* **319**, 168 (2000).
- [88] S. Chandrasekhar, *Astrophys. J.* **97**, 255 (1943).
- [89] V. Springel, R. Pakmor, O. Zier, and M. Reinecke, *Mon. Not. R. Astron. Soc.* **506**, 2871 (2021).
- [90] J. F. Navarro, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **462**, 563 (1996).
- [91] B. Diemer and M. Joyce, *Astrophys. J.* **871**, 168 (2019).
- [92] J. Wang, S. Bose, C. S. Frenk, L. Gao, A. Jenkins, V. Springel, and S. D. M. White, *Nature (London)* **585**, 39 (2020).
- [93] L. Jiang, S. Cole, T. Sawala, and C. S. Frenk, *Mon. Not. R. Astron. Soc.* **448**, 1674 (2015).
- [94] X. Du, B. Schwabe, J. C. Niemeyer, and D. Bürger, *Phys. Rev. D* **97**, 063507 (2018).