

Supermassive black hole formation in the initial collapse of axion dark matter

Pierre Sikivie[✉] and Yuxin Zhao[✉]

Department of Physics, University of Florida, Gainesville, Florida 32611, USA



(Received 15 July 2024; accepted 30 January 2026; published 20 March 2026)

Axion dark matter thermalizes by gravitational self-interactions and forms a Bose-Einstein condensate. We show that the rethermalization of the axion fluid during the initial collapse of large scale overdensities near cosmic dawn transports angular momentum outward sufficiently fast that black holes form with masses ranging from approximately 10^5 to a few times $10^{10} M_\odot$. This conclusion holds for QCD axions and for axionlike particles of mass larger than 10^{-16} eV/ c^2 .

DOI: [10.1103/m461-xxqg](https://doi.org/10.1103/m461-xxqg)

It has long been established [1] that most galaxies have supermassive black holes at their centers with masses ranging from approximately $10^5 M_\odot$ to a few times $10^{10} M_\odot$. The event horizons of the black holes at the center of the Milky Way and the center of the large galaxy M87 have been imaged using very long baseline interferometry [2]. Active galactic nuclei (AGN) are understood to be powered by accretion onto supermassive black holes [3]. Recently, low frequency gravitational waves consistent with emission from supermassive black hole mergers have been detected by pulsar timing arrays [4].

How the supermassive black holes form has been an enduring puzzle. The main impediment to their formation is conservation of angular momentum. If, for example, a $10^8 M_\odot$ black hole condenses out of a region of density, say 10^{-23} gr/cc, which is a value typical of galactic disks, the material forming the black hole must shrink in all directions by eight orders of magnitude. Angular momentum conservation makes this difficult by introducing a distance of closest approach to the black hole. Angular momentum can be transported outward if the material falling toward the black hole has viscosity but in that case the material heats up and acquires pressure opposing its compression. See Ref. [5] for a review of the issues involved in supermassive black hole formation.

To sidestep the problems that arise when only conventional physics is involved, it has been proposed that the supermassive black holes form as a result of the gravothermal collapse of overdensities of dark matter with very strong self-interactions [6] or by accretion onto dark stars,

i.e., stars that are powered by dark matter annihilation [7]. In such scenarios, seed black holes form that have masses of order 10^5 – $10^6 M_\odot$. Larger supermassive black holes with mass 10^9 – $10^{10} M_\odot$ are then supposed to be the result of accretion onto, and mergers of, the seed black holes. These proposals have been challenged by the recent discovery, using the James Webb Space Telescope, of powerful AGN near cosmic dawn, i.e., at redshifts $z \sim 10$ [8]. There is very little time to grow the black holes powering the AGN observed at $z \sim 10$. Finally, one may contemplate the possibility that the supermassive black holes are primordial in nature, i.e., that they formed long before cosmic dawn. This proposal runs afoul of constraints from cosmic microwave background observations, although a recent paper [9] indicates how it may still be viable.

The purpose of our paper is to show that supermassive black holes form naturally near cosmic dawn if the dark matter is axions or axionlike particles. No additional assumptions are required. The crucial step is to recognize that cold dark matter axions thermalize by their gravitational self-interactions [10,11]. Indeed, as explained in more detail below, their density fluctuations are generically large ($\delta\rho \sim \rho$) and correlated over long distances. When the axions thermalize they form a Bose-Einstein condensate, meaning that most axions go to the lowest energy state available to them through the thermalizing interactions. When an axion overdensity collapses near cosmic dawn, the gravitational self-interactions among the axions produce a long range viscosity that causes outward transport of angular momentum. The heat produced in the axion case flows into the thermal distribution accompanying the condensate whereas the condensate, containing most of the axions, stays in the lowest energy particle state available. That state is one of rigid rotation where most of the angular momentum resides far from the central overdensity, which may then perhaps collapse into a black hole.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

First let us indicate how axions differ from other cold dark matter candidates, such as weakly interacting massive particles (WIMPs) and sterile neutrinos, which we will refer to collectively as “ordinary cold dark matter (CDM).” Before density perturbations go nonlinear and multistreaming begins, in particular during the recombination era, the state of ordinary CDM is customarily given by a mass density $\rho(\vec{x}, t)$ and a velocity field $\vec{v}(\vec{x}, t)$. In linear order of perturbation theory, axions behave in the same way as ordinary CDM on length scales longer than their Jeans length [12]. When the axion fluid has mass density $\rho(\vec{x}, t)$ and velocity field $\vec{v}(\vec{x}, t)$, all axions are in the particle state of wave function

$$\Psi(\vec{x}, t) = \sqrt{\frac{\rho(\vec{x}, t)}{Nm}} e^{i\beta(\vec{x}, t)}, \quad (1)$$

where m is the axion mass, N is the number of axions, and $\beta(\vec{x}, t)$ is such that

$$\vec{v}(\vec{x}, t) = \frac{\hbar}{m} \vec{\nabla} \beta(\vec{x}, t). \quad (2)$$

For both ordinary CDM and axions, the above description is exact only when the dark matter has zero velocity dispersion. In reality both axions and ordinary CDM have a finite velocity dispersion. In the case of ordinary CDM this distinction is irrelevant because ordinary CDM is non-degenerate. But it is important in the case of axions because CDM axions form a degenerate Bose gas [11]. WIMPs of mass 100 GeV have primordial velocity dispersion $\delta v_W \sim 10^{-12} c a(t_0)/a(t)$ and hence correlation length $\ell_W \sim 2 \mu\text{m} a(t)/a(t_0)$, where $a(t)$ is the scale factor, t is cosmic time, and t_0 is the present age of the universe. Sterile neutrinos with mass a few keV have primordial velocity dispersion $\delta v_\nu \sim 10^{-8} c a(t_0)/a(t)$ and hence $\ell_\nu \sim \text{cm} a(t)/a(t_0)$. The correlation lengths of ordinary CDM are far too short to be relevant in large scale structure formation. Axions are different.

