Black hole superradiance of self-interacting scalar fields

Masha Baryakhtar[*]
Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, New York 10003, USA

Marios Galanis,[†] Robert Lasenby,[‡] and Olivier Simon[§]
Stanford Institute for Theoretical Physics, Stanford University, Stanford, California 94305, USA

(Received 19 February 2021; accepted 29 March 2021; published 19 May 2021)

Black hole superradiance is a powerful probe of light, weakly coupled hidden sector particles. Many candidate particles, such as axions, generically have self-interactions that can influence the evolution of the superradiant instability. As pointed out in [A. Gruzinov, arXiv:1604.06422.] in the context of a toy model, much of the existing literature on spin-0 superradiance does not take into account the most important self-interaction-induced processes. These processes lead to energy exchange between quasi-bound levels and particle emission to infinity; for large self-couplings, superradiant growth is saturated at a quasi-equilibrium configuration of reduced level occupation numbers. In this paper, we perform a detailed analysis of the rich dynamics of spin-0 superradiance with self-interactions, and the resulting observational signatures. We focus on quartic self-interactions, which dominate the evolution for most models of interest. We explore multiple distinct regimes of parameter space introduced by a nonzero self-interaction, including the simultaneous population of two or more bound levels; at large coupling, we confirm the basic picture of quasiequilibrium saturation and provide evidence that the "bosenova" collapse does not occur in most of the astrophysical parameter space. Compared to gravitational superradiance, we find that gravitational wave "annihilation" signals and black hole spin-down are parametrically suppressed with increasing interactions, while new gravitational wave "transition" signals can take place for moderate interactions. The novel phenomenon of scalar wave emission is less suppressed at large couplings, and if the particle has Standard Model interactions, then coherent, monochromatic axion wave signals from black hole superradiance may be detectable in proposed axion dark matter experiments.

DOI: 10.1103/PhysRevD.103.095019

I. INTRODUCTION

As discovered by Penrose [1], it is possible to extract energy and angular momentum from rotating black holes. While the Penrose thought experiments were in terms of mechanical scattering, equivalent processes were developed by the Zeldovich group for bosonic waves [2–4]. This phenomenon, termed "superradiance," is expected to occur in nature and, for certain initial conditions, amplify photon and graviton waves passing near rotating black holes. Moreover, if there exists a new bosonic particle with a small mass, bound states of this particle could be exponentially amplified around astrophysical black holes, forming very high occupation number "clouds" that could lead to a range of observational signatures.

Black hole (BH) superradiance as a probe of new ultralight particles was first proposed in [5], which has given rise to an extensive literature. Superradiance of new particles, including spin-0 [6–15], spin-1 [16–21], and spin-2 [22,23] fields, have been investigated, with observational signatures including black hole spin-down, gravitational wave emission, and modified black hole in-spiral dynamics; see the above for further references and [14] for a review.

Gravitational interactions are all that is necessary for BH superradiance, which makes superradiance a unique window on new particles that are otherwise inaccessible to experimental probes. However, many beyond-Standard-Model particle candidates have other interactions. These can include self-interactions, interactions with Standard Model (SM) states, and interactions with other hidden sector states. For some new particles, including the well-motivated QCD axion [24–26], both self-interactions and interactions with the SM are required by the model.

[*]mbaryakhtar@nyu.edu
[†]mgalanis@stanford.edu
[‡]rlasenby@stanford.edu
[§]osimon@stanford.edu

Therefore, it is important to understand the consequences of such interactions for the growth and behavior of superradiant bound states.

In this paper, we analyze in detail the consequences of a quartic self-interaction for the superradiance phenomenology of a light scalar around astrophysical black holes. We find that over a large range of parameter space of interest to light axion models, the addition of a quartic coupling leads to rich dynamics in the evolution of the superradiant instability, and new observational consequences. These dynamics include limiting the maximum number of particles in a bound level, populating levels inaccessible through gravitational superradiance alone, saturation to quasiequilibrium configurations of two or more levels, and emission of nonrelativistic and relativistic scalar waves to infinity. As we demonstrate, an effective quartic term is generically the most important effect driving the evolution, for much of the astrophysically relevant parameter space.

BH superradiance of a self-interacting scalar was first introduced in Ref. [6], which discussed phenomena including relativistic scalar emission, level mixing, and the possibility of a "bosenova"—a rapid, nonperturbative collapse of the cloud due to attractive self-interactions. The bosenova process was studied numerically in Ref. [27,28], and these results were used in subsequent phenomenological investigations [29,30]. However, as we will discuss, these previous analyses did not take into account self-interaction-induced energy transfers between different superradiant levels. This was pointed out (for a toy model) in [31], which showed that these energy transfer processes, along with scalar emission, can result in saturation to a two-level equilibrium configuration before the cloud has had a chance to grow large enough for a bosenova. We provide evidence that during evolution from astrophysical initial conditions, a "bosenova" does not occur in much of the phenomenologically relevant parameter space: scalar field values remain small and the cloud size required for collapse is not reached.

For small enough self-couplings—including much of the superradiance parameter space for the QCD axion—self-interaction effects are unimportant. Superradiance proceeds as in the purely gravitational case: a nonrelativistic bound state of scalars is populated by extracting energy and angular momentum from the rotating black hole, and subsequently annihilates to gravitational radiation.

Slightly larger self-interactions result in nonrelativistic scalar radiation to infinity. This new energy loss mechanism reduces the power emitted over time in gravitational wave "annihilation" signals. The interactions also populate higher angular momentum levels; the simultaneous occupation of several bound states can give rise to gravitational wave "transition" signals, in which scalars emit lower frequency gravitational waves by transitioning between two occupied levels.

Large enough self-interactions, including those typical of axion dark matter produced through the misalignment mechanism, significantly reduce the occupation number of the cloud. Instead of being limited by angular momentum conservation, superradiant growth is cut off early by self-interactions. The smaller cloud size suppresses the peak gravitational wave signal strains. For even larger self-couplings, the occupation of the cloud reaches quasi-equilibrium at parametrically smaller occupation values, as found in [31]. In this regime, the self-interactions parametrically slow the spin-down of the BH compared to the purely gravitational case.

Throughout, a new phenomenon of almost-monochromatic, nonrelativistic scalar wave emission occurs; for large self-interactions, the signal amplitude is constant on timescales up to the age of the universe. If couplings to Standard Model particles are present in addition to the self-interaction, then this scalar radiation may be detectable in proposed axion dark matter experiments. For a range of models, the self-interaction and SM interactions are controlled by the same scale; consequently, the signal in Earth-based detectors can persist for arbitrarily small occupation numbers, as long as the classical scalar field description holds.

Many of our analyses in this paper use hydrogenic approximations for bound states around BHs. Consequently, they are valid for scalar Compton wavelengths bigger than a few times the black hole light-crossing time. Understanding the behavior of more massive scalars would require numerical techniques. Since some of the most dramatic superradiance signatures may occur for slightly heavier scalars, further investigations of this kind are strongly motivated.

We review purely gravitational superradiance of scalar (spin-0) fields in Sec. II, and discuss the new processes introduced by quartic (and cubic) interactions in Sec. III. In Sec. IV, we explore in detail the evolution of the superradiant cloud in the presence of quartic self-interactions, which lead to several distinct regions in mass-coupling parameter space. In Sec. V, we discuss the maximum amplitude reached by the axion field, and whether this is large enough to cause nonperturbative behavior such as a "bosenova." We study the observable signatures of axion superradiance in the presence of self-interactions: spin down of astrophysical black holes (Sec. VI), gravitational wave annihilations and transitions (Sec. VII), and axion waves (Sec. VIII). We provide more detailed calculations related to both self-interactions and gravitational superradiance in the Appendixes A–K. We conclude and comment on directions for future investigations in Sec. IX.

II. SPIN-0 SUPERRADIANCE

In this section, we give a brief review of BH superradiance for a scalar with purely gravitational interactions. There is a broad literature on this topic; for a review, see [14]. We take our signature to be $-+++$, and assume

natural units with $c = \hbar = 1$ unless otherwise indicated. We use the convention $M_{\rm pl} \equiv 1/\sqrt{G}$ throughout.

In the Kerr background, the Killing vector tangent to the horizon, in static (Boyer-Lindquist) coordinates, is $\xi = \partial_t + \Omega_H \partial_\phi$. Here, $\Omega_H = \frac{1}{2r_g}\left(\frac{a_*}{1+\sqrt{1-a_*^2}}\right)$ is the angular velocity of the horizon and $a_* = J/GM^2$ is the dimensionless spin of the BH, where J is the BH's angular momentum, M is its mass, and $r_g \equiv GM$. Consequently, a wave with frequency ω, and angular momentum m about the BH spin axis, has energy flux $\propto \omega(\omega - m\Omega_H)$ across the horizon, relative to distant observers (the energy flux is necessarily ingoing for local observers near the horizon). For $\omega < m\Omega_H$, there is energy and angular momentum extraction from the BH, as measured at infinity.

Massive bosonic fields have quasibound states around a BH. In a Schwarzschild background, all of these states are unstable to decay. However, in a Kerr background, states with $\omega < m\Omega_H$ are unstable to *growth* [8,10,32,33].[1] Exponential growth of these superradiant states, starting either from a preexisting astrophysical population in the field, or from quantum fluctuations, will occur given enough time. If we start from the vacuum state, then ignoring the BH interior gives effectively nonunitary evolution of the field outside (due to the absorbing boundary conditions at the horizon), producing a mixed state. Interactions with external systems will generally decohere this into an almost-coherent state, with well-defined phase and amplitude. This process is analogous to the growth of a large-occupation-number laser field from quantum fluctuations [34].

The energy flux across the horizon, for a scalar field φ, is $\dot{E}_\infty \sim A_H |\varphi_H|^2 \omega(\omega - m\Omega_H)$, where $|\varphi_H|$ is the amplitude of the field at the horizon (in in-going coordinates, for which φ is smooth at the horizon), and A_H is the area of the BH horizon. This flux determines the growth rate of a quasibound state. For a scalar of mass $\mu \ll r_g^{-1}$, the lowest energy states are analogous to hydrogenic bound states, since the effect of the BH at large radii is that of a point source with a $1/r$ potential. The hydrogenic level with principal quantum number n, total angular momentum l, and azimuthal angular momentum m (around the BH spin axis) has frequency $\omega = \omega_r + i\omega_i$, where

$$\omega_r \simeq \mu\left(1 - \frac{\alpha^2}{2n^2} + \mathcal{O}(\alpha^4)\right) \tag{1}$$

with $\alpha \equiv GM\mu$ acting as the equivalent of the fine-structure constant [12,21]. The imaginary part of the frequency is

$$\omega_i \propto \alpha^{4l+5}(m\Omega_H - \omega_r)(1 + \mathcal{O}(\alpha)). \tag{2}$$

Strictly speaking, for $m \neq 0$, the leading-α form of this expression is simply $\alpha^{4l+5} m\Omega_H$. However, if $m\Omega_H$ is also small relative to r_g^{-1}, then the expression in Eq. (2) is appropriate (and more generally, changes sign at the correct ω_r). The 211 ($n = 2$, $l = 1$, $m = 1$) level, which has the fastest growth rate at small α, has $\omega_i = \frac{a_*}{48}\alpha^8\mu$ at leading order in α. The "superradiance rate," which is usually defined as the growth rate of the occupation number, is $\Gamma_{\rm SR} \equiv 2\omega_i$. The α^{4l+5} scaling for the growth rate corresponds to the field amplitude at the BH horizon—for higher-l modes, the amplitude is suppressed by the angular momentum barrier, leading to exponentially smaller growth rates for higher l modes [6,8,9,21].

While the expansions above were phrased in terms of α being small, it is actually the case that α/l is a good expansion parameter. Whenever a level is superradiant, we must have $\alpha < m/2$, so $\alpha/l < 1/2$, and the hydrogenic approximation can be used.

If the Compton wavelength of the particle is very large, i.e., $\alpha \ll 1$, then all of the superradiance rates are suppressed by a high power of α, $\Gamma \propto \alpha^{4l+4}\mu$, so are very small. Conversely, if the Compton wavelength of the particle is significantly smaller than the size of the BH, i.e., $\alpha \gg 1$, then only modes with $m \gg 1$ can be superradiant; however, these have exponentially suppressed growth rates. Consequently, for observationally relevant superradiance rates, the Compton wavelength of the particle should approximately match the size of the BH. For stellar-mass black holes, $M_{\rm BH} \sim 10\ M_\odot$, this corresponds to $\mu \sim 10^{-13}$–10^{-11} eV. While the superradiant growth rates around such BHs are rather slow on particle physics scales—with e-folding times a few minutes or longer—they can still be much faster than other astrophysical processes and timescales, allowing superradiance to occur in realistic astrophysical environments.

Once a Kerr BH is "born," e.g., in a binary merger or a supernova, the superradiant bound states start growing in amplitude. The fastest-growing level, which usually has the minimum m satisfying the superradiance condition (except close to the $\omega_r = m\Omega_H$ threshold), is the first to extract a significant amount of angular momentum from the BH, spinning it down to $\Omega_H \simeq \omega/m$.[2] For modes with the same m, the most tightly bound mode is often (for small m) the one with the largest growth rate, since it has larger amplitude at the horizon. Consequently, if $\omega = m\Omega_H$ for that mode, then $\omega > m\Omega_H$ for the other modes, and they are not growing (this is not always true for $m \geq 3$; see Sec. VII).

[1] For complex ω, as appropriate for an unstable state, the energy flux across the horizon is negative if $\frac{|\omega|^2}{{\rm Re}\omega} < m\Omega_H$ [10].

[2] Strictly speaking, a_* asymptotes toward the $\Omega_H = \omega/m$ threshold, since the superradiance rate is $\propto(m\Omega_H - \omega)$, so vanishes at the threshold. However, we will mostly ignore this small effect in the rest of the paper, and will refer to the BH being spun down "to the spin threshold."

Since the angular momentum of an astrophysical BH is very large,

$$J = a_* GM^2 = a_* \frac{M^2}{{M_{\rm pl}}^2} \simeq 10^{78} a_* \left(\frac{M}{10\ M_\odot}\right)^2, \quad (3)$$

it takes $\sim \log(J/m) \sim 180$ e-folds of superradiant growth to cause $\mathcal{O}(1)$ BH spin-down. Correspondingly, the fully grown superradiant cloud has an extremely high occupation number $\sim\mathcal{O}(1)J$. This corresponds to an energy density which is significantly higher than astrophysical DM densities (assuming that DM is not in extremely dense clumps), Appendix K. Consequently, the presence or absence of an astrophysical scalar field abundance makes little difference to its superradiant growth.

The oscillating scalar field sources gravitational wave (GW) radiation, at a frequency $\simeq 2\mu$—on a particle level, this corresponds to scalars annihilating to gravitons in the black hole background. The emitted power scales as $P \propto GN^2\mu^4\alpha^{16+4l}$, where N is the occupation number of the mode [11–13]. The smallness of G, and the high power of α, mean that this process is slow; in particular, it is always too slow to disrupt the initial superradiant growth of the level [11].

The superradiant growth of higher l levels will also take place. Once lower-l modes have grown to saturation, higher-l modes can still be superradiant, but their growth rate is slower, so there is a parametric separation between the growth times of successive levels. The annihilation process generally depletes the majority of the scalar cloud before the next level grows. Once the next level significantly spins down the BH, the first mode now has $\omega > m\Omega_H$, so is decaying with a rate comparable to its initial growth rate, and its remaining density falls back into the BH. Over sufficiently long times, a similar process will repeat for the next level.

There are a number of observational signatures of purely gravitational scalar superradiance. The first is a lack of old, fast-spinning BHs, at masses for which the scalar would have spun them down in the time available. There have been $\sim$10 measurements of stellar-mass BH spins in x-ray binary systems [35]; for high-spin BHs, these measurements can be accurate to a few percent, and have been used to set constraints the mass of weakly interacting scalars [12]. LIGO observations of binary BH mergers also enable spin measurements of the premerger BHs [15,36]. While most of these measurements are currently too imprecise to provide evidence for existence of a scalar [15,36,37], initial bounds are already possible [37] (see Sec. VI for a more detailed discussion).

Another possibility is the observation of gravitational radiation from the scalar cloud. For stellar-mass black holes, this radiation could potentially be observed at LIGO [12,15,38–40]; for heavier BHs, lower-frequency observatories such as LISA or atom interferometers [41] could have sensitivity [12,38,39]. The presence of a scalar cloud during a binary merger could also change inspiral dynamics, yielding further gravitational wave signatures [42–45]. While LIGO only observes the last few periods of BBH mergers, making such observations difficult, lower-frequency detectors will observe many more cycles, which will likely improve their chances of observing such effects.

III. QUARTIC SELF-INTERACTIONS

For a spin-0 particle, the simplest nongravitational interaction is a quartic self-interaction. This is generic in the sense that, if we expand a potential about a symmetric minimum, then the quartic is the most important interaction term for small amplitudes.

More specifically, a naturally small mass for a scalar field, as required for superradiance around astrophysical black holes, can be achieved through the breaking of a shift symmetry at some high energy scale f_a. A potential of the form $V(\varphi) = \Lambda^4 g(\varphi/f_a)$ can be generated from nonperturbative physics, so that $\Lambda \ll f_a$. For the case of a generic potential g, expanding around the minimum of the potential gives a mass scale $\mu^2 = g''\Lambda^4/f_a^2$ and a self-interaction term of order $\lambda = g^{(4)}\Lambda^4/f_a^4$.