The correlation length and velocity dispersion of CDM axions are set by the horizon when the axion mass turns on during the QCD phase transition,

$$\ell(t) \sim c t_1 \frac{a(t)}{a(t_1)}, \quad \delta v(t) \sim \frac{\hbar}{m c t_1} \frac{a(t_1)}{a(t)} \quad (3)$$

where $t_1 \simeq 4 \times 10^{-7} \text{s} (\frac{\mu\text{eV}}{mc^2})^{\frac{1}{3}}$ is the time when axion field oscillations begin [13]. The thermal relaxation rate of the axion fluid is of order [10,11]

$$\Gamma(t) \sim 4\pi G \rho(t) m \ell(t)^2 / \hbar. \quad (4)$$

This result can be understood by noting that the axion fluid has density fluctuations $\delta\rho = \rho$ correlated over distances of order ℓ ; see Appendix A. The density fluctuations source

gravitational fields $\delta g \sim 4\pi G \rho \ell$. Since the fluid momentum dispersion is $\delta p \sim \hbar/\ell$, the fluctuating gravitational forces $m\delta g$ modify the momentum distribution entirely in a time $\tau \sim \delta p / m\delta g$. The thermal relaxation rate Γ is the inverse of τ . The reader may be surprised that the relaxation rate given in Eq. (4) is of order G rather than of order G^2 . Reference [14] discusses in detail why the familiar formula for the relaxation rate, which scales as G^2 , does not apply in the situation discussed here. The order G^2 formula would overestimate the relaxation rate by many orders of magnitude. It is inapplicable in the situation of interest because it relies on Fermi’s Golden Rule, which assumes that each individual axion scattering happens separately from all the other axion scatterings, so that energy is conserved in each scattering separately. This assumption fails here because the scatterings happen so fast that they overlap in time. In spite of its unfamiliarity, we assume Eq. (4) is correct, both for the reasons given in Ref. [11] and because it follows, as explained above, from the fact that the axion fluid has density fluctuations $\delta\rho = \rho$ correlated over distances of order ℓ .

One may attempt a classical field theory description of the thermalization of the cold axion fluid. In such a description, all axions are in a single state whose wave function satisfies the Schrödinger-Poisson equations. The wave function is a superposition of many waves with random phases. In that description the axion fluid also has $\delta\rho \simeq \rho$ fluctuations correlated over distances of order the inverse of the wave vector dispersion, and the resulting fluctuations in the gravitational field also thermalize the axion fluid. The thermalization rate is the same rate, Eq. (4), as in the quantum field theory. However, the outcome of the thermalization process is different. The outcome of thermalization in classical field theory is energy equipartition among all the field modes. The outcome of thermalization in the quantum field theory is Bose-Einstein condensation. Bose-Einstein condensation plays an essential role in the process we describe because the lowest energy state for given angular momentum is one of rigid rotation where most of the angular momentum has moved toward the periphery.

$\Gamma(t)$ exceeds the Hubble rate $H(t)$ at some time well before matter-radiation equality [10]. At that time all conditions for Bose-Einstein condensation are satisfied, and almost all axions go to the lowest energy particle state available to them through the thermalizing interactions. The axion fluid correlation length consequently grows to be of order the horizon at the time. In linear order of perturbation theory, the axion condensate evolves as ordinary CDM. However, in higher orders and in particular when it acquires angular momentum by tidal torquing, the condensate can and will lower its energy by acquiring vorticity whereas ordinary CDM remains vorticity-free [15].

Consider a large overdensity of dark matter that dominates its environment and is about to collapse near

cosmic dawn. We ignore at first all density perturbations on scales smaller than that of the overdensity itself, effectively smoothing it. Its smoothed density field has the general form

$$\rho(\vec{r}, t) = \rho(\vec{0}, t) \left[1 - \left(\frac{x_1}{R_1(t)} \right)^2 - \left(\frac{x_2}{R_2(t)} \right)^2 - \left(\frac{x_3}{R_3(t)} \right)^2 \right], \quad (5)$$

where (x_1, x_2, x_3) are appropriately chosen Cartesian coordinates centered on the peak, and the R_i ($i = 1, 2, 3$) give the peak's extent in the three spatial directions. Reference [16] derived the properties of such peaks in Gaussian random fields, e.g., the number of peaks per unit volume of a given size, and probability distributions for R_i .

Figure 1 shows the evolution of such an overdensity in a two-dimensional cut $(x, \dot{x})$ of its six-dimensional phase space. x is the spatial coordinate along an arbitrary direction through the overdensity and $\dot{x}$ the corresponding velocity. Curve (b) shows the distribution of particles at time t_{in} defined as the time when the central part of the overdensity is at turnaround, i.e., when it stops expanding and is about to contract. The central density at that time $\rho(0, t_{\text{in}})$ is related to t_{in} by

$$t_{\text{in}} = \frac{\pi}{2} \sqrt{\frac{3}{8\pi G \rho(0, t_{\text{in}})}}. \quad (6)$$

The central part of the overdensity collapses at time $t_{\text{coll}} \simeq 2t_{\text{in}}$. Curve (d) shows the distribution of particles

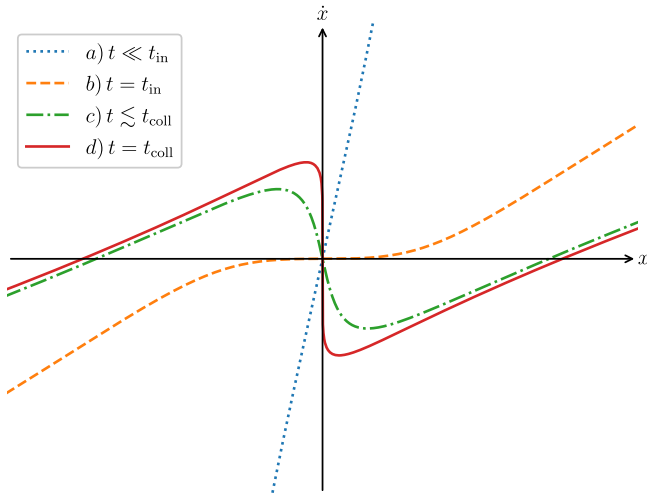


FIG. 1. Phase space distribution of cold collisionless particles during the collapse of a large smooth overdensity near cosmic dawn, at four different times: (a) just after the big bang, (b) when the central part of the overdensity is at turnaround, (c) just before, and (d) at the time t_{coll} of collapse of the central overdensity. x is the spatial coordinate along an arbitrary direction through the overdensity. An actual overdensity has small scale structure, which has been smoothed out in the figure.

at time t_{coll} . At collapse time the density is very large near the center but no black hole forms in case of ordinary CDM because the particles have acquired angular momentum through tidal torquing.

The amount of angular momentum acquired by galaxies through tidal torquing is commonly given by a dimensionless number λ called the galactic spin parameter [17]. Spin parameters in the range $0.01 \lesssim \lambda \lesssim 0.18$ are predicted [18] and found to be consistent with the amount of angular momentum baryons are observed to carry in galaxies [19]. We take this range to be a guide to the amount of angular momentum that the overdensity under consideration acquired through tidal torquing. If the dark matter is ordinary CDM, the angular momentum introduces an average distance of closest approach to the center of the overdensity of order $\lambda^2 R$, far too large for a black hole to form.

If axions are the dark matter, the axions at a distance r from the center of the overdensity thermalize at the rate given in Eq. (4) with $\ell \sim r$ since the gravitational fields due to the axion fluid outside the region of radius r do not help the thermalization within that region [11]. Since $\rho(\vec{r}, t)$ exceeds the average cosmological energy density $\bar{\rho}(t) = 1/6\pi G t^2$,

$$\Gamma(t) \gtrsim 2 \times 10^{16} H(t) \left(\frac{mc^2}{\mu\text{eV}} \right) \left(\frac{r}{10^{22} \text{ cm}} \right) \left(\frac{220 \text{ Myr}}{t} \right), \quad (7)$$

where $H(t) = \frac{2}{3t}$ near cosmic dawn, the thermalization rate is therefore very large compared to the dynamical evolution rate at time t_{in} . One may readily verify that it remains much larger than the dynamical evolution rate during the collapse. However, thermalization does not suffice to justify the angular momentum transport necessary for black hole formation. Since the axions can only change their momenta by an amount $\delta p \sim \hbar/\ell$ in a time $\tau \sim 1/\Gamma$, they can only change their specific angular momentum L , i.e., their angular momentum per unit mass, by an amount $\delta L \sim \hbar/m$ in that time. Hence,

$$\dot{L}_{\text{max}}(r, t) \sim 4\pi G \rho(r, t) r^2 \quad (8)$$

is the maximum rate at which axions at radius r can gain or lose specific angular momentum.

The initial overdensity is not spherically symmetric, and hence its collapse is not isotropic. Instead, its asphericity grows during the collapse [20] as the overdensity tends to become a pancake or spindle. In our treatment below, we ignore the fact that the infall is highly anisotropic because anisotropy only produces velocities in the angular (i.e., nonradial) directions that are at most of order $\sqrt{GM}/r$ where M is the mass enclosed by the shell of radius r . Since this is always much less than c during the infall, the angular velocities do not greatly affect whether a shell falls within the black hole horizon of the mass M it contains. Although we ignore the effect of anisotropy on the infall motion we

do not ignore its effect on the tidal torquing experienced by the infalling overdensity. Tidal torquing is important because it produces angular momentum and hence a minimum radius

$$r_{\min} = \frac{L^2}{2GM} + \mathcal{O}(L^4) \quad (9)$$

for each shell.

In case of isotropic radial infall the radius of the shell containing mass M at time t is given by the parametric equations

$$\begin{aligned} r(M, t) &= r_0(M) \sin^2 \sigma, \\ t &= \sqrt{\frac{r_0(M)^3}{2GM}} \left[\sigma - \frac{1}{2} \sin(2\sigma) \right], \end{aligned} \quad (10)$$

where $r_0(M)$ is the shell's turnaround radius. The shell's turnaround time is

$$t_0(M) = \frac{\pi}{2} \sqrt{\frac{r_0(M)^3}{2GM}}. \quad (11)$$

The axion mass density at shell M is

$$\rho(M, t) = \frac{1}{4\pi r(M, t)^2} \frac{dM}{dr}(M, t). \quad (12)$$

By using Eq. (10) to describe the collapse of an axion dark matter overdensity four approximations are made in addition to ignoring the fact that the infall is highly anisotropic. First, we ignore all forms of matter and energy (baryons, photons, neutrinos, dark energy) other than the axionic dark matter. Second, we use Newtonian gravity and Newton's laws of motion to describe the infall. Relativistic corrections are unimportant until a shell approaches the black hole horizon and are unlikely to change the final black hole mass by more than a factor two or so, which is within the uncertainty that we tolerate throughout. Third, although we keep track of the specific angular momentum $L(M, t)$ of each shell, we ignore its effect on the radial motion of the shell during its infall. This is justified because a shell can only fall into a black hole if its angular momentum is extremely small. Fourth, we ignore the corrections to the radial motion of the shells due to the wave nature of the axion fluid. This is equivalent to describing the infalling axion waves in the Wentzel-Kramers-Brillouin (or eikonal) approximation, and is tantamount to ignoring the role of so-called “quantum pressure.” It is justified to do so provided that the axion Compton wavelength is much less than the black hole size. For our smallest black holes, which have mass of order $10^6 M_\odot$, the axion mass must be more than of order 10^{-16} eV/ c^2 . The condition is amply satisfied by QCD axions.

Equation (10) describe the evolution of an overdensity entirely in terms of the function $r_0(M)$ that gives the turnaround radius of each shell. We defined our initial time t_{in} to be the turnaround time of the innermost shells in Eq. (6), which can be alternatively expressed as

$$t_{\text{in}} = \lim_{M \rightarrow 0} t_0(M).$$

At that time, the density near the center has the form

$$\rho(r, t_{\text{in}}) = \rho(0, t_{\text{in}}) \left[1 - \left(\frac{r}{R} \right)^2 + \mathcal{O}\left(\frac{r^3}{R^3} \right) \right]. \quad (13)$$

We require $r_0(M)$ to be such that Eq. (13) is reproduced for $r \ll R$, and such that $\rho(r, t_{\text{in}})$ approaches the contemporary average cosmological energy density $\bar{\rho}(t_{\text{in}})$ for $r \gg R$. The requirements are met by the choice,

$$r_0(M) = R \left[\left(\frac{M}{M_f} \right)^{\frac{1}{3}} + \frac{1}{5} \frac{M}{M_f} + \frac{1}{2} \left(\frac{M}{M_f} \right)^2 \right] \quad (14)$$

with

$$M_f = \frac{4\pi}{3} \rho(0, t_{\text{in}}) R^3. \quad (15)$$

The resulting initial density profile $\rho(r, t_{\text{in}})$ is shown in Fig. 2.

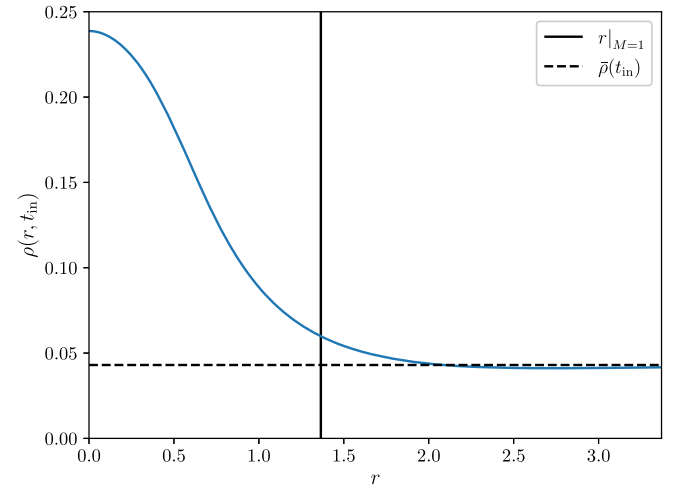


FIG. 2. Density profile at time t_{in} , when the central part of the overdensity is at turnaround, in units where $M_f = 1$ and $R = 1$. In these units the central density $\rho(\vec{0}, t_{\text{in}}) = 3/4\pi$ and the contemporary average cosmological energy density $\bar{\rho}(t_{\text{in}}) = 4/3\pi^3$. The average cosmological energy density is indicated by the horizontal dashed line. The vertical solid line indicates the radius that contains mass M_f .