A well-known example is the QCD axion; given a coupling $\mathcal{L} \supset \frac{\varphi}{f_a}\frac{g_s^2}{32\pi^2} G^a_{\mu\nu}\tilde{G}^{a,\mu\nu}$ of the axion φ to the QCD pseudoscalar field strength, it acquires a potential of the form [46]

$$V(\varphi) \simeq -m_\pi^2 f_\pi^2 \sqrt{1 - \frac{4m_u m_d}{(m_u + m_d)^2}\sin^2(\varphi/(2f_a))}. \quad (4)$$

resulting in a mass $\mu \simeq 6 \times 10^{-12}\ \text{eV}\frac{10^{18}\ \text{GeV}}{f_a}$, and quartic self-interaction [46],

$$\lambda \simeq 0.3\mu^2/f_a^2 \simeq 10^{-80}\left(\frac{\mu}{10^{-12}\ \text{eV}}\right)^4 \quad (5)$$

For more general axionlike particles, the natural parametric value of the quartic coupling is

$$\lambda \sim \frac{\mu^2}{f_a^2} \simeq 10^{-74}\left(\frac{\mu}{10^{-12}\ \text{eV}}\right)^2\left(\frac{10^{16}\ \text{GeV}}{f_a}\right)^2, \quad (6)$$

where we chose the nominal value of μ to be in the range of interest for stellar-mass BHs, and f_a to be around the Grand Unification (GUT) scale, for illustration. For example, a motivated target model is an axionlike particle which makes up $\mathcal{O}(1)$ of the dark matter abundance. If it is produced in the early universe by the misalignment mechanism, and starts out with a field value that is $\sim\mathcal{O}(1)f_a$, then the scale for which we obtain the correct

DM abundance is $f_a \simeq 3 \times 10^{14}\ \mathrm{GeV}(10^{-12}\ \mathrm{eV}/\mu)^{1/4}$ (assuming a time-independent potential, unlike the QCD axion case). This gives a typical quartic coupling of

$$\lambda \sim 10^{-71}\left(\frac{\mu}{10^{-12}\ \mathrm{eV}}\right)^{5/2}. \tag{7}$$

We will see that even such tiny self-coupling values can have important consequences for the dynamics and phenomenology of spin-0 superradiance.

The Lagrangian for a scalar field φ with a quartic coupling λ in a fixed background spacetime is given by

$$\mathcal{L} = -\frac{1}{2}(D_\mu\varphi)(D^\mu\varphi) - \frac{1}{2}\mu^2\varphi^2 + \frac{1}{4!}\lambda\varphi^4, \tag{8}$$

where D^μ is the covariant derivative and μ is the mass of φ. This gives the equation of motion

$$(D^2 - \mu^2)\varphi = -\frac{\lambda}{6}\varphi^3. \tag{9}$$

The quartic interaction strength λ can have either sign; $\lambda > 0$ corresponds to an attractive self-interaction, as is the case for axionlike particles, while $\lambda < 0$ is repulsive. For future convenience, we also define an energy scale f such that the quartic $\lambda \equiv \mu^2/f^2$; for an axionlike particle, we expect $f \sim f_a$, where f_a is the symmetry-breaking scale.

The states that dominate the evolution of superradiance are generally nonrelativistic, hydrogen-like wave functions; these have the fastest growth rates and so obtain the largest amplitudes. Consequently, it is helpful to perform a nonrelativistic reduction, writing

$$\varphi = \frac{1}{\sqrt{2\mu}}(\psi e^{-i\mu t} + \mathrm{c.c}). \tag{10}$$

Here, the "wave function" ψ is a complex scalar field, with $\int dV|\psi|^2 \simeq N$ the occupation number. The equation of motion is

$$(D^2 - \mu^2)\psi e^{-i\mu t} + \mathrm{c.c.} = \frac{-\lambda}{12\mu}(\psi^3 e^{-3i\mu t} + 3\psi^2\psi^* e^{-i\mu t}) + \mathrm{c.c.} \tag{11}$$

If ψ changes slowly with time, compared to μ^{-1}, then we can ignore the $\partial_t^2\psi$ terms, and extract the $e^{-i\mu t}$ part of the EoM to obtain the Gross-Pitaevskii equation [6],

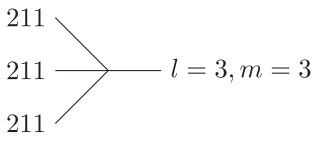
$$\left(i\partial_t + \frac{\nabla^2}{2\mu} + \frac{\alpha}{r}\right)\psi \simeq \frac{-3}{24\mu^2}\lambda\psi^2\psi^*. \tag{12}$$

The $\psi^3 e^{-3i\mu t}$ term in Eq. (11) leads to additional subdominant processes, such as the emission of relativistic φ waves, that are not captured by Eq. (12) (see Sec. III A and Appendix B 4).

As a visual aid for understanding the λ-induced interactions, we can use a diagrammatic notation for the terms of

$$\frac{\lambda}{4!}\varphi^4 = \frac{\lambda}{96\mu^2}(\psi e^{-i\mu t} + \psi^* e^{i\mu t})^4 \tag{13}$$

in close analogy to Feynman diagrams. If we expand $\psi = \sum \alpha_i\psi_i$ in some basis $\{\psi_i\}$, then legs on the left-hand side of the diagram will correspond to ψ_i terms in Eq. (13), while legs on the right-hand side will correspond to ψ_i^* terms. For example, relativistic emission sourced by the 211 hydrogenic level corresponds to the diagram



in the sense that the relevant terms in the equation of motion are obtained from terms involving ψ_{211}^3 in the Lagrangian, which source a $l, m = 3, 3$ relativistic mode. We will make use of these diagrams throughout this section.

The (typically tiny) values of λ introduced in Eq. (6) have very little effect on processes involving only a few φ quanta. In particular, if we start in a vacuum (or near-vacuum) state, the first process of interest is the superradiant growth of the most unstable hydrogenic levels, exactly as in the purely gravitational case. However, since the occupation number N of a superradiant level can reach exponentially large values [Eq. (3)], the large field amplitude can compensate for a small self-interaction, and the quartic term's effects can qualitatively alter the dynamics of superradiance. We investigate these effects below.

Higher-dimensional interactions, corresponding to higher powers of the field, will be present in general. However, we will see that, in much of the astrophysically relevant parameter space, the field never reaches large enough amplitudes for them to be important, for natural hierarchies between the mass, quartic, and higher-order terms (see Sec. V A). The case of an additional cubic coupling leads to qualitatively similar dynamics as for the quartic alone, as discussed in Sec. III D.

In the presence of a quartic interaction, three types of perturbative processes affect the evolution of the levels (here, perturbative is meant in the sense that dynamics can be treated as involving approximately hydrogenic modes, interacting on timescales long compared to their oscillation times). These are relativistic emission of axions to infinity (Sec. III A), nonrelativistic emission of axions to infinity (Sec. III B), and bound-state interactions leading to energy exchange between levels (Sec. III C). We will see in the

following sections that the latter two processes will be most important for determining the dynamics of the scalar cloud.

A. Relativistic scalar emission

One of the simplest kinds of process arising from the equation of motion [Eq. (9)] is the $3 \to 1$ process in which bound-state particles "annihilate" into a relativistic φ. In terms of the nonrelativistic reduction, the relativistic mode φ_∞ is sourced by

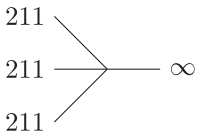
$$(D^2 - \mu^2)\varphi_\infty \simeq \frac{-\lambda/6}{(2\mu)^{3/2}}\psi^3 e^{-3i\mu t} + \text{c.c.} \tag{14}$$

This can be solved via Green's function methods, using the solution of $(D^2 - \mu^2)\varphi = 0$ in the Kerr background. For small α, when the wavelength $\sim\mu^{-1}$ of the emitted radiation is much larger than the horizon scale r_g, we can ignore the near-horizon structure of the Kerr metric, and consider only its $1/r$ behavior. These calculations are discussed in more detail in Appendix B.

For radiation sourced by the 211 hydrogenic level, which we write as $211 \times 211 \times 211 \to \infty$, the emitted power to infinity is (see Table VII)

$$P \simeq 1.5 \times 10^{-8}\alpha^{17}\mu^2\lambda^2 N_{211}^3, \tag{15}$$

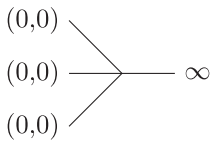
at leading order in α. The corresponding diagram is



In principle, the emitted mode has $\omega < m\Omega_H$ when the 211 level is superradiant, and so will extract additional energy from the BH. However, like the SR rate of bound states, this horizon flux is suppressed by the small overlap between the BH and the radiation, and is consequently a subleading effect in the small-α limit.

Equation (15) is $\sim$15 times larger than the estimate in [6]. The latter effectively solved the equation $\partial^2\varphi_\infty = \frac{-\lambda/6}{(2\mu)^{3/2}}\psi^3 e^{-3i\mu t} + \text{c.c.}$; that is, they approximated the emitted radiation as being massless, and propagating on a flat-space background.

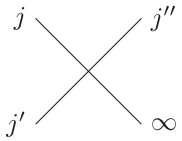
If there is some occupation number in states other than 211, then any combination of three initial states can result in relativistic radiation. If the bound states have orbital angular momenta l, l', l'', then the emitted power scales as $P \propto \alpha^{11+2(l+l'+l'')}\mu^2 NN'N''$, where N, N', N'' are the respective occupation numbers. In particular, as we will see below, populations in multiple superradiant levels can lead to forced oscillations in the $l = 0$, $m = 0$ mode. This might lead us to wonder whether the less severe α suppression in the



process, as compared to $211 \times 211 \times 211 \to \infty$, can compensate for the smaller amplitude of the 00 mode in comparison to 211. However, for the 211 and 322 occupation numbers attained in the evolution of the cloud (see Sec. IV), the emitted power via $211 \times 211 \times 211 \to \infty$, Eq. (15), is suppressed by fewer powers of α, and numerically always much larger.

B. Nonrelativistic scalar emission

Emission to unbound states can also occur in the nonrelativistic regime. Suppose that we have bound oscillations $\psi_j(t) = \psi_j e^{-i\tilde{\omega}_j t}$, where j labels a particular bound state, with frequencies $\tilde{\omega}_{j,j',j''} < 0$ (i.e., the physical frequencies are $\omega = \mu + \tilde{\omega} < \mu$). If $\tilde{\omega}_j + \tilde{\omega}_{j'} - \tilde{\omega}_{j''} > 0$, then the $\psi_j\psi_{j'}\psi^*_{j''}$ term in the equation of motion will source unbound, nonrelativistic radiation, corresponding to the diagram



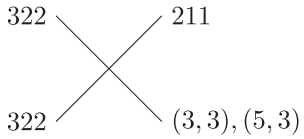
Since the emitted state is also nonrelativistic, we can consistently use the Gross-Pitaevskii equation [Eq. (12)]. Writing ψ for the radiated wave, we want to solve

$$\left(\tilde{\omega} + \frac{\nabla^2}{2\mu} + \frac{\alpha}{r}\right)\psi = \frac{-3}{12\mu^2}\lambda\psi_j\psi_{j'}\psi^*_{j''} \tag{16}$$

(with the appropriate multiplicity factors). For each of the different spherical harmonic components in the right-hand side of Eq. (16), we can write a one-dimensional radial equation for the part of Ψ with the corresponding angular dependence. These radial equations can be nondimensionalized [31], showing that the power emitted in nonrelativistic modes is given by $P \propto \alpha^4\lambda^2 N_j N_{j'} N_{j''}\mu^2$ at leading order in α, where $N_j, N_{j'}, N_{j''}$ are the occupation numbers of the bound modes. The constant factors can be found by numerically solving the radial equations, as reviewed in Appendix B 3.

Considering an example which will, in many circumstances, be very important for the cloud's evolution, suppose that we have some population in the 211 and 322 modes. Taking $\psi_{j,j'} = \psi_{322}$ and $\psi_{j''} = \psi_{211}$, we have $2\tilde{\omega}_{322} - \tilde{\omega}_{211} \simeq \frac{\alpha^2\mu}{72} > 0$, so emission to infinity is possible. As reviewed in Appendix B 3, this emission is dominantly sourced at radii $r \sim r_c \equiv r_g/\alpha^2$, i.e., where most of the cloud's mass sits. Since the dominant part of the BH potential is $\sim 1/r$ at large distances, which is spherically

symmetric, both the bound modes and the emitted wave will have approximately spherical harmonic angular dependence. For this particular case, $Y^2_{22}Y^*_{11} = \sqrt{\frac{5}{42\pi^2}}Y_{33} - \sqrt{\frac{5}{1848\pi^2}}Y_{53}$, so the emitted quanta are in the $l=3$, $m=3$ and $l=5$, $m=3$ modes. At leading order in α, the total emitted power for the



(17)

process is

$$P \simeq 10^{-8}\alpha^4\lambda^2\mu^2 N^2_{322}N_{211} \tag{18}$$

with the $(l,m)=(3,3)$ radiation dominating the emitted power.[3] This is a factor 4 smaller than the rate given in [31], due to the hydrogenic wave functions used in the latter having a normalization that is a factor $\sqrt{2}$ too large. The rates for processes involving different bound states are discussed in Appendix B 3, and tabulated in Table VI.

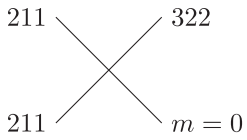
At larger α, deviations from the nonrelativistic approximation become more important. However, at small enough α such that 211 is still superradiant, the ψ_{211} and ψ_{322} wave functions are still well-approximated by the hydrogenic form, except near the origin. Since the source term $\psi^2_{322}\psi^*_{211}$ for the nonrelativistic radiation is largest at the characteristic radius of the bound states, $a \sim r_g/\alpha^2$, where the potential is dominantly $\sim 1/r$, we would expect the corrections to the nonrelativistic calculation to be small. This can be confirmed by performing a numerical computation in the Kerr background, the results of which match the leading-order formula for the emitted power [Eq. (18)] at the few percent level.

As well as relativistic effects, there will also be higher-order effects of λ; for example, self-interaction-induced distortions to the bound state wave functions, and to the radiated wave. For $\varphi/f \ll 1$, these effects will be small. In much of the astrophysically relevant parameter space, this condition holds, as we discuss in Sec. V.

C. Bound state interactions

If we have bound oscillations $\psi_{j,j',j''}$ for which $\tilde{\omega} = \tilde{\omega}_j + \tilde{\omega}_{j'} - \tilde{\omega}_{j''} < 0$, then the oscillation that they source is also bound. For example, the $\psi^2_{211}\psi^*_{322}$ term has frequency $2\tilde{\omega}_{211} - \tilde{\omega}_{322} \simeq -\frac{7\alpha^2\mu}{36} < 0$. In general, $\tilde{\omega}$ will not be very close to the frequency of any of the hydrogenic bound levels (with some exceptions that we review below) so the oscillation that they source will be forced.

Depending on the angular properties of the driving modes, the forced oscillation may gain or lose energy from the BH. If it loses energy to the BH, then for a forcing term $\psi_j\psi_{j'}\psi^*_{j''}$, this corresponds to energy loss from the $\psi_j, \psi_{j'}$ modes, but energy *gain* for the $\psi_{j''}$ mode. The example that will be the most important for us is when $\psi_j, \psi_{j'} = \psi_{211}$, and $\psi_{j''} = \psi_{322}$:



(19)

The forced oscillation has $m=0$, so loses energy through the BH horizon. Given some amplitude in the 211 and 322 modes, each $\sim\mu$ of energy lost from the forced oscillation into the BH corresponds to $\sim 2\mu$ loss from the 211 mode, and $\sim\mu$ gain in the 322 mode. The energy loss rate is proportional to the squared amplitude of the forced oscillation, which is $\propto N^2_{211}N_{322}$. Consequently, if we have a large initial occupation number in 211, and a small initial occupation number in 322, then this process will lead to the exponential growth of N_{322}, at the expense of 211.

This picture makes intuitive sense when the amplitudes of the "forcing" modes (211 and 322 in the above example) are large. However, if we are interested in e.g., the growth of 322 from quantum fluctuations, we might worry about the validity of treating it as a forcing for the $m=0$ oscillation. A more systematic approach (reviewed in Appendix A) is to assume that we have some large-amplitude ψ_c, and treat this as the source for only two of the "legs", i.e., to solve

$$(i\partial_t + \mathcal{M})\psi = \frac{-3\lambda}{24\mu^2}(\psi_c^2\psi^* + |\psi_c|^2\psi) \tag{20}$$

(here, $\mathcal{M}$ represents the other terms in the nonrelativistic Hamiltonian, including an absorbing term corresponding to the BH horizon) with ψ_c acting as a *parametric* driving term, rather than a simple forcing. When the amplitude of this driving term is small, its effects can be described as perturbations to the usual modes, "mixing" them with others. The key point is that, if the $\psi_c^2\psi^*$ term induces a mixing with a decaying mode, then this contributes a *growing* term to the original ψ mode. In our $211\times 211 \to 322\times$ BH example, if we take $\psi_c = \psi_{211}$, then this acts as a parametric driving, which mixes 322 with decaying modes such as 100. This results in the same growth rate for 322 as we would calculate from the forced oscillation picture above. Quantitatively, the energy flux into the BH is, at leading order in α,

$$P \simeq 4\times 10^{-7}\alpha^7\lambda^2\left(1+\sqrt{1-a_*^2}\right)\mu^2 N^2_{211}N_{322} \tag{21}$$

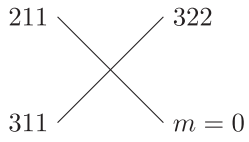
[3]This expression corresponds to the classical wave equation; in the quantum case, the final state occupation number N_{211} should be replaced by $N_{211}+1$. We use the classical expression for brevity in the remainder of the text, though the quantum version is important in allowing levels to grow from vacuum fluctuations.

More generally, for $\psi_{j,j',j''}$ such that the forced oscillation has a $m = 0$ component, the energy flux through the BH horizon is $P \propto \alpha^7\lambda^2(1+\sqrt{1-a_*^2})\mu^2 N_j N_{j'} N_{j''}$.

These calculations are discussed in Appendix B 2, and rates for different processes are tabulated in Table V. The listed processes all correspond to forced oscillations with $m = 0$. Forced oscillations with larger $|m|$ have smaller energy fluxes into (or out of) the horizon, corresponding to bound state interaction rates that are suppressed by higher powers of α.

At larger α, there will be deviations from the leading power-law behavior of Eq. (21). Since the energy lost through the forced oscillation depends on its value at the horizon, i.e., on the behavior at small distances, we would expect these deviations to be relatively greater than those for nonrelativistic radiation in the previous subsection. As we discuss in Appendix A, the behavior is similar to that of the 100 level's decay rate, with the rate a factor few larger than the leading-order value at $\alpha \sim 0.2$. While we provide leading-α expressions in the text, the semianalytic and numerical results from Appendix A are used for our results.

If all four legs of the interaction are almost on-shell, then the α scaling of the energy flux can be different from that of Eq. (21). An example, that will be of interest in Sec. IV, is



(22)

Since $\omega_r = \mu(1-\alpha^2/(2n^2)+\mathcal{O}(\alpha^4))$, we have $\omega_{211}+\omega_{311}-\omega_{322} = \omega_{200}+\mathcal{O}(\alpha^4)$ (whereas for $211\times 211 \to 322\times \mathrm{BH}$, $2\omega_{211}-\omega_{322}$ is $\mathcal{O}(\alpha^2)$ away from the frequency of any quasi-bound level). Consequently, the 200 forced oscillation dominates the energy flux into the BH, and we obtain

$$P \simeq 3\times 10^{-10}\alpha^3\lambda^2\left(1+\sqrt{1-a_*^2}\right)\mu^2 N_{211}N_{311}N_{322} \tag{23}$$

This parametrically faster rate means that any 311 occupation can be quickly depleted by this process, as we will see in Sec. IV C 2.

D. Cubic couplings

In the above, we assumed that the self-interactions consist of a quartic $\lambda\varphi^4$ interaction. A generic scalar can also have a cubic term,

$$\mathcal{L} \supset -\frac{1}{2}\mu^2\varphi^2 + \frac{g}{3!}\varphi^3 + \frac{\lambda}{4!}\varphi^4. \tag{24}$$

If we write $\lambda = \mu^2/f^2$, then a natural value for the cubic is $g = C\mu^2/f$, $C \sim \mathcal{O}(1)$. For example, if we take a cosine potential and add a slope

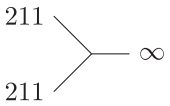
$$V(\varphi) = \mu^2 f^2(1-\cos(\varphi/f) - C\varphi/f), \tag{25}$$

then the expansion of the potential around its minimum is

$$V(\varphi_0+\delta\varphi) = \frac{\mu^2}{2}\delta\varphi^2 - \frac{C}{3!}\frac{\mu^2}{f}\delta\varphi^3 - \frac{1}{4!}\frac{\mu^2}{f^2}\delta\varphi^4 + \ldots \tag{26}$$

to leading order in small C and $\delta\varphi$.

At leading order in g, the only relevant process is relativistic $2 \to 1$ emission, in analogy to the relativistic $3 \to 1$ emission discussed in Sec. III A. For definiteness, consider again the situation for the level with the fastest superradiant rate, 211. The leading order cubic process is



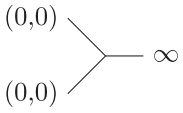
(27)

with power (see Table VII):

$$P \simeq 10^{-4}\alpha^{12}C^2(\mu^4/f^2)N_{211}^2. \tag{28}$$

More generally, for radiation sourced by quasibound levels with orbital angular momentum l and l', the emitted power scales as $P \propto \alpha^{8+2(l+l')}C^2(\mu^4/f^2)NN'$. Unlike for the case of relativistic $3 \to 1$ emission via a quartic coupling (Sec. III A), the leading-α contribution can be obtained by treating the radiation as propagating in flat space, i.e., by solving $(\partial^2-\mu^2)\varphi_\infty = \text{source}$.

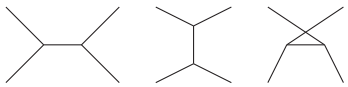
Similarly to the discussion in Sec. III A, we can ask whether the smaller α suppression of the



process, sourced by forced oscillations in the $l = 0$, $m = 0$ mode, can compensate for its smaller source amplitude compared to $211\times 211 \to \infty$. For the 211 and 322 occupation numbers attained (Sec. IV), the power from the latter process is again parametrically and numerically larger.

In the next section, we will show that, at the very least for large parts of parameter space, relativistic processes in general (from cubic or quartic vertices) are less important than quartic self-interactions between nonrelativistic states.

As well as these leading-order processes, interactions between nonrelativistic modes are generated at order g^2:



In terms of interactions between nonrelativistic modes, these are equivalent to a quartic interaction

$\lambda_{\rm eff} = \frac{5}{3}\frac{g^2}{\mu^2} = \frac{5C^2}{3}\frac{\mu^2}{f}$, which is always attractive.[4] It should be noted that this is only true for nonrelativistic modes; other processes induced at order g^2, such as $3 \to 1$ emissions, will not be captured by the same effective quartic. Nevertheless, as we will discuss in Sec. IV, in many circumstances, only processes involving nonrelativistic states are important for the evolution of the field around the BH.

Since the most important behavior can generally be captured by an effective quartic coupling, we will ignore the cubic coupling for most of this paper, setting $C = 0$. For $C \neq 0$, one can use the replacement rule

$$\frac{1}{f^2} \to \frac{1}{f_{\rm eff}^2} = \left(1 + \frac{5}{3}C^2\right)\frac{1}{f^2} \tag{29}$$

for processes involving only nonrelativistic states.

E. Summary

In "gravitational" superradiance, there are two generic ways for bound states to gain or lose energy and thus particle number: superradiance itself, in which the black hole acts as an energy and angular momentum source, and gravitational radiation, which carries energy and angular momentum to infinity. We have seen that in the presence of quartic self-interactions, three new classes of processes are introduced: emission of relativistic axion waves to infinity, emission of nonrelativistic axion waves to infinity, and excitation of forced oscillations which typically are absorbed back into the black hole.

A nonzero cubic self-interaction can act as an additional source of relativistic emission, as well as contributing to an effective quartic term. We will see that, unless the cubic coupling is tuned so as to suppress the effective quartic coupling, or the cubic is rather large compared to its natural value ($|C| \gg 1$), relativistic emission generally does not have an important effect on the dynamics.

The first investigation of scalar self-interactions in BH superradiance was in Ref. [6], which carried out a very similar analysis to ours; for example, Eq. (50) in Ref. [6] corresponds to our Eq. (20) describing bound-state interactions. However, in considering whether a perturbation grows or shrinks, Ref. [6] focused on the energy flux through the BH horizon, and did not take into account energy transfer, through the parametric forcing term, between bound states. Since the BH absorbs energy in e.g., the $211 \times 211 \to 322 \times$ BH process, the conclusion was that interaction between modes suppresses occupation number growth. This seems to account for the discrepancy between our analysis and the conclusions of Ref. [6].

The processes outlined in this section create new energy loss mechanisms for bound states, thereby typically limiting their occupation numbers below those of gravitational superradiance. They also create the ability to exchange particles efficiently between bound states with different energy and angular momentum, enabling the growth of high angular momentum states on timescales much faster than the growth possible through gravitational superradiance alone. In the following section, we will discuss in detail the new dynamics for a range of self-interaction strengths.

Finally, similarly to the emission processes discussed above, there will also be effects that are higher order in λ. In particular, if the amplitude of the cloud becomes too large, then the attractive self-interactions will lead to a rapid, nonperturbative collapse, the "bosenova" [6]. However, we will see that, for most parts of parameter space, the leading order in λ processes that we have described will prevent the field from reaching such large amplitudes. We discuss such nonperturbative behavior in more detail in Sec. V.

IV. PERTURBATIVE EVOLUTION

In this section, we study the evolution of the cloud-BH system, when the new dynamics introduced by self-interactions can be treated perturbatively. That is, we treat the cloud as consisting of approximately hydrogenic levels, interacting on timescales long compared to their oscillation timescales. Although the processes are individually simple, the number of them involved can make the narrative hard to follow. Accordingly, we have collated some of the most important information into a number of tables and figures. Table I lists the most important processes affecting level evolution, and gives their rates. Figure 3 is an important guide to how our discussion is structured, showing the four qualitatively different regimes of parameter space that we analyze. Table II gives approximate expressions for the boundaries of these regions, and points to their definitions in the text. Figure 4 shows examples of the time evolution of the cloud-BH system, drawn from the four different regions. Table III summarizes the level occupation numbers, observational signatures, and characteristic timescales associated with each region.

A. Evolution of occupation numbers

The evolution of the scalar field around the BH is driven by the gravitational processes discussed in Sec. II—superradiant growth or decay, and GW emission—and by the interaction-mediated processes discussed in Sec. III. As we have seen, when these processes can be treated perturbatively, they can be viewed as transferring energy to and from the quasi-bound states of the field (which are themselves only slightly perturbed from their hydrogenic forms). Putting everything together, we can write down a set of coupled differential equations, governing the evolution of the occupation numbers of the modes.

Schematically, if we write the occupation number of level j as N_j (where we index the different quasibound states by a single index j), then

[4]We find that the contribution of the cubic coupling to the effective quartic is greater than the one in [31] by a factor of 5/4.

TABLE I. Rates for the most important processes involved in the evolution of the 211 and 322 hydrogenic levels, at leading order in α. The second column shows the rate constants appropriate for occupation numbers N_{211} etc, as per equations (30) and (31), while the third column shows the rate constants for normalized occupation numbers $\varepsilon_{211} \equiv N_{211}/(GM_{\rm BH}^2)$ etc, as per Eq. (32).

Process	Rate constant (occupation numbers N)	Rate constant (normalized occupation numbers ε)
211 superradiance	$\Gamma^{\rm SR}_{211} \simeq 4\times10^{-2}\alpha^8(a_* - 2\alpha(1+\sqrt{1-a_*^2})\mu$	$\gamma^{\rm SR}_{211} = \Gamma^{\rm SR}_{211}$
211 → ∞ + 211	$\Gamma^{\rm GW}_{211\times211} \simeq 10^{-2}\alpha^{12}(\frac{\mu}{M_{\rm pl}})^2\mu$	$\gamma^{\rm GW}_{211\times211} \simeq 10^{-2}\alpha^{14}\mu$
211 + 211 → 322 + BH	$\Gamma^{322\times{\rm BH}}_{211\times211} \simeq 4\times10^{-7}\alpha^7\lambda^2(1+\sqrt{1-a_*^2})\mu$	$\gamma^{322\times{\rm BH}}_{211\times211} \simeq 4\times10^{-7}\alpha^{11}(\frac{M_{\rm pl}}{f})^4(1+\sqrt{1-a_*^2})\mu$
322 + 322 → 211 + ∞	$\Gamma^{211\times\infty}_{322\times322} \simeq 10^{-8}\alpha^4\lambda^2\mu$	$\gamma^{211\times\infty}_{322\times322} \simeq 10^{-8}\alpha^8(\frac{M_{\rm pl}}{f})^4\mu$
322 superradiance	$\Gamma^{\rm SR}_{322} \simeq 8\times10^{-5}\alpha^{12}(a_* - \alpha(1+\sqrt{1-a_*^2})\mu$	$\gamma^{\rm SR}_{322} = \Gamma^{\rm SR}_{211}$

TABLE II. Approximate expressions for the boundaries between different regions in μ, f parameter space, as diagrammed in the bottom-right panel of Fig. 3. The first column identifies the section in the text discussing the particular parameter space region, while the third column presents the f range (for given μ) corresponding to that region, along with references to the relevant equations in the text. The expressions given are to leading order in small α, and numerical coefficients are approximate; the reader should refer to the text for more precise expressions.

Coupling strength	Fig 3	Boundary in parameter space
Small (IV B 2)	A	$f > f_{\rm AB} \approx \min[3\times10^{16}\ {\rm GeV}(\frac{T_{\rm BH}}{10^{10}\ {\rm yr}})^{\frac{1}{4}}(\frac{\mu}{10^{-13}\ {\rm eV}})^{\frac{1}{4}}(\frac{\alpha}{0.01})^{\frac{11}{4}},$ $8\times10^{18}\ {\rm GeV}(\frac{0.01}{\alpha})^{\frac{3}{4}}(\frac{a_*}{0.9})^{\frac{1}{4}}]$ (Eqs. (41), (42)
Moderate (IV B 3)	B	$f_{\rm AB} > f > f_{\rm BC} \approx 2\times10^{16}\ {\rm GeV}(\frac{a_*(t_0)}{0.9})^{\frac{1}{4}}\min[(\frac{\alpha}{0.04})^{\frac{3}{4}}, (\frac{\alpha}{0.04})^{\frac{3}{2}}]$ (Eqs. (53), (54), (56)
Large (IV B 4)	C	$f_{\rm BC} > f > f_{\rm CD} \approx 3\times10^{14}\ {\rm GeV}(\frac{10^{10}\ {\rm yr}}{T_{\rm BH}})^{\frac{1}{2}}(\frac{10^{-13}\ {\rm eV}}{\mu})^{\frac{1}{2}}(\frac{0.01}{\alpha})^{\frac{5}{2}}(\frac{0.9}{a_*(t_0)})^{\frac{3}{4}}$ (Eqs. (62)
No spindown (IV B 5)	D	$f_{\rm CD} > f \gg \mu$

TABLE III. Summary of important quantities in the parameter space regimes A-D (Fig 3, Table II). The second column lists the ratio of the peak value $\varepsilon^{\rm peak}_{211}$ attained in the corresponding region to the maximum value attained through gravitational superradiance $\varepsilon^{\rm max}_{211}$. The fourth column describes the most important observational signatures of superradiance in each regime. For regions A and B, these are BH spindown (see Sec. VI), the emission of gravitational radiation (see Sec. VII) from $211\times211 \to$ GW annihilations and from $322 \to 211\times$ GW transitions (only in region B). For regions C and D, gravitational radiation is suppressed, but nonrelativistic scalar radiation ("AW", for "axion waves") from the $322\times322 \to 211\times\infty$ process may be detectable, if the scalar field couples to SM states (see Sec. VIII). The right-most column gives approximate expressions for the relevant dynamical timescales, which also correspond to typical signal timescales of GW radiation (for A and B) and scalar radiation (for C and D). The expressions given are to leading order in small α, and numerical coefficients are approximate; the reader should refer to the text for more precise expressions.

Coupling strength	$\varepsilon^{\rm peak}_{211}/\varepsilon^{\rm max}_{211}$	$\eta = \varepsilon_{322}/\varepsilon_{211}$	Signatures	Timescales
Small (IV B 2), A	1	$\simeq 0$	Spindown, GW	$\tau_{\rm ann} \approx 10^5\ {\rm yr}(\frac{0.1}{\alpha})^{14}(\frac{10^{-12}\ {\rm eV}}{\mu})$ [Eq. (39)]
Moderate (IV B 3), B	1	$10^{-5}(\frac{\alpha}{0.01})^3$	Spindown, GW	$\tau_{\rm scalar} \approx 10^{-1}\ {\rm yr}(\frac{0.1}{\alpha})^{14}(\frac{10^{-12}\ {\rm eV}}{\mu})(\frac{f}{10^{17}\ {\rm GeV}})^4$ [Eq. (49)]
Large (IV B 4), C	$(\frac{f}{f_{\rm BC}})^2$	$10^{-5}(\frac{\alpha}{0.01})^3$	Slow spindown, AW	$\tau_{\rm sd} \approx 10^7\ {\rm yr}(\frac{0.01}{\alpha})^5(\frac{10^{-12}\ {\rm eV}}{\mu})(\frac{0.9}{a_*})^{\frac{3}{2}}(\frac{10^{15}\ {\rm GeV}}{f})^2$ [Eq. (60)]
No spindown (IV B 5), D	$(\frac{f}{f_{\rm BC}})^2$	$10^{-5}(\frac{\alpha}{0.01})^3$	No spindown, AW	$\tau_{\rm sd} \gtrsim T_{\rm BH}$ [Eq. (63)]

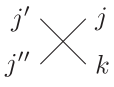
$\varepsilon^{\rm eq}_{211} \approx \frac{2}{\sqrt{3}}\frac{\sqrt{\kappa^\infty\kappa^{\rm SR}(a_*-2\alpha\tilde{r}_+)}}{\alpha^3\kappa^{\rm BH}\tilde{r}_+}(\frac{f}{M_{\rm pl}})^2 = 2.5\times10^{-1}(\frac{0.01}{\alpha})^3(\frac{a_*}{0.9})^{1/2}(\frac{f}{10^{15}\ {\rm GeV}})^2$ [Eq. (55a)];

$\varepsilon^{\rm eq}_{322} \approx \sqrt{\frac{1}{3}\frac{\kappa^{\rm SR}(a_*-2\alpha\tilde{r}_+)}{\kappa^\infty}}(\frac{f}{M_{\rm pl}})^2 = 6.9\times10^{-6}(\frac{a_*}{0.9})^{1/2}(\frac{f}{10^{15}\ {\rm GeV}})^2$ [Eq. (55b)].