The central overdensity collapses at time

$$t_{\text{coll}} = 2t_{\text{in}} = \pi \sqrt{\frac{3}{8\pi G\rho(0, t_{\text{in}})}}, \quad (16)$$

with the outer shells collapsing later. We define $M_\star(t)$ such that all axions within shell $M_\star(t)$ (but outside any black hole that may have formed) thermalize sufficiently fast that they rotate rigidly, in the three-dimensional sense. Let $\omega(t)$ be their angular rotation frequency at that time. For $M < M_\star(t)$ the specific angular momentum of the axions at the equator of shell M is therefore

$$L(M, t) = \omega(t)r(M, t)^2. \quad (17)$$

Its rate of change following the motion is

$$\dot{L}(M, t) = \frac{d\omega}{dt}(t)r(M, t)^2 + 2\omega(t)r(M, t)v_r(M, t), \quad (18)$$

where $v_r(M, t) = \dot{r}(M, t)$ is the radial velocity of the shell. The maximum rate at which the shell can shed its specific angular momentum $L(M, t)$ while collapsing is given by Eq. (8) with $r = r(M, t)$.

If there were no relaxation, $\dot{L}(M, t) = 0$ and the angular frequency of each shell would increase as $r(M, t)^{-2}$. Instead, relaxation allows the shells within $M_\star(t)$ to collapse while hardly increasing their angular rotation frequency $\omega(t)$ because their angular momentum is transported outward. The condition for a shell to collapse without hardly increasing its angular rotation frequency is that the rhs of Eq. (8) is larger than the second term on the rhs of Eq. (18), or equivalently that

$$\omega(t) \lesssim 2\pi G\rho(M, t)r(M, t) \frac{1}{v_r(M, t)} \equiv \omega_{\text{max}}(M, t). \quad (19)$$

$M_\star(t)$ is therefore the largest shell such that

$$\omega(t) < \omega_{\text{max}}(M, t) \quad (20)$$

for all $M < M_\star(t)$. Because $\omega_{\text{max}}(M, t)$ is, at all times, a decreasing function of M near $M = 0$, $M_\star(t)$ is the smallest solution of

$$\omega_{\text{max}}(M_\star(t), t) = \omega(t). \quad (21)$$

All axions within the volume enclosed by $M_\star(t)$ exchange angular momentum sufficiently fast that they can, and therefore do, rotate with the common angular frequency $\omega(t)$. On the other hand all axions outside shell $M_\star(t)$ conserve their angular momentum. Of course, the transition at $M_\star(t)$ is not sudden as we take it to be but smoothing it out is not expected to change the final outcomes significantly.

$\omega(t)$ increases with time for two distinct reasons,

$$\frac{d\omega}{dt} = \frac{d\omega}{dt}\bigg|_L + \frac{d\omega}{dt}\bigg|_T. \quad (22)$$

The first term is due to the conservation of angular momentum within the volume of axions that rotate rigidly, and the second term is due to tidal torquing. The first term is

$$\frac{d\omega}{dt}\bigg|_L = -\omega(t) \frac{\dot{I}(t)}{I(t)} \quad (23)$$

with $I(t)$ the moment of inertia of all axions between shells $M_{\text{bh}}(t)$ and $M_\star(t)$, where $M_{\text{bh}}(t)$ is the black hole mass at time t ,

$$I(t) = \frac{2}{3} \int_{M_{\text{bh}}(t)}^{M_\star(t)} dMr(M, t)^2, \quad (24)$$

and

$$\dot{I}(t) = \frac{4}{3} \int_{M_{\text{bh}}(t)}^{M_\star(t)} dMr(M, t)v_r(M, t) \quad (25)$$

is its time derivative following the motion. The second term on the rhs of Eq. (22) is estimated in Appendix B. Let us write the initial value of $\omega(t)$ as

$$\omega(t_{\text{in}}) \equiv \omega_{\text{in}} = \frac{j}{t_{\text{in}}}. \quad (26)$$

In case of rigid rotation, the relationship between j and the spin parameter is

$$\lambda = \frac{4}{5\pi} \sqrt{\frac{6}{5}} j = 0.279j + \mathcal{O}(j^2). \quad (27)$$

We therefore expect j to be in the approximate range of 0.03 to 0.8. We find in Appendix B that

$$\frac{d\omega}{dt}\bigg|_T \simeq 2.2j \frac{1}{t_{\text{in}}^2} \left(\frac{t_{\text{in}}}{t}\right)^{\frac{4}{3}} \quad (28)$$

during the interval $t_{\text{in}} < t < t_{\text{coll}}$.

Equations (19), (21), (22), (23), and (28) were solved numerically. Initially, $M_\star$ is of order M_f or larger. As $\omega(t)$ increases, it becomes more and more difficult to maintain rigid rotation, $M_\star(t)$ starts to decrease and then accelerates toward $M = 0$. Provided ω is sufficiently small at time t_{coll} , a black hole forms and grows. Soon thereafter, at a time t_f , $r(M_\star(t), t)$ reaches zero and relaxation stops. After t_f , $L(M, t)$ of each shell is conserved. The black hole mass is the largest M that satisfies $r_{\text{min}}(M, t_f) < 2GM/c^2$ or equivalently $L(M, t_f) < 2GM/c$. All the shells that do

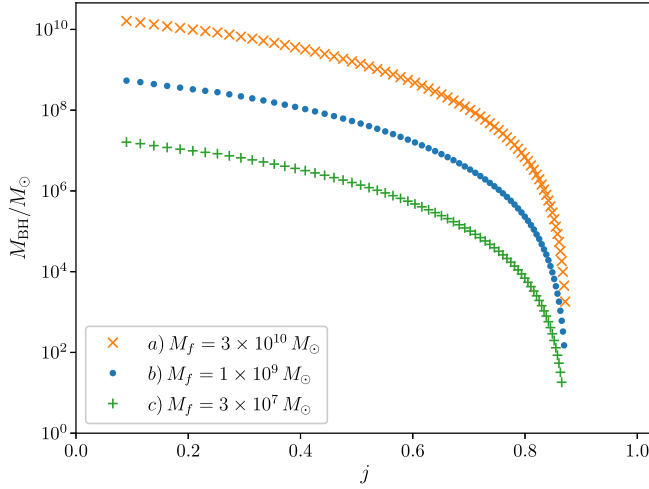


FIG. 3. Black hole mass as a function of $j \equiv \omega_{\text{in}} t_{\text{in}}$ for three values of M_f . For given M_f there is only a very slight dependence of the black hole mass on t_{in} . The values shown were computed for $z_{\text{coll}} = 10$ and hence $t_{\text{in}} = 240$ Myr.

not fall into the black hole move back out and start to form the halo that will later surround the galaxy. Although for the smallest expected j values ($j \sim 0.03$) a large fraction of the initial overdensity falls into the black hole, the total fraction of dark matter that ends up in supermassive black holes is very small since the initial overdensity is only a small fraction (of order 10^{-4}) of the galaxy it grows into.