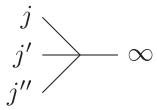
$$\dot{N}_j = \Gamma^{\rm SR}_j N_j + \sum_{j'}(-c\Gamma^{\rm GW}_{j\times j'} + \Gamma^{\rm GW}_{j'\to j} - \Gamma^{\rm GW}_{j\to j'})N_jN_{j'} + \sum_{j',j''}(\Gamma^{j\times k}_{j'\times j''} - c\Gamma^{j''\times k}_{j\times j'} - c\Gamma^{j'\times k}_{j\times j''} - c\Gamma^{\infty}_{j\times j'\times j''})N_jN_{j'}N_{j''} \tag{30}$$

where the $c\Gamma$ notation encodes the appropriate multiplicity factors, and

(i) $\Gamma_j^{\rm SR}$ is the growth(/decay) rate corresponding to the mode's flux across the BH horizon
(ii) $\Gamma^{\rm GW}_{j\times j'}$ is the annihilation rate of $j\times j'$ to gravitational radiation.
(iii) $\Gamma^{\rm GW}_{j\to j'}$ is the rate of transitions, via gravitational-wave emission, from j' to j.
(iv) $\Gamma^{j\times k}_{j'\times j''}$ is the rate of the



process, where the k leg corresponds to nonrelativistic scalar emission, or to bound forced oscillation damped by the BH. For emission to infinity, we will sometimes write $\Gamma^{j''\times\infty}_{j\times j'}$, while for a bound forced oscillation, we will write $\Gamma^{j''\times{\rm BH}}_{j\times j'}$.
(v) $\Gamma^{\infty}_{j\times j'\times j''}$ is the rate of the



relativistic emission process. Repeated indices will sometimes be abbreviated using an exponential (i.e., $\Gamma^{\infty}_{j\times j\times j}=\Gamma^{\infty}_{j^3}$)

For example, the evolution of the fastest-growing level is given by

$$\begin{aligned}\dot N_{211} =& \Gamma^{\rm SR}_{211}N_{211} - 2\Gamma^{\rm GW}_{211\times 211}N_{211}^2 - \Gamma^{\rm GW}_{211\to 322}N_{211}N_{322} + \ldots \\ & - 2\Gamma^{322\times{\rm BH}}_{211\times 211}N_{211}^2N_{322} + \Gamma^{211\times\infty}_{322\times 322}N_{211}N_{322}^2 + \ldots \\ & - 3\Gamma^{\infty}_{(211)^3}N_{211}^3 - 2\Gamma^{\infty}_{(211)^2\times 322}N_{211}^2N_{322} + \ldots \end{aligned} \tag{31}$$

Some of the key rates, at leading order in α, are listed in Table I.

While, as we observed above, λ is often extremely small, the N_j can become extremely large. From Eq. (3), the angular momentum of a BH is $J=a_*GM^2\simeq 10^{78}a_*(\frac{M}{10M_\odot})^2$. To spin it down by $\mathcal{O}(1)$, as is necessary to saturate the superradiant instability, we need N_j to be of this order. Consequently, it is often more convenient to work in terms of "normalized" occupation numbers, $\varepsilon_j\equiv N_j/(GM^2_{\rm BH})<1$, and normalized rates γ such that

$$\begin{aligned}\dot\varepsilon_j =& \gamma_j^{\rm SR}\varepsilon_j + \sum_{j'}(-c\gamma^{\rm GW}_{j\times j'} + \gamma^{\rm GW}_{j'\to j} - \gamma^{\rm GW}_{j'\to j})\varepsilon_j\varepsilon_{j'} \\ & + \sum_{j',j''}(\gamma^{j\times k}_{j'\times j''} - c\gamma^{j''\times k}_{j\times j'} - c\gamma^{j'\times k}_{j\times j''} - c\gamma^{\infty}_{j\times j'\times j''})\varepsilon_j\varepsilon_{j'}\varepsilon_{j''} \end{aligned} \tag{32}$$

Similarly, it is helpful to write $\lambda\equiv\mu^2/f^2$, as motivated around Eq. (6). In terms of these, the scalings with α and f of the different γ are

(i) For growth (or decay) of a bound oscillation via the BH horizon $\gamma_j^{\rm SR}\propto\alpha^{4l+4}$
(ii) For nonrelativistic scalar emissions to infinity, $\gamma^{j\times\infty}_{j'\times j''}\propto\alpha^8(M_{\rm pl}/f)^4$
(iii) For the absorption of energy from a forced bound oscillation with angular momentum l damped by the BH, $\gamma^{j\times{\rm BH}}_{j'\times j''}\propto\alpha^{11+4l}(M_{\rm pl}/f)^4$ (except in the case of "resonant" processes, as discussed in Sec. III C).
(iv) For 3-to-1 relativistic scalar emissions to infinity $\gamma^{\infty}_{j\times j'\times j''}\propto\alpha^{2(l+l'+l'')+15}$
(v) For annihilation to gravitational waves $\gamma^{\rm GW}_{j\times j'}\propto\alpha^{10+2(l+l')}$
(vi) For transitions between bounds states with gravitational wave emission, $\gamma^{\rm GW}_{j\to j'}$, see Sec. VII B.

In addition, a nonzero cubic interaction contributes to the evolution equations (32) as

$$\dot\varepsilon_j = -\sum_{j'}c\gamma^{\infty}_{j\times j'}\varepsilon_j\varepsilon_{j'} + \ldots \tag{33}$$

with rate $\gamma^{\infty}_{j\times j'}\propto\alpha^{2(l+l')+10}|C|^2(M_{\rm pl}/f)^2\mu$.

The rates that determine the evolution in large parts of the parameter space are listed in Table I, at leading order in α. As discussed above, for some of these processes, this approximation can be quite poor at α values of interest, and for the computations involved in producing our plots, we use more accurate numerical or semianalytic expressions.

When all of the ε_j are very small, then only the $\gamma_j^{\rm SR}$ are important, and evolution proceeds as in the purely gravitational case, with the fastest-growing level increasing exponentially in amplitude. Since the ε_j for this level will usually dominate exponentially over the other $\varepsilon_{j'}$, other levels can only be built up (faster than their superradiance rates) through[5]

where the BH leg corresponds to a bound oscillation. If interaction processes are strong enough to significantly

[5]If the occupied level j is higher-frequency than some other level j', then transitions from j to j' via GW emission can also occur. However, as discussed in Sec. VII, the fastest-growing superradiant level is also the most tightly bound superradiant level, for $l<3$. Consequently, transitions from a superradiant level would have to be to decaying levels. Since the decay rate through the BH horizon is generally significantly larger than the growth rate due to GW transitions, this does not give rise to exponential growth of j'. For example, if we consider $322\to 200+$ GW transitions, the evolution equation for the 200 level is $\dot\varepsilon_{200}/\mu\simeq 4\times10^{-6}\alpha^8\varepsilon_{322}\varepsilon_{200}-0.5\alpha^5\varepsilon_{200}$, so the 200 level is still damped even for large ε_{322}.

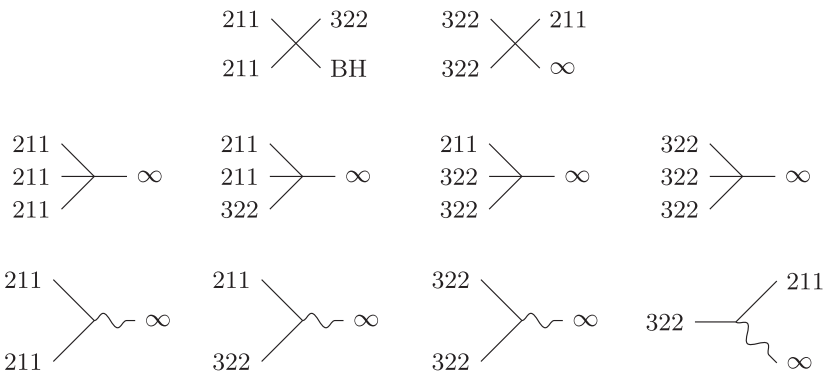


FIG. 1. Processes relevant to the evolution of the 211 and 322 hydrogenic modes. The first row corresponds to the interactions between nonrelativistic modes (Sec. III B and III C) and the second corresponds to the emission of relativistic scalar radiation (Sec. III A), both mediated by the quartic self-interaction. The third row corresponds to the emission of gravitational radiation (indicated by wavy legs), also present in gravitational superradiance.

affect the evolution, then the j' for which this growth rate is fastest will be the next level to become important.

For small α, the fastest superradiant growth is for the 211 level, and the fastest quartic process, given a 211 amplitude, is

as discussed in Sec. III C. It turns out that, similarly to the toy model discussed in [31], there is a large parameter space for which *only* the (perturbed) 211 and 322 levels are ever significantly populated. This regime will be the main focus of our paper.

Situations in which 211 is the first superradiant level generally lead to the strongest radiative signals, either in gravitational or scalar waves. However, superradiance into higher levels can be important for other phenomenological signatures, such as BH spin-down. In such circumstances, levels other than 211 and 322 will be important. For example, if 322 is the first level to grow through superradiance, then 544 will generally be the next level to be built up through self-interactions. Though we do not investigate such scenarios in detail in this paper, they represent an important subject for future work.

B. Two-level system

If the (suitably perturbed) 211 and 322 modes are the only ones with significant occupation numbers, then the relevant processes are illustrated in Fig. 1. Given this multitude of processes, the behavior of the system seems potentially very complicated. However, we will see that, because the relativistic emission rates are suppressed by high powers of α (and the gravitational radiation rates have an additional relative suppression of $(f/M_{\rm pl})^4$, which will turn out to be small when self-interactions are important), only the two nonrelativistic processes (along with superradiance) are generally significant.

Assuming that 211 is the fastest-growing mode at the start of the evolution, these give rise to fairly simple qualitative behavior, for large enough couplings λ. Initially, the 211 mode grows through superradiance. Once its occupation number is large enough, the growth rate of the 322 mode, through the $211 \times 211 \to 322 \times \mathrm{BH}$ process, becomes significant. This stops the growth of the 211 mode. Since 322 is depleted via the $322 \times 322 \to 211 \times \infty$ process, but built up via $211 \times 211 \to 322 \times \mathrm{BH}$ (and vice versa for 211), the 211 and 322 modes reach a quasiequilibrium configuration, in which their occupation numbers are almost constant. This evolution is illustrated schematically in Fig. 2, and is the regime that was studied in the toy model of [31].

The above picture holds for the case of large enough self-couplings; in the opposite limit of very small self-couplings, the evolution will be almost the same as the purely gravitational case. For intermediate values of λ, there can be more complicated behaviors. In the rest of this section, we will make all of these statements precise, by investigating in detail the evolution of the cloud, for different μ and f. Figure 3, and Tables II and III, serve as guides to this discussion. Readers more interested in the observational effects of superradiance around astrophysical BHs can skip ahead to Secs. VI and VII, referring back to this section when necessary.

1. Evolution equations

As discussed above, only the processes in Table I are generally important in the evolution of the 211/322 system. We highlight these rates (which are presented outside the parentheses) in the full evolution equations for the occupation numbers of the 211 and 322, which are (at leading order in α)

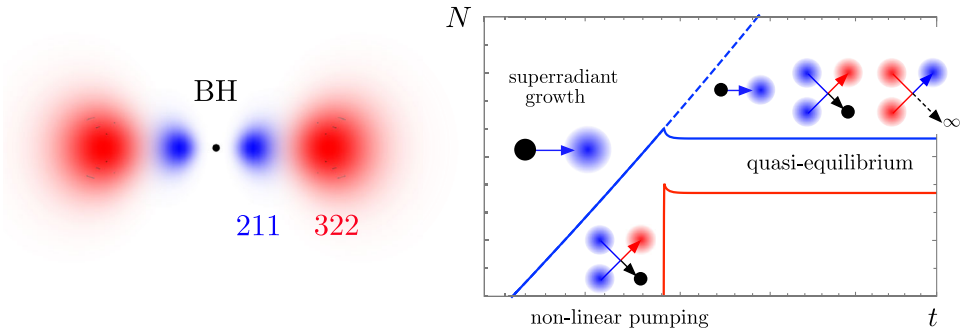


FIG. 2. Schematic illustration of the effects of a large quartic self-interaction on the growth of scalar fields around a spinning BH. The left-hand figure shows the energy densities of the 211 (blue) and 322 (red) modes in the (x, z) plane, taking the BH spin to be in the z direction. The right-hand panel shows the evolution of the 211 (blue) and 322 (red) occupation numbers with time (where the N axis is taken to be logarithmic). We assume that the initial BH spin is high enough that the first process to occur is superradiant growth of 211. In the absence of self-interactions, this growth would continue until the BH was spun down to the $m = 1$ threshold (as indicated by the dashed blue line). When sufficiently large self-interactions are present, the 322 mode is built up from the 211 mode, via the nonlinear pumping process described in Sec. III C. This stops the growth of 211, and the levels quickly reach a quasiequilibrium configuration, in which the processes of 211 superradiance, $211 \times 211 \to 322 \times \mathrm{BH}$ and $322 \times 322 \to 211 \times \infty$ emission (Sec. III B) keep the 211 and 322 occupation numbers almost constant.

$$\frac{\dot{\varepsilon}_{211}}{\mu} = \kappa^{\mathrm{SR}}_{211}\alpha^8(a_* - 2\alpha\tilde{r}_+)\varepsilon_{211} - 2\kappa^{322\times\mathrm{BH}}_{211\times211}\alpha^{11}(M_{\mathrm{pl}}/f)^4\tilde{r}_+\varepsilon^2_{211}\varepsilon_{322} + \kappa^{211\times\infty}_{322\times322}\alpha^8(M_{\mathrm{pl}}/f)^4\varepsilon^2_{322}\varepsilon_{211} - 2\kappa^{\mathrm{GW}}_{211\times211}\alpha^{14}\varepsilon^2_{211}$$
$$+ (-\kappa^{\mathrm{GW}}_{211\times322}\alpha^{16}\varepsilon_{211}\varepsilon_{322} + \kappa^{\mathrm{GW}}_{322\to211}\alpha^{10}\varepsilon_{211}\varepsilon_{322} - 3\kappa^{\infty}_{(211)^3}\alpha^{21}(M_{\mathrm{pl}}/f)^4\varepsilon^3_{211} - 2\kappa^{\infty}_{(211)^2\times(322)}\alpha^{23}(M_{\mathrm{pl}}/f)^4\varepsilon^2_{211}\varepsilon_{322}$$
$$- \kappa^{\infty}_{(211)\times(322)^2}\alpha^{25}(M_{\mathrm{pl}}/f)^4\varepsilon_{211}\varepsilon^2_{322}), \tag{34}$$

$$\frac{\dot{\varepsilon}_{322}}{\mu} = \kappa^{\mathrm{SR}}_{322}\alpha^{12}(a_* - \alpha\tilde{r}_+)\varepsilon_{322} + \kappa^{322\times\mathrm{BH}}_{211\times211}\alpha^{11}(M_{\mathrm{pl}}/f)^4\tilde{r}_+\varepsilon^2_{211}\varepsilon_{322} - 2\kappa^{211\times\infty}_{322\times322}\alpha^8(M_{\mathrm{pl}}/f)^4\varepsilon^2_{322}\varepsilon_{211}$$
$$+ (-2\kappa^{\mathrm{GW}}_{322\times322}\alpha^{18}\varepsilon^2_{322} - \kappa^{\mathrm{GW}}_{211\times322}\alpha^{16}\varepsilon_{211}\varepsilon_{322} - \kappa^{\mathrm{GW}}_{322\to211}\alpha^{10}\varepsilon_{211}\varepsilon_{322} - 3\kappa^{\infty}_{(322)^3}\alpha^{27}(M_{\mathrm{pl}}/f)^4\varepsilon^3_{322}$$
$$- \kappa^{\infty}_{(211)^2\times(322)}\alpha^{23}(M_{\mathrm{pl}}/f)^4\varepsilon^2_{211}\varepsilon_{322} - 2\kappa^{\infty}_{(211)\times(322)^2}\alpha^{25}(M_{\mathrm{pl}}/f)^4\varepsilon_{211}\varepsilon^2_{322}), \tag{35}$$

where $\tilde{r}_+ \equiv r_+/r_g = 1 + \sqrt{1 - a_*^2}$, and the κ values correspond to the γ rates, with the leading α, f and a_* dependence factored out (e.g., $\gamma^{322\times\mathrm{BH}}_{211\times211} = \kappa^{322\times\mathrm{BH}}_{211\times211}\alpha^{11}(M_{\mathrm{pl}}/f)^4\tilde{r}_+\mu$, etc). We also need to keep track of the BH's mass and spin, for which

$$\frac{\dot{a}_*}{\mu} = -\kappa^{\mathrm{SR}}_{211}\alpha^8(a_* - 2\alpha\tilde{r}_+)\varepsilon_{211} - 2\kappa^{\mathrm{SR}}_{322}\alpha^{12}(a_* - \alpha\tilde{r}_+)\varepsilon_{322}, \tag{36}$$

and

$$\frac{\dot{M}}{\mu^2 GM^2} \simeq -\kappa^{\mathrm{SR}}_{211}\alpha^8(a_* - 2\alpha\tilde{r}_+)\varepsilon_{211} - \kappa^{\mathrm{SR}}_{322}\alpha^{12}(a_* - \alpha\tilde{r}_+)\varepsilon_{322} + \kappa^{322\times\mathrm{BH}}_{211\times211}\alpha^{11}(M_{\mathrm{pl}}/f)^4\tilde{r}_+\varepsilon^2_{211}\varepsilon_{322}. \tag{37}$$

A simplifying assumption at small α is to neglect the change in the mass of the black hole; we will often use this approximation in the text. This is equivalent to setting the maximum 211 fractional occupation value attained through purely gravitational evolution, $\varepsilon^{\mathrm{max}}_{211}$, to $|\Delta a_*| = a_*(t_0) - 4\alpha/(1 + 4\alpha^2)$. At larger α, the mass of the BH changes more significantly and $\varepsilon^{\mathrm{max}}_{211} > |\Delta a_*|$. Our expressions can still be used, however, with the correct value of $\varepsilon^{\mathrm{max}}_{211}$, for which we derive good analytic approximations in Appendix F.

2. Small self-coupling: Gravitational superradiance

In the limit of very small coupling, $f \to \infty$, the system evolves under purely gravitational dynamics, as summarized in Sec. II. As long as the fastest and second-fastest growing superradiant levels have sufficiently different growth rates, the former will grow first, and attain exponentially larger occupation numbers than other modes. For

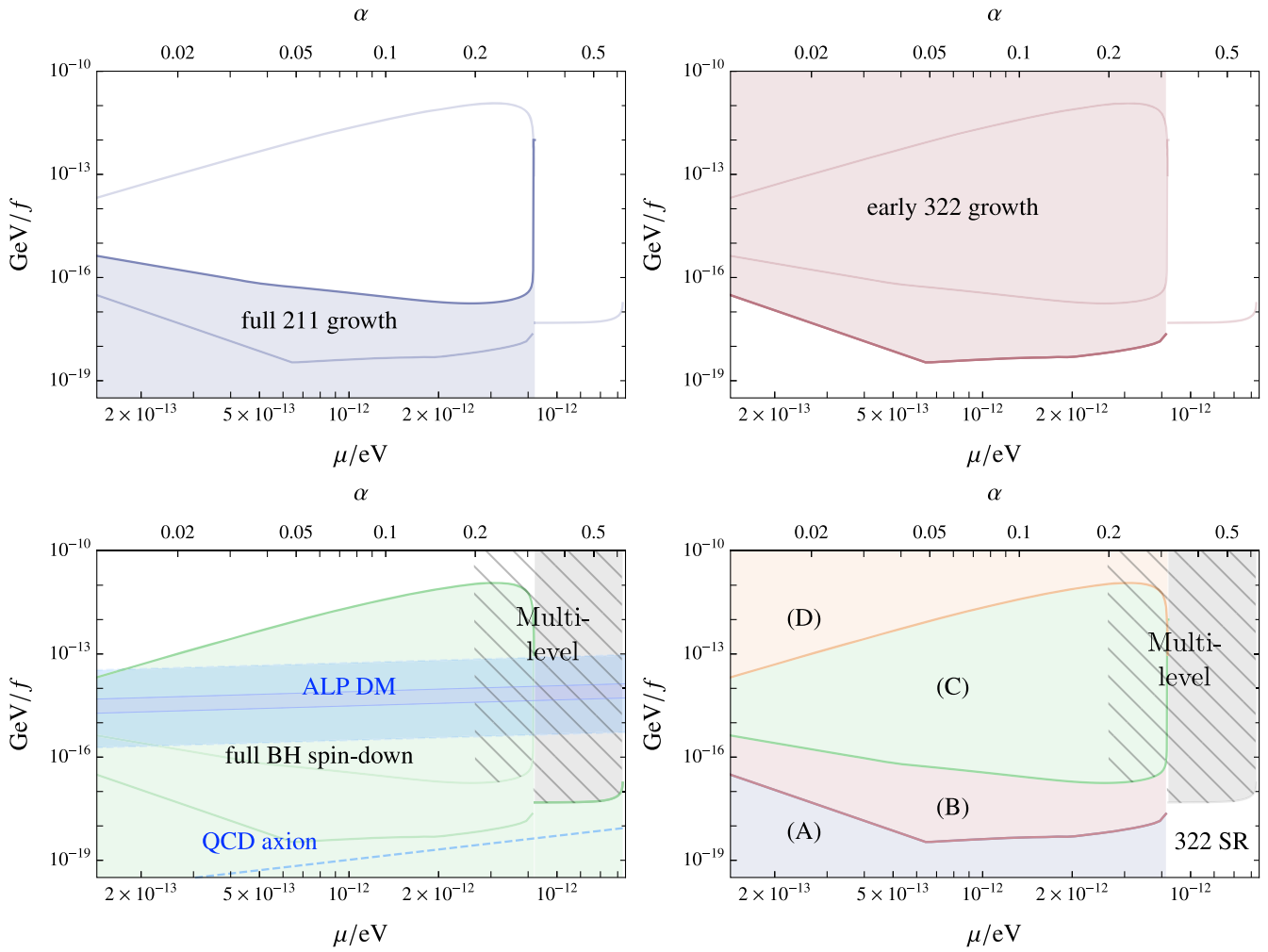


FIG. 3. Parameter space for superradiance of a scalar with mass μ and quartic coupling $\lambda = \mu^2/f^2$, around a BH with $M_{\rm BH} = 10\ M_\odot$ and $a_* = 0.9$ (initially), given a total evolution time of 10^{10} yr. Top-left: parameter space in which the 211 level grows to saturation through superradiance. Top-right: parameter space in which the 322 level grows faster due to self-interactions than it would have through superradiance alone. Bottom-left: parameter space in which the BH is spun down to the threshold of 211 superradiance. For $\mu \gtrsim 4 \times 10^{-12}$ eV (i.e., past the threshold for 211 superradiance), we show the parameter space region in which 322 superradiance is not cut off by self-interactions, and we can be confident that the BH is spun down to the threshold of 322 superradiance. The gray hatched region corresponds to parameter space in which levels other than 211 and 322 are expected to grow; we have not fully analyzed the behavior in these regimes. The blue dashed line corresponds to the quartic coupling for the QCD axion. The "ALP DM" band corresponds to the range of quartic couplings that, for an axion with a time-independent cosine potential, allow the observed DM abundance to be produced by the early universe misalignment mechanism. The darker middle band corresponds to $\mathcal{O}(1)$ values of the initial misalignment angle, while the lighter bands above and below correspond to "tuned" initial values (see Sec. VI A for details). Bottom-right: parameter space regions discussed in the text. (A) corresponds to the "small self-coupling" regime discussed in Sec. IV B 2, (B) corresponds to the "moderate self-coupling" regime discussed in Sec. IV B 3, (C) corresponds to the "large self-coupling" regime discussed in Sec. IV B 4, and (D) corresponds to the "lack of BH spindown" regime discussed in Sec. IV B 5. The "322 SR" region is where 322 superradiance is not cut off by self-interactions, while the gray parameter space above this is when this does occur, and further analysis would be required.

most of this paper, we focus on situations where the initially fastest-growing mode is the 211 level. This grows to maximum size, and spins the BH down to the $m = 1$ superradiance threshold, in a time

$$\frac{\log GM^2}{\Gamma^{\rm SR}_{211}} \simeq \left(\frac{M_{\rm BH}}{10\,M_\odot}\right) \times \begin{cases} 9\ \text{hour} & \alpha = 0.4 \\ 6\times 10^3\ \text{yr}\left(\frac{0.05}{\alpha}\right)^9 & \alpha \lesssim 0.2 \end{cases} \tag{38}$$

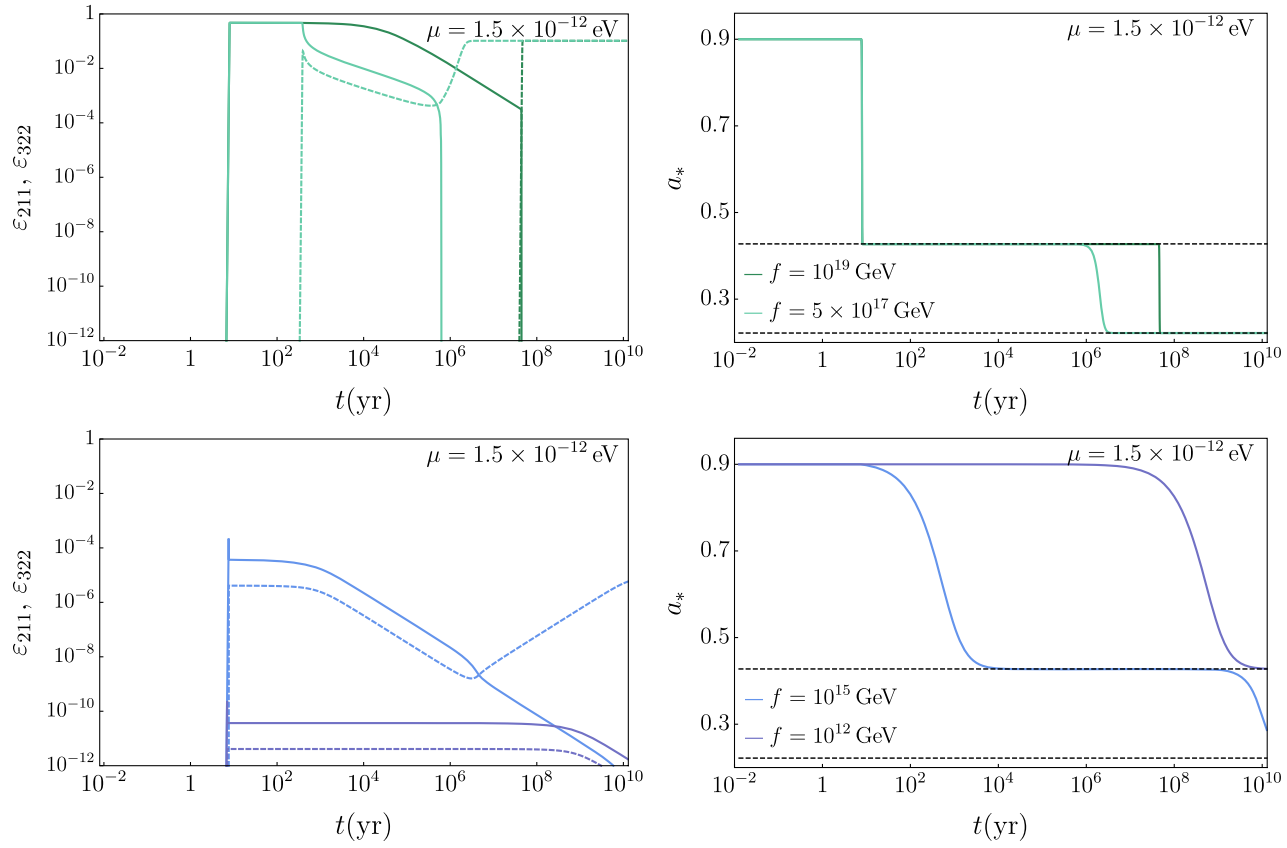


FIG. 4. Left panel: fractional occupation numbers of 211 (solid lines) and 322 (dashed lines) levels, and Right panel: BH spin, as a function of time, for a BH of mass 10 $M_\odot$ and initial spin $a_* = 0.9$, given a scalar of mass $\mu = 1.5 \times 10^{-12}$ eV. The different colors correspond to the different self-interaction strengths indicated in the right-hand plots (see Sec. IV for explanations of the behaviors at different couplings).

for high spin ($a_* = 0.99$). On a timescale that, for small α, is parametrically larger, the 211 level is depleted through gravitational wave annihilations, with a decay time of

$$\tau_{\rm ann} \approx \frac{1}{2\Gamma^{\rm GW}_{211\times211} N_{211,\rm max}} \approx \left(\frac{M_{\rm BH}}{10\ M_\odot}\right) \times \begin{cases} 4 \text{ hour} & \alpha = 0.4, \\ 3\times10^9 \text{ yr}\left(\frac{0.05}{\alpha}\right)^{15} & \alpha \lesssim 0.2. \end{cases} \quad (39)$$

On even longer timescales, the fastest-growing $m = 2$ level (i.e., 322) spins down the BH via superradiance,

$$\frac{\log GM^2}{\Gamma^{\rm SR}_{322}} \simeq \left(\frac{M_{\rm BH}}{10\ M_\odot}\right) \times \begin{cases} 4 \text{ yr} & \alpha = 0.4, \\ 10^{11} \text{ yr}\left(\frac{0.05}{\alpha}\right)^{13} & \alpha \lesssim 0.5. \end{cases} \quad (40)$$

By this point, only a small fraction of the initial 211 occupation generally remains (for α large enough that growth occurs on relevant timescales), so gravitational wave transition signals from $322 \to 211 \times$ GW events are small. The upper panels of Fig. 4 illustrate this evolution, for $f \simeq M_{\rm pl}$. For BHs with long enough lifetimes, a similar story applies to the growth of higher-m levels.

As we discuss below, the purely gravitational story describes the evolution well if the self-interaction-induced $211 \times 211 \to 322 \times$ BH process is always slow compared to superradiant growth processes. The parameter space for which this is true is plotted as region (A) in the bottom-right panel of Fig. 3.

3. Moderate self-coupling: Early growth of 322 and late equilibrium

If we decrease f, while holding other parameters fixed, the first significant difference from purely gravitational evolution that arises is earlier growth of the 322 level. We label this regime, where 211 still grows to saturation, but 322 grows sooner than it would have if $\lambda = 0$, the "moderate self-coupling" regime. The upper-left panel of Fig. 4 illustrates the evolution of the 211 and 322

occupation numbers for an f value in this regime (as well as for a larger f in the small self-coupling regime).

The parameter space for moderate self-coupling is plotted as region (B) in the bottom right-hand panel of Fig. 3, and corresponds to the intersection of the shaded regions in the upper two panels. In this subsection, we will focus on the threshold between the small self-coupling and moderate self-coupling regimes, deferring the small-f boundary of the moderate regime (i.e., the point at which 211 no longer grows to saturation) to the next subsection.

For the $211 \times 211 \to 322 \times \mathrm{BH}$ process to build up 322 within the lifetime of the BH, we need

$$\gamma_{211\times211}^{322\times\mathrm{BH}}(\varepsilon_{211}^{\max})^2 \gtrsim \frac{\log(\varepsilon_{322}^{\mathrm{final}}/\varepsilon_{322}^{\mathrm{initial}})}{T_{\mathrm{BH}}} \simeq \frac{\log(GM_{\mathrm{BH}}^2)}{T_{\mathrm{BH}}} \tag{41}$$

where $\varepsilon_{211}^{\max} \approx a_*(t_0) - a_*^{\mathrm{thresh}} \approx a_*(t_0) - 4/\alpha(1+4\alpha^2)$ is the occupation number of the saturated 211 level. Parametrically, if we start from very small fluctuations in the 322 level, and $\varepsilon_{322}^{\mathrm{final}}$ is not exponentially small, then $\varepsilon_{322}^{\mathrm{final}}/\varepsilon_{322}^{\mathrm{initial}} \sim GM^2$. For this growth to be faster than 322 superradiance, we need $\gamma_{211\times211}^{322\times\mathrm{BH}}(\varepsilon_{211}^{\max})^2 \gtrsim \gamma_{322}^{\mathrm{SR}}$.

The condition (41) is necessary for early 322 growth to occur, but not sufficient, since annihilations to gravitational waves may deplete 211 before 322 can grow. In order for this not to happen, we need

$$\frac{\gamma_{211\times211}^{322\times\mathrm{BH}}(\varepsilon_{211}^{\max})^2}{\log(\varepsilon_{322}^{\mathrm{final}}/\varepsilon_{322}^{\mathrm{initial}})} \gtrsim 2\gamma_{211\times211}^{\mathrm{GW}}\varepsilon_{211}^{\max} \tag{42}$$

Replacing the rates by their small-α expansions, this is equivalent to

$$\frac{\kappa^{\mathrm{BH}}\tilde{r}_+(M_{\mathrm{pl}}/f)^4\varepsilon_{211}^{\max}}{\log(GM^2)} \gtrsim 2\kappa_{211\times211}^{\mathrm{GW}}\alpha^3. \tag{43}$$

The combination of the conditions (41) and (42) is responsible for the shape of the (A)-(B) boundary in Fig. 3. At small α, (41) is more constraining, while at larger α, (42) takes over. The parametric form of this threshold value f_{AB} is given in Table II.