For given j the black hole mass M_{bh} is found to be very nearly proportional to M_f . It has only a small dependence on t_{in} for fixed M_f and j . Figure 3 shows M_{bh} as a function of j for a) $M_f = 3 \times 10^{10} M_{\odot}$, b) $M_f = 10^9 M_{\odot}$, and c) $M_f = 3 \times 10^7 M_{\odot}$. For the sake of definiteness, collapse was assumed to occur at redshift $z = 10$, so that $t_{\text{coll}} \simeq 0.5$ Gyr, and hence $t_{\text{in}} \simeq 0.25$ Gyr, implying $\rho(0, t_{\text{in}}) \simeq 7.0 \times 10^{-26}$ gr/cc. The three cases were chosen to correspond to overdensities that will evolve later into a) very large galaxies, b) galaxies similar to the Milky Way, and c) small galaxies. Since the ratio M_{bh}/M_f , a function of j , has little dependence on the collapse time, it is straightforward to generate results for all plausible cases.

Figure 3 shows that no black hole forms for $j > j_c \sim 0.87$. It also implies that black holes of mass less than 10^{-5} of the initial overdensity are unlikely because this occurs only for a very small range of j values, $0.85 < j < 0.87$. Furthermore, near the cutoff at $j = 0.87$ the formation of a black hole would be easily disrupted by the various complications, such as the anisotropy of the infall, that we have neglected.

The black hole masses that result from the initial collapse of axion overdensities near cosmic dawn are in surprisingly good agreement with observations. First, the range of black hole masses formed, from approximately $10^5 M_{\odot}$ to a few times $10^{10} M_{\odot}$ is the mass range of observed supermassive black holes. As mentioned already, a cutoff is predicted at

low masses. Specifically the theory predicts that a black hole mass less than approximately $10^5 M_{\odot}$ is unlikely on the scale of Milky Way size galaxies. Also, black hole masses larger than $10^{11} M_{\odot}$ are unlikely. They occur only in the largest overdensities and only if ω_{in} is unexpectedly small. Second, the predicted supermassive black holes form near cosmic dawn. The puzzle of why supermassive black holes appear at high redshifts is removed. Although they may merge and accrete later, mergers and accretion are not necessary to explain their size. Third, there is a strong correlation between black hole mass and galaxy size. On the other hand, for a given galaxy size the black hole mass ranges widely, by a factor of one hundred or so. Both the correlation with galaxy size and the intrinsic variability are in qualitative agreement with observation. Fourth, the black holes form for approximately the range of galactic spin parameters expected from tidal torquing. Since there are galaxies with black holes close to the cutoff in j , e.g., the Milky Way whose central black hole has mass $4.3 \times 10^6 M_{\odot}$, the theory suggests that there are galaxies without a supermassive black hole because their j happens to be larger than the cutoff. This too appears consistent with observation. For example, the nearby small galaxy M33 appears not to have a supermassive black hole [21].

Our description of supermassive black hole formation does not make any *ad hoc* assumptions. No new particle is postulated, other than the standard QCD axion or an axionlike particle with similar properties. The invoked processes of thermalization and angular momentum transport in the cold axion dark matter fluid are the same as were described previously in Refs. [10,11,15]. The theory is predictive and can be tested further. It should be possible to predict the distribution of supermassive black hole masses from the distribution of overdensities of a given size at cosmic dawn and the distribution of galactic spin parameters. The amplitude and spectrum of gravitational waves produced can be calculated, to be compared with observation. By including baryons, it may be possible to determine whether the theory is consistent with the observed relation [22] between the stellar velocity dispersion in a galactic bulge and the mass of the supermassive black hole mass at its center.

ACKNOWLEDGMENTS

We are grateful to Laura Blecha for critical feedback and to Jeff Andrews, Abhishek Chattaraj, Jeff Dror, Antonios Kyriazis, Wei Xue, and Fengwei Yang for useful discussions. This work was supported in part by the U.S. Department of Energy under Grant No. DE-SC0022148 at the University of Florida.

APPENDIX A: AXION FLUID FLUCTUATIONS

It has been shown in a variety of contexts that degenerate Bosonic systems have generically large fluctuations in

intensity or density [23]. In this Appendix we show this for the cold dark matter axion fluid.

In the nonrelativistic limit, the scalar quantum field $\phi(\vec{x}, t)$ describing axions in a volume V may be written as

$$\phi(\vec{x}, t) = \frac{1}{\sqrt{2m}} [\psi(\vec{x}, t)e^{-imt} + \text{H.c.}] \quad (\text{A1})$$

and expanded

$$\psi(\vec{x}, t) = \sum_{\vec{a}} u^{\vec{a}}(\vec{x}, t) a_{\vec{a}}(t) \quad (\text{A2})$$

where the $u^{\vec{a}}(\vec{x}, t)$ are any set of orthonormal and complete (ONC) wave functions in that volume,

$$\begin{aligned} \int_V d^3x u^{\vec{a}}(\vec{x}, t)^* u^{\vec{b}}(\vec{x}, t) &= \delta^{\vec{a}\vec{b}}, \\ \sum_{\vec{a}} u^{\vec{a}}(\vec{x}, t) u^{\vec{a}}(\vec{y}, t)^* &= \delta^{(3)}(\vec{x} - \vec{y}). \end{aligned} \quad (\text{A3})$$

The $a_{\vec{a}}(t)$ and their Hermitian conjugates $a_{\vec{a}}(t)^\dagger$ satisfy canonical equal time commutation relations. The most general axion system state is given by a linear combination

$$|c_{\{\mathcal{N}\}}\rangle = \sum_{\{\mathcal{N}\}} c_{\{\mathcal{N}\}} |\{\mathcal{N}\}\rangle \quad (\text{A4})$$

of all possible particle state occupation number eigenstates

$$|\{\mathcal{N}\}\rangle = \prod_{\vec{a}} \frac{1}{\sqrt{\mathcal{N}_{\vec{a}}!}} (a_{\vec{a}}^\dagger)^{\mathcal{N}_{\vec{a}}} |0\rangle. \quad (\text{A5})$$

Here $|0\rangle$ is the empty state and $\{\mathcal{N}\} = \{\mathcal{N}_{\vec{a}} : \forall \vec{a}\}$ is the set of integers giving the occupation number of each particle state. $N = \sum_{\vec{a}} \mathcal{N}_{\vec{a}}$ is the total number of axions in volume V .

For the purpose of describing the cold dark matter axion fluid with average number density $\frac{1}{m}\rho(\vec{x}, t)$ and average velocity field $\vec{v}(\vec{x}, t)$ we choose a set of ONC wave functions as the wave function of Eq. (1) and spatial modulations thereof with wave vector $\vec{k}$,

$$u^{\vec{k}}(\vec{x}, t) = e^{i\vec{k}\cdot\vec{x}} \Psi(\vec{x}, t) \sim e^{i\vec{k}\cdot\vec{x}} \Psi(\vec{x}, t). \quad (\text{A6})$$

If the axion fluid is homogeneous and at rest, this ONC set of wave functions would simply be

$$u^{\vec{k}}(\vec{x}) = \frac{1}{\sqrt{V}} e^{i\vec{k}\cdot\vec{x}}. \quad (\text{A7})$$

When the axion fluid is inhomogeneous and/or moving, there are many ways to construct suitable $u^{\vec{k}}(\vec{x}, t)$. A particular method is described in Ref. [24]. Another way

is to start with all the wave functions $e^{i\vec{k}\cdot\vec{x}} \Psi(\vec{x}, t)$ and orthonormalize them in succession. In any such basis, the state of the axion fluid is one where most particle states have low occupation numbers but those with wave vector magnitude $k \lesssim m\delta v/\hbar$ are hugely occupied. The occupation numbers of those states that are occupied are of order 10^{61} or larger [10,11].

The number density of axions is the operator

$$n(\vec{x}, t) = \psi(\vec{x}, t)^\dagger \psi(\vec{x}, t). \quad (\text{A8})$$

In occupation number eigenstates, it has quantum mechanical average

$$\langle \{\mathcal{N}\} | n(\vec{x}, t) | \{\mathcal{N}\} \rangle = \sum_{\vec{k}} \mathcal{N}_{\vec{k}} |u^{\vec{k}}(\vec{x}, t)|^2 = \frac{1}{m} \rho(\vec{x}, t). \quad (\text{A9})$$

Let us define the operator

$$\delta n(\vec{x}, t) = n(\vec{x}, t) - \frac{1}{m} \rho(\vec{x}, t). \quad (\text{A10})$$

One readily finds

$$\langle \{\mathcal{N}\} | \delta n(\vec{x}, t) \delta n(\vec{y}, t) | \{\mathcal{N}\} \rangle = |D(\vec{x}, \vec{y}; t)|^2 \left[1 + \mathcal{O}\left(\frac{1}{\mathcal{N}}\right) \right], \quad (\text{A11})$$

where

$$D(\vec{x}, \vec{y}; t) = \sum_{\vec{k}} \mathcal{N}_{\vec{k}} u^{\vec{k}}(\vec{x}, t)^* u^{\vec{k}}(\vec{y}, t). \quad (\text{A12})$$

We have therefore

$$\langle \{\mathcal{N}\} | (\delta n(\vec{x}, t))^2 | \{\mathcal{N}\} \rangle = \left(\frac{1}{m} \rho(\vec{x}, t) \right)^2 \left[1 + \mathcal{O}\left(\frac{1}{\mathcal{N}}\right) \right]. \quad (\text{A13})$$

Thus, in eigenstates of the occupation numbers, the root-mean-square deviation from the average density at every space-time point equals the average density there. Moreover, these deviations are correlated over distances of order $\ell = \frac{1}{\delta k} = \frac{\hbar}{m\delta v}$ since $D(\vec{x}, \vec{y}; t)$ cannot vary much over distances shorter than $1/\delta k$.

In the general system states of Eq. (A4) it is not possible to make such strong statements as Eq. (A13) because they include system states in which all axions are in a single state whose wave function is a linear combination of several $u^{\vec{k}}(\vec{x}, t)$. In these very special system states, the quantum mechanical uncertainty in measuring $n(\vec{x}, t)$ vanishes. Consider nonetheless the quantum mechanical average in a general state of any operator Ω

$$\langle c_{\{\mathcal{N}\}} | \Omega | c_{\{\mathcal{N}\}} \rangle = \sum_{\{\mathcal{N}\}} \sum_{\{\mathcal{N}'\}} c_{\{\mathcal{N}\}}^* c_{\{\mathcal{N}'\}} \langle \{\mathcal{N}\} | \Omega | \{\mathcal{N}'\} \rangle. \quad (\text{A14})$$

Even in case the initial values of the complex coefficients $c_{\{\mathcal{N}\}}$ are very special, they acquire random phases after some time, especially as a result of thermal relaxation of the fluid. We have

$$\langle c_{\{\mathcal{N}\}} | \Omega | c_{\{\mathcal{N}\}} \rangle = \sum_{\{\mathcal{N}\}} |c_{\{\mathcal{N}\}}|^2 \langle \{\mathcal{N}\} | \Omega | \{\mathcal{N}\} \rangle, \quad (\text{A15})$$

where $\langle \dots \rangle$ means quantum mechanical average followed by an average over the phases of the coefficient $c_{\{\mathcal{N}\}}$. Using Eq. (A13) we have

$$\langle c_{\{\mathcal{N}\}} | (\delta n(\vec{x}, t))^2 | c_{\{\mathcal{N}\}} \rangle = \left(\frac{1}{m} \rho(\vec{x}, t) \right)^2. \quad (\text{A16})$$

That degenerate Bose fluids generically have large fluctuations is not a new result. In a 1909 [25] paper A. Einstein showed that the fluctuations in Bose fluids are not Poisson distributed and that the particles in such fluids tend to be more bunched than classical particles. The phenomenon has been thoroughly investigated, theoretically and experimentally, in the case of light beams. For a review, see Ref. [26]. In particular, Purcell [27] showed that highly degenerate photon beams have $\delta\rho = \rho$ fluctuations correlated over a timescale of order the inverse of the photon frequency dispersion. The phenomenon can be seen also in a classical field theory description of degenerate Bose fluids. It is straightforward to show that a linear combination of many plane waves has $\delta\rho = \rho$ density fluctuations if the wave amplitudes are Gaussian distributed and their phases are random.