Evolution of levels.—Unlike in the gravitational scenario, where the growth of 322 via superradiance is accompanied by a rapid drop in 211 occupation, here both levels eventually reach roughly comparable occupation numbers. Subsequently, the joint cloud is slowly depleted by the combination of nonrelativistic scalar emission and damping by the BH. Other processes, including gravitational annihilations and transitions as well as relativistic scalar emission, are small perturbations to this overall evolution.

As discussed above, only a few rates drive the dynamics in the regions of parameter space for which self-interactions modify the purely gravitational scenario. These are $\kappa_{211}^{\mathrm{SR}}$, $\kappa_{211\times211}^{322\times\mathrm{BH}}$, and $\kappa_{322\times322}^{211\times\infty}$ (and $\kappa_{322}^{\mathrm{SR}}$, in some circumstances). To streamline our notation, we will refer to them as κ^{SR}, κ^{BH}, and κ^{∞} respectively.

In the regime of moderate self-coupling, the growth of the 211 level occurs as in the purely gravitational case; both the occupation number and the BH angular momentum change "suddenly," with almost all of the change happening in the last few e-folds of superradiant growth. This is illustrated in the top panels of Fig. 4. The BH spin decreases to $a_* \approx 4\alpha/(1+4\alpha^2)$, and ε_{211} stays at $\approx\varepsilon_{211}^{\max}$ for a long time. In the purely gravitational scenario, the cloud would then slowly self-annihilate to gravitational waves until $\sim$200 e-folds of 322 superradiance have passed. Here, however, the quartic process dominates, and the 322 growth rate is higher:

$$\frac{\dot{\varepsilon}_{322}}{\mu} \approx \kappa^{\mathrm{BH}}\tilde{r}_+\alpha^{11}(M_{\mathrm{pl}}/f)^4(\varepsilon_{211}^{\max})^2\varepsilon_{322}. \tag{44}$$

Eventually, the 322 occupation number becomes large enough that the quartic vertex $322 \times 322 \to 211 \times \infty$ becomes important and a quasiequilibrium is established, roughly after time

$$t_* \simeq GM\frac{\log(GM^2)}{\kappa^{\mathrm{BH}}\tilde{r}_+\alpha^{12}(M_{\mathrm{pl}}/f)^4(\varepsilon_{211}^{\max})^2} \tag{45}$$

has passed.

At this point, superradiance to 211 has effectively shut down, and 322 superradiance is too slow to be significant. Particles are leaving the combined cloud, going back to the BH (via $211 \times 211 \to 322 \times \mathrm{BH}$) and to infinity (via $322 \times 322 \to 211 \times \infty$). Gravitational and relativistic scalar processes are suppressed by high powers of α. Accordingly, the coupled dynamics of the two-level system simplifies to

$$\frac{\dot{\varepsilon}_{211}}{\mu} \approx -2\kappa^{\mathrm{BH}}\tilde{r}_+\alpha^{11}(M_{\mathrm{pl}}/f)^4\varepsilon_{211}^2\varepsilon_{322} + \kappa^{\infty}\alpha^8(M_{\mathrm{pl}}/f)^4\varepsilon_{322}^2\varepsilon_{211}, \tag{46a}$$

$$\frac{\dot{\varepsilon}_{322}}{\mu} \approx \kappa^{\mathrm{BH}}\tilde{r}_+\alpha^{11}(M_{\mathrm{pl}}/f)^4\varepsilon_{211}^2\varepsilon_{322} - 2\kappa^{\infty}\alpha^8(M_{\mathrm{pl}}/f)^4\varepsilon_{322}^2\varepsilon_{211}, \tag{46b}$$

$$a_* \approx \frac{4\alpha}{1+4\alpha^2}. \tag{46c}$$

Since there are no processes (except for the negligible superradiance of 322) which contribute particles to the

cloud, particles are only leaving. Accordingly, the system has no true equilibrium occupations. However, (46) still admits a time-independent equilibrium *ratio* of occupation numbers, $\varepsilon_{322}/\varepsilon_{211} = \eta^B$ to which the system flows,

$$\eta^B \simeq \frac{1}{2}\frac{\kappa^{\rm BH}\alpha^3\tilde{r}_+}{\kappa^\infty} \simeq 4\times 10^{-5}\left(\frac{\alpha}{0.01}\right)^3. \tag{47}$$

For the regime of moderate self-coupling, the scalings in (47) are only representative at leading orders in α. A more accurate expression is derived in Appendix D.

When the equilibrium ratio is obtained at time t_*, the occupations evolve as

$$\varepsilon_{211}(t) \simeq \frac{\varepsilon_{211}(t_*)}{\sqrt{1+2\varepsilon_{211}^2(t_*)(t-t_*)/\tau_{\rm scalar}}}, \tag{48}$$

where

$$\tau_{\rm scalar} \equiv \frac{4}{3\mu}\frac{\kappa^\infty}{(\kappa^{\rm BH}\tilde{r}_+)^2\alpha^{14}}\left(\frac{f}{M_{\rm pl}}\right)^4 \approx 10^{-1}\ {\rm yr}\left(\frac{0.1}{\alpha}\right)^{14}\left(\frac{10^{-12}\ {\rm eV}}{\mu}\right)\left(\frac{f}{10^{17}\ {\rm GeV}}\right)^4, \tag{49}$$

and $\varepsilon_{322}(t) = \varepsilon_{211}(t)\eta^B$.

The joint cloud continues to deplete until the occupation of 211 has diminished enough that the superradiance rate of 322 outcompetes the "stimulated" emission process $322\times 322 \to 211\times\infty$, and the cloud starts growing again. A large occupation builds up in 322, causing rapid 211 depletion via $211\times 211 \to 322\times{\rm BH}$. Moreover, as superradiance extracts angular momentum from the BH to 322, the BH's spin decreases further, making 211 (and other $m=1$ states) *damped*. This sequence of events is illustrated in the top panels of Fig. 4 (where the green curves correspond to moderate self-coupling, and the blue to small self-coupling).

In the $\lambda = 0$ case, $m=2$ superradiance must proceed from zero-point quantum fluctuations, or from a small preexisting astrophysical density. Here, superradiance gets to act on the *preexisting* occupation ε_{322}, since 322 has already been populated by self-interaction-mediated processes. In this way, self-interactions "assist" superradiance, sometimes leading to more rapid saturation of the $m=2$ instability than allowed in the purely gravitational story. The $f = 5\times 10^{17}$ GeV curves in the upper panels of Fig. 4 show an example of this, with 322 spin-down occurring after only ~few $\times 10^6$ yr, compared to almost 10^8 yr in the purely gravitational case.

The above discussion summarizes the evolution of the cloud in the moderate self-coupling regime. Before moving on, we will discuss the effects of processes other than $211\times 211\to 322\times{\rm BH}$, $322\times 322\to 211\times\infty$, and superradiance, and review why they are (in most cases) subdominant.

Annihilations to GWs.—An important point is that, to be in the moderate self-coupling regime for astrophysical BH masses, we need $f \lesssim M_{\rm pl}$ (as illustrated in Fig. 3). This is evident from the form of the threshold $f_{\rm AB}$ given in Table II, $f_{\rm AB} = \min(f_1, f_2)$. The first term f_1 comes from the condition $\gamma_{211\times 211}^{322\times{\rm BH}}(\varepsilon_{211}^{\rm max})^2 \gtrsim \frac{\log(GM_{\rm BH}^2)}{T_{\rm BH}}$; to make $f_1 \geq M_{\rm pl}$, we need to take $\alpha \gtrsim 0.07$ [for $M_{\rm BH} = \mathcal{O}(10\ M_\odot)$]. Such large values of α make the f_2, coming from the condition that GW annihilations are not too fast (42), much less than $M_{\rm pl}$. Consequently, gravitational wave emission processes suffer a *suppression* $\sim(f/M_{\rm pl})^4$, relative to self-interaction-mediated quartic processes. This means that, once 322 has reached its equilibrium ratio with 211 [Eq. (47)], even the fastest GW emission process, $211\times 211\to{\rm GW}$, is generally slower than $211\times 211\to 322\times{\rm BH}$ and $322\times 322\to 211\times\infty$ (at least until the levels have depleted significantly).

GW transitions.—From Table IV, gravitational wave transitions $322\to 211+{\rm GW}$ contribute a term $\dot{\varepsilon}_{322} \simeq -3\times 10^{-6}\alpha^{10}\varepsilon_{322}\varepsilon_{211}\mu + \cdots$ to the evolution equations. If we take $\varepsilon_{322} = \eta^B\varepsilon_{211}$ [Eq. (47)], this gives

$$\dot{\varepsilon}_{322}/\mu \simeq -3\alpha^{13}\varepsilon_{211}^2 + 0.4\alpha^{14}\varepsilon_{211}^3\left(\frac{M_{\rm pl}}{f}\right)^4 + \ldots \tag{50}$$

where we have also included the $211\times 211\to 322\times{\rm BH}$ term for comparison. While the GW transition term is suppressed by one less power of α, Fig. 3 illustrates that, as α decreases, the maximum f for the moderate self-coupling regime decreases (from Table II, $f_{AB}\propto\alpha^{11/4}$ for small α). Consequently, the relative $(M_{\rm pl}/f)^4$ enhancement of the quartic self-interaction terms always wins out.

Even though gravitational wave emission no longer dominates the evolution compared to the small self-interactions regime of gravitational superradiance, GW annihilation signals can still be strong enough for detection in this regime. In addition, the simultaneous occupation of the two levels allows for the possibility of GW signals from transitions. We explore potential signatures in more detail in Sec. VII.

Relativistic $3\to 1$ *emission.*—As discussed in Sec. III A, quartic self-interactions also lead to processes emitting relativistic scalar waves, such as $211\times 211\times 211\to\infty$. This contributes

$$\dot{\varepsilon}_{211}/\mu \simeq -5\times 10^{-9}\alpha^{21}\left(\frac{M_{\rm pl}}{f}\right)^4\varepsilon_{211}^3 + \ldots \tag{51}$$

Because of the high power of α this is suppressed by, its effect is small compared to the nonrelativistic quartic processes.

Relativistic cubic emission.—In Sec. III D, we discussed how, in addition to a quartic self-interaction, there may also be a cubic interaction term, $\mathcal{L} \supset \frac{1}{6} C \frac{\mu^2}{f} \varphi^3$, which can lead to relativistic emission processes such as $211 \times 211 \to \infty$. This contributes

$$\dot{\varepsilon}_{211}/\mu \simeq -2 \times 10^{-4} \alpha^{14} |C|^2 \left(\frac{M_{\rm pl}}{f}\right)^2 \varepsilon_{211}^2 + \ldots \quad (52)$$

Compared to the quartic-induced term in Eq. (50), the lower power of $M_{\rm pl}/f$, and the smaller constant factor, mean that unless $|C| \gg 1$, relativistic emission from the cubic coupling will be a subdominant effect.

4. Large self-coupling: early equilibrium and halted extraction of angular momentum

If we further decrease f, we reach a point where 322 grows large enough, early enough, that 211 superradiance is disrupted, and 211 does not reach its saturation value. We call this the regime of "large self-coupling"; it corresponds to regions (C) and (D) in the bottom-right panel of Fig. 3, and to the bottom panels in Fig. 4.

For the $211 \times 211 \to 322 \times {\rm BH}$ process to disrupt 211 superradiance, we need that $2\gamma_{211\times211}^{322\times{\rm BH}} \varepsilon_{211} \varepsilon_{322} \gtrsim \gamma_{211}^{\rm SR}$ before ε_{211} has grown to its saturation value. This does not necessarily preclude 211 reaching $\varepsilon_{211}^{\rm max}$ (ε_{211} can still grow after that point, albeit more slowly than it would have with $\lambda = 0$), but it is necessary to have a significant effect. Parametrically, this condition is approximately equivalent to

$$\gamma_{211\times211}^{322\times{\rm BH}} (\varepsilon_{211}^{\rm max})^2 \gtrsim 2 \log (GM^2) \gamma_{211}^{\rm SR}, \quad (53)$$

where we neglect the dependence of the rates on the BH spin [i.e., set $a_*(t) = a_*(t_0)$]. A more precise condition is derived in Appendix E.

The condition (53) can be expressed as a condition on f. 211 superradiance is basically unaffected if $f \gtrsim f_{\rm thresh}$, where

$$f_{\rm thresh} \approx M_{\rm pl} \left(\frac{\alpha^3}{2\log(GM^2)} \frac{\kappa^{\rm BH} \tilde{r}_+ (\varepsilon_{211}^{\rm max})^2}{\kappa^{\rm SR} a_*(t_0)}\right)^{1/4} \approx 6 \times 10^{15}\ {\rm GeV} \left(\frac{\alpha}{0.01}\right)^{3/4} \left(\frac{a_*(t_0)}{0.9}\right)^{1/4}. \quad (54)$$

The scalings in (54) are only representative when $\alpha \ll a_*(t_0)$. For larger values of α, rates obtained numerically, and a more precise version of (53) (Appendix E), can be used.

As pointed out in [31], if a_* is held fixed, the system admits equilibrium occupations for which $\dot{\varepsilon}_{211} = \dot{\varepsilon}_{322} = 0$:

$$\varepsilon_{211}^{\rm eq}(a_*) \approx \frac{2}{\sqrt{3}} \frac{\sqrt{\kappa^\infty \kappa^{\rm SR}(a_* - 2\alpha\tilde{r}_+)}}{\alpha^3 \kappa^{\rm BH} \tilde{r}_+} \left(\frac{f}{M_{\rm pl}}\right)^2 \equiv \left(\frac{f}{f_{\rm eq}}\right)^2 \varepsilon_{211}^{\rm max} = 2.5 \times 10^{-1} \left(\frac{0.01}{\alpha}\right)^3 \left(\frac{a_*}{0.9}\right)^{1/2} \left(\frac{f}{10^{15}\ {\rm GeV}}\right)^2, \quad (55a)$$

$$\varepsilon_{322}^{\rm eq}(a_*) \approx \sqrt{\frac{1}{3} \frac{\kappa^{\rm SR}(a_* - 2\alpha\tilde{r}_+)}{\kappa^\infty}} \left(\frac{f}{M_{\rm pl}}\right)^2 = 6.9 \times 10^{-6} \left(\frac{a_*}{0.9}\right)^{1/2} \left(\frac{f}{10^{15}\ {\rm GeV}}\right)^2, \quad (55b)$$

where

$$f_{\rm eq} \approx M_{\rm pl} \left(\frac{\sqrt{3}}{2} \frac{\alpha^3 \kappa^{\rm BH} \tilde{r}_+ \varepsilon_{211}^{\rm max}}{\sqrt{\kappa^{\rm SR} \kappa^\infty (a_* - 2\alpha\tilde{r}_+)}}\right)^{1/2} \approx 2 \times 10^{15}\ {\rm GeV} \left(\frac{\alpha}{0.01}\right)^{3/2} \left(\frac{a_*(t_0)}{0.9}\right)^{1/4}. \quad (56)$$

Note that the ratio $\eta^{\rm eq} \equiv \varepsilon_{322}^{\rm eq}/\varepsilon_{211}^{\rm eq}$ is

$$\eta^{\rm eq} = \frac{\gamma^{\rm BH}}{2\gamma^\infty} \approx (\eta^B)_{{\rm small}\alpha}, \quad (57)$$

according to the approximation (47) valid for small α. At larger values of α, $\eta^B > \eta^{\rm eq}$. See Appendix D for more details.

We now consider what happens in the physical case, where a_* can change. If $\varepsilon_{211}^{\rm eq}$ is much less than its saturation value, then the timescale to extract an $\mathcal{O}(1)$ fraction of the BH's spin is much longer than the characteristic timescale of the processes maintaining the equilibrium. Consequently, we expect the quasiequilibrium to be maintained to a good approximation, as a_* undergoes a slow descent. The equilibrium occupation numbers $\varepsilon_{211}^{\rm eq}(a_*)$ and $\varepsilon_{322}^{\rm eq}(a_*)$ stay almost constant, with the angular momentum extracted from the BH via 211 superradiance being emitted to infinity via the $322 \times 322 \to 211 \times \infty$ process. This is in contrast to the regimes of small and moderate self-interactions, where the angular momentum lost from the BH builds up in the cloud.

Close to the transition from moderate to large self-interactions, there is a sliver of parameter space for which the exponential growth of 211 is maintained for some time and $\mathcal{O}(1)$ of the maximum spin extraction occurs, before getting cut short by the equilibrium. Deep inside the region of small f, however, the spin of the BH is essentially

unchanged at the time the equilibrium is established, and most of the extraction of angular momentum happens adiabatically.