As mentioned already, there are special system states in which the fluctuations are absent. In particular $\delta\rho = 0$ if all the particles are in a single state. (A photon beam stops fluctuating in time if its frequency dispersion vanishes, but of course a perfectly monochromatic photon beam is an idealization.) Now, Bose-Einstein condensation does move a large fraction of all particles into a single state, to wit the lowest energy available state. If the condensate is stable and its environment is static, the fluctuations disappear when the Bose-Einstein condensation is complete. However, this does not apply to the axion fluid under consideration here because it is collapsing by gravitational instability. The Bose-Einstein condensation of the axions is never complete because the lowest energy particle state keeps changing. Rigid rotation can only be maintained if the axions keep moving between states of different angular momentum. In the collapsing overdensity numerous axion states have enormous occupancy, which, as shown above, is the condition for the $\delta\rho = \rho$ fluctuations to be present.

APPENDIX B: TIDAL TORQUING OF THE OVERDENSITY

Consider the tidal torque on all the particles between shells $M_{\text{bh}}(t)$ and $M_{\star}(t)$

$$\vec{\tau}(t) = \int d\Omega \int_{r(M_{\text{bh}}(t), t)}^{r(M_{\star}(t), t)} r^2 dr \rho(\vec{r}, t) \vec{r} \times (-\vec{\nabla} \Phi(\vec{r}, t) + \vec{\nabla} \Phi(\vec{0}, t)), \quad (\text{B1})$$

where $\Phi(\vec{r}, t)$ is the gravitational potential due to density perturbations outside the overdensity of interest. The region surrounding the overdensity near cosmic dawn is an expanding Einstein–de Sitter space-time where $\vec{r} = \vec{x}(\frac{t}{t_{\text{in}}})^{\frac{2}{3}}$, $\vec{x}$ are comoving coordinates, and $\Phi(\vec{r}, t) = \Phi(\vec{x})$ [28]. We expand the gravitational potential in Taylor series about the center of the overdensity

$$\Phi(\vec{x}) = \Phi(\vec{0}) - \vec{g} \cdot \vec{x} + \frac{1}{2} \vec{x}^T \mathbb{T} \vec{x} + \dots, \quad (\text{B2})$$

where $\vec{g}$ is a constant vector and $\mathbb{T}$ a constant symmetric matrix. Equation (B1) becomes

$$\vec{\tau}(t) \simeq - \left(\frac{t}{t_{\text{in}}} \right)^{\frac{4}{3}} \int d\Omega \int_{r(M_{\text{bh}}(t), t)}^{r(M_{\star}(t), t)} r^4 dr \rho(\vec{r}, t) \hat{n}(\theta, \phi) \times \mathbb{T} \hat{n}(\theta, \phi), \quad (\text{B3})$$

where $\hat{n}(\theta, \phi) = \frac{1}{r} \vec{r}$ is the unit vector in the direction of spherical coordinates (θ, ϕ) . We expect the asphericity of the smoothed overdensity, as in Eq. (5), to dominate the integral in Eq. (B3). Let us rewrite Eq. (5) as

$$\rho(\vec{r}, t_{\text{in}}) = \rho(\vec{0}, t_{\text{in}}) \left[1 - \left(\frac{r}{R} \right)^2 A(\theta, \phi) \right], \quad (\text{B4})$$

where $A(\theta, \phi)$ is a linear combination of the spherical harmonics up to second order.

For $0 < t < t_{\text{in}}$, we estimate the integral in Eq. (B3) using linear perturbation theory in a homogeneous expanding universe of density $\rho(\vec{0}, t)$. In such a universe

$$r(M, t) = b(t) r(M, t_{\text{in}}), \quad (\text{B5})$$

where the scale factor $b(t)$ is given by

$$b(t) = \sin^2 \sigma$$

$$t = \sqrt{\frac{3}{8\pi G \rho(\vec{0}, t_{\text{in}})}} \left[\sigma - \frac{1}{2} \sin(2\sigma) \right]. \quad (\text{B6})$$

We have therefore

$$\frac{\delta\rho(r(M, t)\hat{n}, t)}{\rho(\vec{0}, t)} = \delta_+(t) \frac{\delta\rho(r(M, t_{\text{in}})\hat{n}, t_{\text{in}})}{\rho(\vec{0}, t_{\text{in}})}, \quad (\text{B7})$$

where $\delta_+(t)$ is the appropriate growth function [28] normalized so that $\delta_+(t_{\text{in}}) = 1$,

$$\delta_+(\sigma) = \frac{1}{2} \left(\frac{-3\sigma \cos \sigma}{\sin^3 \sigma} + \frac{3}{\sin^2 \sigma} - 1 \right). \quad (\text{B8})$$

Since

$$\delta\rho(r(M, t_{\text{in}})\hat{n}, t_{\text{in}}) = -\rho(\vec{0}, t_{\text{in}}) \left(\frac{r(M, t_{\text{in}})}{R} \right)^2 A(\theta, \phi) \quad (\text{B9})$$

we have

$$\delta\rho(r(M, t)\hat{n}, t) = -\delta_+(t) \rho(\vec{0}, t) \left(\frac{r(M, t)}{R} \right)^2 A(\theta, \phi). \quad (\text{B10})$$

We substitute this for $\rho(\vec{r}, t)$ in Eq. (B3) and define

$$\frac{3}{8\pi} \int d\Omega A(\theta, \phi) \hat{n}(\theta, \phi) \times \mathbb{T} \hat{n}(\theta, \phi) \equiv \Sigma \hat{z}. \quad (\text{B11})$$

Upon integrating the radial coordinate r from 0 to $R(t) = b(t)R$, Eq. (B3) becomes

$$\vec{\tau}(t) \simeq \frac{8\pi}{3} \Sigma \hat{z} \left(\frac{t_{\text{in}}}{t} \right)^{\frac{4}{3}} \frac{1}{7} R^5 b(t)^7 \rho(\vec{0}, t) \delta_+(t). \quad (\text{B12})$$

Since

$$\vec{\tau}(t) = I(t) \hat{z} \frac{d\omega}{dt} \Big|_T \quad (\text{B13})$$

with

$$I(t) = \frac{2}{3} \int_0^{R(t)} 4\pi r^4 dr \rho(\vec{0}, t) = 4\pi \frac{2}{15} R^5 b(t)^5 \rho(\vec{0}, t), \quad (\text{B14})$$

we have

$$\frac{d\omega}{dt} \Big|_T \simeq \frac{5}{7} \Sigma \left(\frac{t_{\text{in}}}{t} \right)^{\frac{4}{3}} b(t)^2 \delta_+(t) \quad (\text{B15})$$

during the time interval $0 < t < t_{\text{in}}$. Hence

$$\omega(t_{\text{in}}) \simeq \frac{5}{7} \Sigma \int_0^{t_{\text{in}}} dt \left(\frac{t_{\text{in}}}{t} \right)^{\frac{4}{3}} b(t)^2 \delta_+(t) \simeq 0.448 \Sigma t_{\text{in}}. \quad (\text{B16})$$