Although (55) is valid at equilibrium, if α is large enough then ε_{211} will "overshoot" its equilibrium value before ε_{322} has caught up with it. Before equilibrium, if we neglect the dependence of $\gamma_{211}^{\rm SR}$ on the BH spin, $\varepsilon_{211} \propto \exp(\gamma_{211}^{\rm SR} t)$. In Appendix E, we derive an estimate for the value of the exponent $\gamma_{211}^{\rm SR} t$ at the time when $211 \times 211 \to 322 \times {\rm BH}$ is comparable to SR. To a good approximation

$$\varepsilon_{211}^{\rm thresh} \approx \sqrt{\frac{2\gamma_{211}^{\rm SR} \log(GM^2)}{\gamma_{211\times 211}^{322\times {\rm BH}}}} \approx \left(\frac{f}{f_{\rm thresh}}\right)^2 \varepsilon_{211}^{\rm max}, \quad (58)$$

where we set $a_*(t) = a_*(t_0)$ in both rates.

Accordingly, the evolution toward equilibrium can happen in two qualitatively different ways. When $\alpha \gtrsim 0.04$, $f_{\rm thresh} < f_{\rm eq}$ and $\varepsilon_{211}^{\rm thresh} > \varepsilon_{211}^{\rm eq}$. In this case, the occupation ε_{211} overshoots its equilibrium value and subsequently evolves toward it from above. This is illustrated in the bottom-left panel of Fig. 4 (for which $\alpha = 0.11$). Conversely, when $\alpha \lesssim 0.04$, then $\varepsilon_{211}^{\rm thresh} < \varepsilon_{211}^{\rm eq}$. There is no overshoot, and ε_{211} evolves toward its equilibrium occupation from below.

Given this, the boundary between the moderate self-coupling regime, where ε_{211} reaches $\varepsilon_{211}^{\rm max}$, and large self-coupling, where it does not, is set by

$$f \lesssim f_{\rm BC} \equiv \min[f_{\rm thresh}, f_{\rm eq}]. \quad (59)$$

To review, the evolution of the superradiant cloud, in the regime of large self-coupling, occurs in different stages:

(1) An initial stage of exponential 211 growth, during which ε_{322} is too small to significantly affect the evolution of ε_{211}.
(2) A "nonequilibrium" stage in which ε_{211} and ε_{322} evolve toward their equilibrium values. The timescale to approach the equilibrium values is at most a logarithmic multiple of $1/\gamma_{211}^{\rm SR}$, since the relevant self-interaction processes are at least as fast as $\gamma_{211}^{\rm SR}$.
(3) Once ε_{211} and ε_{322} are close to their equilibrium values, there is a long period of quasi-adiabatic evolution. The spin-down of the BH due to spin extraction through 211 superradiance, which changes a_* on a timescale $(\dot{a}_*/a_*)^{-1} \sim (\varepsilon_{211}^{\rm max}/\varepsilon_{211}^{\rm eq})/\gamma_{211}^{\rm SR}$, leads to the slow evolution of the equilibrium occupation numbers.
(4) If the BH lifetime is long enough that spin-down to the $m = 1$ threshold occurs, then similar behavior to the moderate self-coupling regime will result. The 211 and 322 levels will maintain a quasiequilibrium ratio, but with decreasing occupation numbers, as scalars are emitted to infinity. Eventually, the occupation numbers will become small enough that 322 superradiance starts to dominate, at which point the 322 occupation number starts growing again (e.g., the $f = 10^{15}$ GeV curves in the bottom-left panel of Fig. 4).

Consequently, when f is appreciably smaller than $f_{\rm BC}$, the first and second stages change a_* by only a small amount, and the majority of the BH's spin-down to the $m = 1$ threshold happens during the period of almost adiabatic, quasiequilibrium evolution.

When the equilibrium occupations (55) are obtained, the angular momentum of the BH decreases according to (36), with $\varepsilon_{211} = \varepsilon_{211}^{\rm eq}(a_*)$ (and we can ignore $\kappa_{322}^{\rm SR}$). The timescale for spindown is therefore set by

$$\begin{aligned}\tau_{\rm sd}(a_*) &\approx \frac{\sqrt{3}}{2\alpha^5\mu} \frac{\kappa^{\rm BH}\tilde{r}_+ (M_{\rm pl}/f)^2}{\sqrt{\kappa^\infty}(\kappa^{\rm SR}(a_* - 2\alpha\tilde{r}_+))^{3/2}} \\ &\approx 10^7 \text{ yr} \left(\frac{0.01}{\alpha}\right)^5 \left(\frac{10^{-12}\text{ eV}}{\mu}\right) \\ &\quad \times \left(\frac{0.9}{a_*}\right)^{\frac{3}{2}} \left(\frac{10^{15}\text{ GeV}}{f}\right)^2. \end{aligned} \quad (60)$$

While in (slowly varying) equilibrium, the cloud emits nonrelativistic axion waves through the $322 \times 322 \to 211 \times \infty$ process. These could, in the presence of axion-SM interactions, be detected by experiments on Earth. Even though the occupation number of the cloud decreases $\propto f^2$ for small f, the coupling strength of axion-SM interactions will generically scale as $\sim 1/f$. Consequently, the interaction rate of the emitted radiation with a laboratory target can be *independent* of f in the small-f regime. This in contrast to gravitational wave signals, which are suppressed at small f. We discuss this possibility more fully in Sec. VIII.

In the previous subsection on the moderate self-coupling regime, we discussed how interaction processes, other than nonrelativistic quartic interactions and superradiance, are generally subdominant in their effects on the evolution of the cloud. Very similar calculations apply to the large self-coupling regime; the equilibrium ratio of $\varepsilon_{322}/\varepsilon_{211}$ is the same, with the difference being that the equilibrium occupation numbers are suppressed, scaling $\propto f^2$.

This scaling only makes a difference to comparisons between processes with different multiplicities. For annihilation to GWs, the $(f/M_{\rm pl})^2$ scaling of the occupation number is not enough to make up for the $(M_{\rm pl}/f)^4$ relative enhancement of the quartic interaction rates, so GW annihilation processes are even less important than they are in the moderate self-coupling regime.

For relativistic cubic emissions, the fastest of which is $211 \times 211 \to \infty$, we can compare the contribution to the evolution rate to that from $211 \times 211 \to 322 \times {\rm BH}$:

$$\dot{\varepsilon}_{211}/\mu \simeq -2\times 10^{-4}\alpha^{14}|C|^2\left(\frac{M_{\rm pl}}{f}\right)^2\varepsilon_{211}^2 - 8\times 10^{-7}\alpha^{11}\left(\frac{M_{\rm pl}}{f}\right)^4\varepsilon_{211}^2\varepsilon_{322} \simeq (-2\times 10^{-4}\alpha^{14}|C|^2 - 10^{-3}\alpha^{11})\times\left(\frac{M_{\rm pl}}{f}\right)^2(\varepsilon_{211}^{\rm eq})^2 \quad (61)$$

where the second equality applies for the equilibrium occupation numbers (55). Consequently, if $|C|\lesssim 16(0.2/\alpha)^{3/2}$, then the effect of the cubic emission term is small compared to that of the nonrelativistic quartic processes.

For $\alpha \gtrsim 0.04$, the equilibrium values of ε_{211} and ε_{322} are smaller than the "overshoot" values at which self-interactions first affect the evolution of 211. Consequently, if the relativistic cubic processes are unimportant in equilibrium, then they are always less important than the quartic $211\times 211\to 322\times$ BH process, whenever the latter has a significant effect on 211 evolution.

For smaller α, the $2\to 1$ process will be relatively most important around the initial time at which 211 growth is slowed down, since the equilibrium occupation numbers are approached from below. Still, even without calculating the thresholds carefully, we can see that as long as $|C|\lesssim 16(0.2/0.04)^{3/2}\simeq 180$, cubic emission will be insignificant in that regime (since decreasing α decreases the relative importance of cubic emission). Overall, we can see that, unless $|C|\gg 1$, relativistic emission through the cubic coupling should always be a subdominant effect on the evolution of the 211 level (cubic emission for higher-l levels is suppressed by higher powers of α, so should generally be less significant again).

5. Large self-coupling: lack of BH spindown

Since $\varepsilon_{211}^{\rm eq}\propto f^2$, and the rate of spin extraction from the BH is $\propto\varepsilon_{211}$, the spin-down rate for small enough f will be so slow that the $m=1$ threshold spin is not reached within the BH lifetime. The $f=10^{12}$ GeV curves in the bottom panels of Fig. 4 show an example, if we take the BH lifetime to be $<10^{10}$ yr. This affects BH spin-down signatures of superradiance, as we discuss in Sec. VI.

The timescale for spin extraction in the large self-coupling regime is set by $\tau_{\rm sd}$ [Eq. (60)]. Setting this equal to the age $T_{\rm BH}$ of the BH gives the threshold value of f

$$f_{\rm CD}\approx 3\times 10^{14}\ {\rm GeV}\left(\frac{10^{10}\ {\rm yr}}{T_{\rm BH}}\right)^{\frac{1}{2}}\left(\frac{10^{-13}\ {\rm eV}}{\mu}\right)^{\frac{1}{2}}\times\left(\frac{0.01}{\alpha}\right)^{\frac{5}{2}}\left(\frac{0.9}{a_*(t_0)}\right)^{\frac{3}{4}} \quad (62)$$

i.e., if $f\lesssim f_{\rm CD}$, then the BH does not have time to fully spin down. The parameter space in which this is the case is plotted as region (D) in the bottom-right panel of Fig. 3, and is illustrated by the smallest-f curve in Fig. 9. For $f\ll f_{\rm CD}$, which gives $T_{\rm BH}\ll\tau_{\rm sd}$, the amount of angular momentum extracted is

$$|\Delta a_*|\simeq\frac{T_{\rm BH}}{\tau_{\rm sd}(a_*(t_0))}. \quad (63)$$

C. Beyond the two-level system

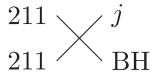
So far, we have focussed on BH-cloud systems which are dominated by the 211 and 322 hydrogenic levels. In this subsection, we consider the effect of other levels on the dynamics, including higher principal number n and higher angular momentum numbers l, m. We continue to assume that the initial conditions are such that 211 satisfies the superradiance condition and is the first level to grow; this is the regime of fastest black hole spindown and the largest gravitational and scalar emission rates, and is thus the most relevant from an observational perspective.

We find that, for $\alpha\lesssim 0.2$, the two-level picture discussed so far is probably sufficient, with only 211 and 322 growing to large occupation numbers. For $\alpha\gtrsim 0.2$, we expect that self-interactions would cause other levels to grow; we leave a full analysis of this regime to future work.

Our analysis in this section focusses on perturbative processes, assuming that evolution is well-approximated by a combination of approximately hydrogenic levels. In Sec. V, we investigate whether nonperturbative processes, such as "bosenova," could change this picture; we find that, for $\alpha\lesssim 0.2$, this seems rather unlikely.

1. Growth mechanisms in the presence of self-interactions

As discussed in Sec. IV A, if 211 is initially the only state with appreciable occupation number, then other states j can be built up through processes of the form



Taking $j=322$ gives the fastest growth rate, since the forced oscillation damped by the BH has $m=0$ (maximizing the damping rate), and the overlap factors are large.

If a 322 and 211 abundance are both present, then other states can also be built up through

However, as well as these processes building up new states, there are also processes reducing their abundance;

To determine whether, starting from very small fluctuations, another level j will start growing, we can look at the linear-in-ε_j evolution terms (i.e., ignore processes such as the last diagram), and see whether the growth rate is positive or negative.

2. $n11$ levels

For a state j with $m = 1$, the quartic processes with j in the final state all have forced oscillations with $m \geq 1$, which are growing rather than decaying (in the parameter space where 211 is superradiant). Consequently, they contribute a negative term to j's growth rate. Hence, growth of j can only come about through superradiance.

In the large self-coupling regime, a quasiequilibrium for 211 and 322 can be reached with very little effect on the BH spin, so the superradiance rates for $m = 1$ states are still positive. The fastest such rates are for the $n11$ states. The linear-order evolution of the occupation number is set by

$$\frac{\dot{\varepsilon}_{n11}}{\varepsilon_{n11}} = \gamma^{\rm SR}_{n11} - (\gamma^{322\times{\rm BH}}_{211\times n11} + \gamma^{211\times\infty}_{n11\times 322})\varepsilon_{211}\varepsilon_{322}. \quad (64)$$

Substituting in the equilibrium values for ε_{211} and ε_{322}, we have

$$\frac{\dot{\varepsilon}_{n11}}{\gamma^{\rm SR}_{n11}\varepsilon_{n11}} \simeq 1 - \frac{2}{3}\frac{\gamma^{\rm SR}_{211}}{\gamma^{\rm SR}_{n11}}\frac{\gamma^{322\times{\rm BH}}_{211\times n11} + \gamma^{211\times\infty}_{n11\times 322}}{\gamma^{322\times{\rm BH}}_{211\times 211}} \quad (65)$$

It is useful to analyse the large-n behavior of this expression. At leading order in small α, the ratio $\frac{\gamma^{\rm SR}_{211}}{\gamma^{\rm SR}_{n11}}\frac{\gamma^{322\times{\rm BH}}_{211\times n11}}{\gamma^{322\times{\rm BH}}_{211\times 211}}$ is independent of α and a_*; it exceeds 1 for $n \gtrsim 10$, and approaches 1.27 at large n (see Appendix C 2 a and Fig. 22). As discussed in Sec. III C, the most important finite-α effects on the quartic BH rates arise via the horizon flux of the associated forced oscillation. Since they are driven by near-horizon behavior, these do not have large effects on *ratios* of rates (Fig. 21). Consequently, the ratio of analytic superradiance rates should be accurate at the few-percent level, except close to the superradiance boundary.

The ratio $\frac{\gamma^{\rm SR}_{211}}{\gamma^{\rm SR}_{n11}}\frac{\gamma^{211\times\infty}_{n11\times 322}}{\gamma^{322\times{\rm BH}}_{211\times 211}}$ scales as α^{-3} at small α. For n large, it approaches

$$\frac{2\gamma^{\rm SR}_{n11}}{3\gamma^{\rm SR}_{211}}\frac{\gamma^{211\times\infty}_{n11\times 322}}{\gamma^{322\times{\rm BH}}_{211\times 211}} \to \left(\frac{0.29}{\alpha\tilde{r}_+^{1/3}}\right)^3, \quad n\to\infty, \quad \alpha \ll 1 \quad (66)$$

(see Appendix C 2 a and Fig. 23).

The combination of these negative contributions means that no $n11$ level with $n \gtrsim 6$ gets populated, at least for $\alpha\tilde{r}_+ \lesssim 0.3$.[6] For $n = 3$, the process $211 \times 311 \to 322 \times {\rm BH}$ is resonant, as discussed in Sec. III C; this makes it more difficult to populate 311. However, for $\alpha \gtrsim 0.2$, we expect that the 411 level will grow, given enough time. This is illustrated in Fig. 5.

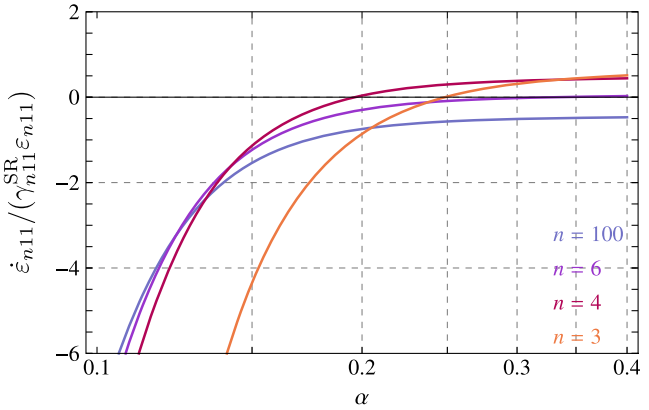


FIG. 5. Growth rates of $n11$ levels once 211/322 quasiequilibrium has been reached, relative to their superradiance rates. At $\alpha \lesssim 0.2$ none of the levels have positive growth rates; levels with $n \gtrsim 10$ have negative growth rates for all α, within our hydrogenic approximation.

Since the 411 superradiance rate is $\mathcal{O}(10)$ smaller than that of 211, the evolution of the 211/322 two-level system should proceed, at first, without modifications. Therefore, in the moderate and large self-coupling regimes we are considering, 211 and 322 will reach their two-level quasiequilibrium occupation numbers, as described in Sec. IV B. After two-level quasiequilibrium is reached, we can initially treat 211 and 322 as constant sources while 411 grows (since the BH spin-down timescale is relatively very long). As a result, 411 grows with an "effective" superradiance rate which is smaller than its usual superradiance rate,

$$\gamma^{\rm SR-eff}_{411} \equiv \gamma^{\rm SR}_{411} - (\gamma^{322\times{\rm BH}}_{211\times 411} + \gamma^{211\times\infty}_{411\times 322})\varepsilon^{\rm eq}_{211}\varepsilon^{\rm eq}_{322} \quad (67)$$

where the quasiequilibrium concentrations are given by Eqs. (55a) and (55b).