During the time interval $t_{\text{in}} < t < t_{\text{coll}}$, tidal torquing continues but perturbation theory breaks down. The mass distribution becomes dominated by its anisotropic component. We estimate the integral in Eq. (B3) by setting

$$\rho(r(M, t)\hat{n}(\theta, \phi), t) = \rho(M, t) A(\theta, \phi), \quad (\text{B17})$$

which yields

$$\vec{\tau}(t) \simeq \left(\frac{t_{\text{in}}}{t} \right)^{\frac{4}{3}} \Sigma I(t) \hat{z} \quad (\text{B18})$$

and hence

$$\frac{d\omega}{dt} \Big|_T \simeq \left(\frac{t_{\text{in}}}{t} \right)^{\frac{4}{3}} \Sigma. \quad (\text{B19})$$

Combining this with Eq. (B16) yields

$$\frac{d\omega}{dt} \Big|_T \simeq 2.2 \frac{\omega(t_{\text{in}})}{t_{\text{in}}} \left(\frac{t_{\text{in}}}{t} \right)^{\frac{4}{3}} \quad (\text{B20})$$

during the time interval $t_{\text{in}} < t < t_{\text{coll}}$.

-
- [1] J. Kormendy and L. C. Ho, *Annu. Rev. Astron. Astrophys.* **51**, 511 (2013).
 - [2] A. Kazunori *et al.* (the EHT Collaboration), *Astrophys. J. Lett.* **875**, L5 (2019); *Astrophys. J. Lett.* **930**, L12 (2022).
 - [3] J. Frank, A. King, and D. J. Raine, *Accretion Power in Astrophysics*, 3rd ed. (Cambridge University Press, Cambridge, England, 2002).
 - [4] G. Agazie *et al.* (the NANOGrav. Collaboration), *Astrophys. J. Lett.* **951**, L8 (2023); D. J. Reardon *et al.*, *Astrophys. J. Lett.* **951**, L6 (2023); H. Xu *et al.*, *Res. Astron. Astrophys.* **23**, 075024 (2023); J. Antoniadis *et al.* (EPTA Collaboration and InPTA Collaboration), *Astron. Astrophys.* **685**, A94 (2024).
 - [5] K. Inayoshi, E. Visbal, and Z. Haiman, *Annu. Rev. Astron. Astrophys.* **58**, 27 (2020).
 - [6] S. Balberg and S. L. Shapiro, *Phys. Rev. Lett.* **88**, 101301 (2002); J. Pollack, D. N. Spergel, and P. Steinhardt, *Astrophys. J.* **804**, 131 (2015); W.-X. Feng, H.-B. Yu, and Y.-M. Zhong, *Astrophys. J. Lett.* **914**, L26 (2021).
 - [7] T. Rindler-Daller, K. Freese, M. H. Montgomery, D. Winget, and B. Paxton, *Astrophys. J.* **799**, 210 (2015).
 - [8] R. Larson *et al.*, *Astrophys. J. Lett.* **953**, L29 (2023); A. Bogdan *et al.*, *Nat. Atron.* **8**, 126 (2024); R. Maiolino *et al.*, *arXiv:2308.01230*; L. J. Furtak *et al.*, *Nature (London)* **628**, 57 (2024); R. Miaolino *et al.*, *Nature (London)* **627**, 59 (2024).
 - [9] D. Hooper, A. Ireland, G. Krujaic, and A. Stebbins, *J. Cosmol. Astropart. Phys.* **04** (2024) 021.
 - [10] P. Sikivie and Q. Yang, *Phys. Rev. Lett.* **103**, 111301 (2009).

- [11] O. Erken, P. Sikivie, H. Tam, and Q. Yang, *Phys. Rev. D* **85**, 063520 (2012).
- [12] M. Y. Khlopov, B. A. Malomed, and Y. B. Zel'dovich, *Mon. Not. R. Astron. Soc.* **215**, 575 (1985).
- [13] J. Preskill, F. Wilczek, and M. Wise, *Phys. Lett.* **120B**, 127 (1983); L. Abbott and P. Sikivie, *Phys. Lett.* **120B**, 133 (1983); M. Dine and W. Fischler, *Phys. Lett.* **120B**, 137 (1983).
- [14] P. Sikivie and Y. Zhao, [arXiv:2601.02321](https://arxiv.org/abs/2601.02321).
- [15] N. Banik and P. Sikivie, *Phys. Rev. D* **88**, 123517 (2013).
- [16] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, *Astrophys. J.* **304**, 15 (1986).
- [17] P. J. E. Peebles, *Astrophys. J.* **155**, 393 (1969).
- [18] G. Efstathiou and B. J. T. Jones, *Mon. Not. R. Astron. Soc.* **186**, 133 (1979); J. Barnes and G. Efstathiou, *Astrophys. J.* **319**, 575 (1987).
- [19] X. Hernanadez, C. Park, B. Cervantes-Sodi, and Y.-Y. Choi, *Mon. Not. R. Astron. Soc.* **375**, 163 (2007).
- [20] C. C. Lin, L. Meistel, and F. H. Shu, *Astrophys. J.* **142**, 1431 (1965); Y. B. Zel'dovich, *Astron. Astrophys.* **5**, 84 (1970); J. Binney, *Astrophys. J.* **215**, 492 (1977).
- [21] K. Gebhardt *et al.*, *Astron. J.* **122**, 2469 (2001).
- [22] F. Ferrarese and D. Merritt, *Astrophys. J.* **539**, L9 (2000); K. Gebhardt *et al.*, *Astrophys. J.* **539**, L13 (2000).
- [23] R. K. Pathria and P. D. Beale, *Statistical Mechanics*, 3rd ed. (Elsevier, New York, 2011), and references therein.
- [24] S. Chakrabarty, S. Enomoto, Y. Han, P. Sikivie, and E. M. Todarello, *Phys. Rev. D* **97**, 043531 (2018).
- [25] A. Einstein, *Phys. Z.* **10**, 185 (1909).
- [26] L. Mandel, *Prog. Opt.* **2**, 181 (1963).
- [27] E. W. Purcell, *Nature (London)* **178**, 1449 (1956).
- [28] S. Weinberg, *Gravitation and Cosmology* (J. Wiley and Sons, New York, 1972).