After $\mathcal{O}(100)$ e-folds, the occupation number of 411 will become comparable to those of 211 and 322, and the three levels reach a new quasiequilibrium. The most striking feature of this is that the equilibrium 411 occupation number is significantly higher than the equilibrium occupation numbers in the two-level 211/322 equilibrium. The 411 evolution equation is

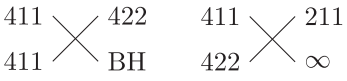
[6] If α is large enough that we are in the "overshoot" regime, where the maximum occupation numbers are reached before the equilibrium phase, the negative contributions to the growth rate during the overshoot are even larger than in equilibrium.

$$\frac{\dot{\varepsilon}_{411}}{\varepsilon_{411}} \simeq \gamma^{\rm SR}_{411} - (\gamma^{322\times{\rm BH}}_{211\times411} + \gamma^{211\times\infty}_{411\times322})\varepsilon_{211}\varepsilon_{322} - \gamma^{322\times{\rm BH}}_{411\times411}\varepsilon_{322}\varepsilon_{411}$$
$$= \gamma^{\rm SR-eff}_{411} - \gamma^{322\times{\rm BH}}_{411\times411}\varepsilon_{322}\varepsilon_{411}$$

Since the numerical coefficient of the $\gamma^{322\times{\rm BH}}_{411\times411}$ rate is significantly smaller than e.g., that of $\gamma^{322\times{\rm BH}}_{211\times411}$ (see Table V), then unless $\gamma^{\rm SR-eff}_{411}$ is significantly smaller than the components of Eq. (67), we need $\varepsilon^{\rm eq}_{411} \gg \varepsilon^{\rm eq}_{211,322}$ to compensate. This is illustrated in Fig. 6, which shows the growth of 411, and development of a new three-level equilibrium, for $\alpha \simeq 0.22$. From numerical calculations, 411 grows to be up to ~50 times larger than the benchmark two-level quasiequilibrium value of 211 [Eq. (55a)].

Given this enhanced occupation number, it is natural to ask whether higher-order or nonperturbative processes could occur, even if they do not for the two-level system. As discussed in Sec. V, the more spread-out wave function of the 411 level makes this unlikely. The emission of scalar radiation will also be enhanced, as discussed in Sec. VIII.

This three-level quasiequilibrium is unlikely to be the full story. As we discuss in the next section, within the two-level equilibrium, we do not expect $n22$ levels to grow. However, the large value of $\varepsilon^{\rm eq}_{411}$ can change this conclusion. For example, the dominant processes building up and depleting the 422 level, in the presence of equilibrium 211, 322 and 411 occupations, are



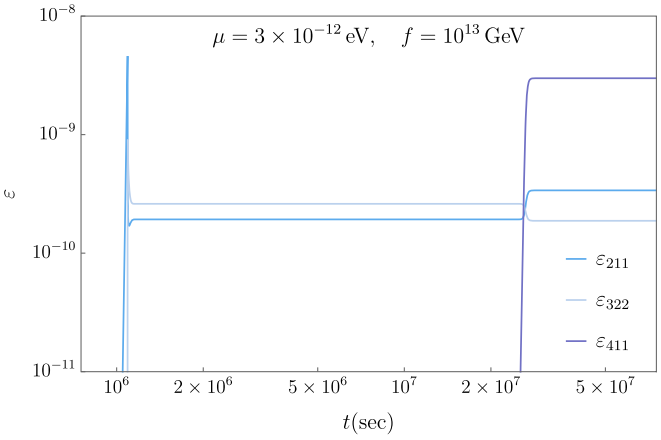


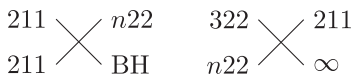
FIG. 6. Example of 411 level growth after a period of 211/322 quasiequilibrium. This plot assumes a 10 $M_\odot$ BH, with $\alpha \simeq 0.22$, and an initial BH spin of 0.9. As discussed in Sec. IV C 2, the three levels reach a new quasiequilibrium state, in which we expect the 422 level to grow, becoming large at later times than those shown here.

The first diagram is almost on-shell for a 400 forced oscillation, so the $411\times411 \to 422\times{\rm BH}$ process is "resonant," like the $211\times311 \to 322\times{\rm BH}$ process discussed in Sec. III C. Consequently, its rate is suppressed by a lower power of α. Along with the large value of ε_{411} relative to ε_{211}, this means that the growth rate of 422 is positive for the three-level equilibrium occupation numbers. As a result, after $\mathcal{O}(100)$ e-folds of this new growth time, the three-level equilibrium would be disrupted by the growth of the 422 level.

We leave a more detailed analysis of evolution in this large-α regime to future work (as well as the evolution being complicated, our hydrogenic approximations are less reliable here). It is possible that further levels will grow after 422 does, leading to a complicated, multi-state superradiant cloud. In particular, is possible that the cloud could reach large enough field amplitudes that higher-order or nonperturbative processes become important, as we discuss in Sec. V.

3. n22 levels

$n22$ states grow and are depleted similarly to the 322 level, via the processes



at linear order in ε_{n22} (the superradiance rate of $n22$ states is small enough not to be important, for parameters of interest). The linear-order growth rate is

$$\dot{\varepsilon}_{n22} = \gamma^{n22\times{\rm BH}}_{211\times211}\left(1 - \frac{\gamma^{211\times\infty}_{n22\times322}}{\gamma^{n22\times{\rm BH}}_{211\times211}}\eta\right)\varepsilon^2_{211}\varepsilon_{n22}, \qquad (68)$$

where $\eta \equiv \varepsilon_{322}/\varepsilon_{211}$.

At early times, $\varepsilon_{322}/\varepsilon_{211} \ll 1$, and $n22$ is sourced in the same way as 322. However, since the 322 growth rate is at least $\mathcal{O}(1)$ larger, it has an exponentially larger occupation number than the other $n22$ levels by the time quasiequilibrium is established. For example,

$$\frac{\gamma^{422\times{\rm BH}}_{211\times211}}{\gamma^{322\times{\rm BH}}_{211\times211}} \simeq 0.36; \qquad \frac{\gamma^{n22\times{\rm BH}}_{211\times211}}{\gamma^{322\times{\rm BH}}_{211\times211}} \propto n^{-3}. \qquad (69)$$

(see Appendix C 2 a and Fig. 24 for further details). For the quasiequilibrium abundances of 211 and 322, the negative term in Eq. (68) dominates, reaching a value of 1.96 for $n = 4$ (1.69 for $n \to \infty$),

$$\frac{\gamma^{211\times\infty}_{n22\times322}}{\gamma^{n22\times{\rm BH}}_{211\times211}}\eta \gtrsim \frac{1}{2}\frac{\kappa^{211\times\infty}_{n22\times322}}{\kappa^{211\times\infty}_{322\times322}}\frac{\kappa^{322\times{\rm BH}}_{211\times211}}{\kappa^{n22\times{\rm BH}}_{211\times211}} \gtrsim 1.69. \qquad (70)$$

Including higher order corrections to the equilibrium ratio of 322 to 211, as well as the superradiance of 322, increases the ratio further. Thus the time derivative of $n22$ becomes

negative at leading order in α, independently of α, n, and a_*.

4. n33 levels

$n33$ states grow and are depleted by

$$211 \times 322 \to n33 \times \mathrm{BH} \qquad 322 \times n33 \to 211 \times \infty$$

giving

$$\begin{aligned}\dot{\varepsilon}_{n33} &= (\gamma^{n33\times\mathrm{BH}}_{211\times322} - \gamma^{211\times\infty}_{322\times n33})\varepsilon_{211}\varepsilon_{322}\varepsilon_{n33} \\ &\simeq (\kappa^{n33\times\mathrm{BH}}_{211\times322}\tilde{r}_+\alpha^{11} - \kappa^{211\times\infty}_{322\times n33}\alpha^8)\left(\frac{M_{\mathrm{pl}}}{f}\right)^4\varepsilon_{211}\varepsilon_{322}\varepsilon_{n33}\end{aligned} \tag{71}$$

at linear order in ε_{n33}. Due to the different α scaling, the grow rate is negative at small enough α. Quantitatively,

$$\left(\frac{\kappa^{211\times\infty}_{322\times n33}}{\kappa^{n33\times\mathrm{BH}}_{211\times322}}\right)^{1/3} = \begin{cases} 0.31 & n = 4 \\ 0.5 & n \to \infty \end{cases} \tag{72}$$

so at high spin, where $\tilde{r}_+ \simeq 1$, the growth rate is always negative for $\alpha \lesssim 0.3$ (see Appendix C 2 a and Fig. 25).

5. n44 levels

For $n44$, we have

$$322 \times 322 \to n44 \times \mathrm{BH} \qquad 322 \times n44 \to 211 \times \infty$$

giving

$$\begin{aligned}\dot{\varepsilon}_{n44} &= \left(\gamma^{n44\times\mathrm{BH}}_{322\times322}\frac{\varepsilon_{322}}{\varepsilon_{211}} - \gamma^{211\times\infty}_{n44\times322}\right)\varepsilon_{211}\varepsilon_{322}\varepsilon_{n44} \\ &\simeq (\kappa^{n44\times\mathrm{BH}}_{322\times322}\tilde{r}_+\alpha^3\eta - \kappa^{211\times\infty}_{n44\times322})\alpha^8\left(\frac{M_{\mathrm{pl}}}{f}\right)^4\varepsilon_{211}\varepsilon_{322}\varepsilon_{n44}\end{aligned} \tag{73}$$

at linear order in ε_{n44}.

With quasiequilibrium occupations for 211 and 322, the growth of $n44$ states occurs when α is large enough that

$$\frac{\kappa^{n44\times\mathrm{BH}}_{322\times322}}{\kappa^{211\times\infty}_{n44\times322}}\tilde{r}_+\alpha^3\eta \approx \frac{1}{2}\frac{\kappa^{n44\times\mathrm{BH}}_{322\times322}}{\kappa^{211\times\infty}_{n44\times322}}\frac{\kappa^{322\times\mathrm{BH}}_{211\times211}}{\kappa^{211\times\infty}_{322\times322}}\alpha^6\tilde{r}_+^2 \gtrsim 1, \tag{74}$$

or equivalently

$$\alpha\tilde{r}_+^{1/3} \gtrsim 0.3 \tag{75}$$

where the right-hand side is as large as 0.34 for $n = 5$ (0.3 for $n \to \infty$) (see Appendix C 2 a and Fig. 26).

6. Other levels

The $n22, n33$ and $n44$ levels considered above are the only ones which can be built up via quartic processes where the forced oscillation has $l = m = 0$.[7] To build up other processes via self-interactions, starting from 211 and 322, we need to use forced oscillations with $l > 0$, which have a parametrically smaller flux through the BH horizon. They therefore stand even less chance of having positive growth rates. For $l \geq 2$, we can often rule out these processes being relevant on astrophysical timescales, simply by estimating the magnitude of the growth rate. For example, for $l = 2$, we have

$$\gamma^{766\times\mathrm{BH}(2,-2)}_{322\times322}(\varepsilon^{\mathrm{eq}}_{322})^2 \sim 10^{-2}\left(\frac{\alpha}{0.3}\right)^{19}\left(\frac{M_\odot}{M}\right)\ \mathrm{Myr}^{-1}, \tag{76}$$

where the superscript $\mathrm{BH}(l, m)$ indicates the angular momentum numbers of the damped leg.

Taking an $l = 1$ example,

$$\dot{\varepsilon}_{655} = \gamma^{655\times\mathrm{BH}(1,-1)}_{322\times322}\left(1 - \frac{\gamma^{211\times\infty}_{655\times322}}{\gamma^{655\times\mathrm{BH}(1,-1)}_{322\times322}}\frac{\varepsilon_{211}}{\varepsilon_{322}}\right)\varepsilon^2_{322}\varepsilon_{655}. \tag{77}$$

The depletion term dominates at equilibrium as long as

$$\alpha\tilde{r}_+^{1/9} \lesssim \left(\frac{\kappa^{211\times\infty}_{655\times322}}{\kappa^{655\times\mathrm{BH}(1,-1)}_{322\times322}}\frac{1}{\eta^B}\right)^{1/9} \approx 0.7. \tag{78}$$

Similar checks can be performed for other processes involving mixing with an $l = 1$ damped state (see Appendix C 2 a). One finds that, for all of them, the depletion process to infinity dominates over the pumping process for the entire range of α for which $m = 1$ states can be superradiant ($\alpha \lesssim 0.5$).

V. NONPERTURBATIVE BEHAVIOR

So far, our analysis has assumed that the scalar field is always well-approximated by a combination of approximately hydrogenic bound states, and that quartic interactions result in the slow transfer of energy to and from these bound states. However, if the field amplitude becomes large enough, we expect this picture to break down. Most directly, for a generic potential, higher-order

[7]This is not strictly true—the Kerr potential breaks spherical symmetry, so l is no longer a good quantum number, and e.g., $n42$ can also be build up via a $m = 0$ forced oscillation. However, in the small-α limit, the overlaps for such processes are suppressed by more powers of α.

field interactions can become important. In addition, for large enough amplitudes, attractive interactions would make hydrogenic bound states unstable to collapse, in a "bosenova" [6,27,28].

As we explored in Sec. IV, for large self-couplings, the quartic interactions lead to the saturation of the cloud to a quasiequilibrium configuration (for much of the parameter space of interest), with field amplitude $\propto f$. For a potential of the form $V(\varphi) \propto g(\varphi/f)$, this means that the relative importance of higher-dimensional interactions becomes independent of f (for small enough f). As we will show below, for small α, the maximum value of $\theta \equiv \varphi/f$ is small, and the quartic-driven behavior we have investigated should be a good approximation. Similarly, for small α, the cloud is always far from the nonperturbative "bosenova" regime. For $\alpha \gtrsim 0.2$, we expect levels beyond 211 and 322 to grow in the small-f regime, as discussed in the previous section, so their behavior would need to be analyzed to draw conclusions about nonperturbative behavior.

A. Maximum field amplitude

When a single hydrogenic level dominates the energy stored in the cloud, the dimensionless field amplitude $\theta = \varphi/f$ is related to the occupation number of that level by $|\theta| \propto \alpha^{5/2}\sqrt{\varepsilon} M_{\rm pl}/f$. In the small and moderate self-coupling regimes, where 211 reaches its saturation occupation number, $|\theta|$ increases $\propto 1/f$ as f decreases. However, once we are in the large-self-coupling regime, the occupation numbers reached are $\propto f^2$, so θ becomes independent of f.

If 211 is the dominant level, then the maximum value of θ is attained at $r = 2a_0$ and $\theta = \pi/2$, with

$$|\theta_{\rm max}| \approx \alpha^{5/2}\sqrt{\varepsilon_{211}}\left(\frac{M_{\rm pl}}{f}\right)\sqrt{\frac{1}{8\pi}}e^{-1}. \tag{79}$$

As we decrease f, this increases until $f \simeq f_{\rm BC}$ [Eq. (59)]. For $\alpha \gtrsim 0.04$, $f_{\rm BC} = f_{\rm thresh}$ and

$$|\theta_{\rm max}(f_{\rm BC})| \approx \alpha^{7/4}\left(\frac{\log(GM^2)\kappa^{\rm SR}a_*(t_0)}{\kappa^{\rm BH}}\right)^{1/4}\frac{e^{-1}}{2\sqrt{\sqrt{2}\pi}} \approx 0.03\left(\frac{\alpha}{0.05}\right)^{7/4}. \tag{80}$$

The scalings in (80) are only representative when $\alpha \ll a_*(t_0)$ (see Appendix E). For $\alpha \lesssim 0.04$, $f_{\rm BC} = f_{\rm eq}$ and the maximum value of θ is equal to its value at equilibrium:

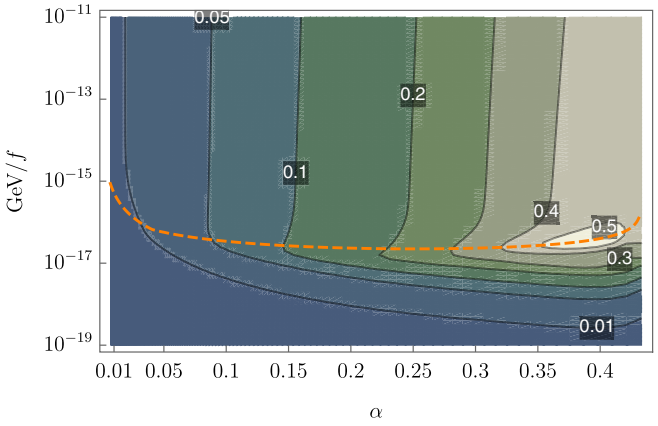


FIG. 7. Maximum value of $|\theta| \equiv |\varphi/f|$ attained during the evolution of the two-level 211/322 system, for a BH with initial spin $a_* = 0.99$ and initial mass 10 $M_\odot$ (the BH mass only affects this plot via the number of e-folds $\sim \log(GM^2)$ a level can grow). The dashed orange line indicates the boundary between the moderate and large self-coupling regimes (corresponding to $f_{\rm BC}$ as defined in Sec. IV). $|\theta_{\rm max}|$ is computed by numerically solving the evolution equations for the 211 and 322 occupation numbers.

$$|\theta^{\rm eq}_{\rm max}| \approx \alpha\left(\frac{\sqrt{\kappa^{\rm SR}a_*(t_0)\kappa^{\infty}}}{\kappa^{\rm BH}}\right)^{1/2}\sqrt{\frac{1}{\sqrt{24}\pi}}e^{-1} \approx 0.005\left(\frac{\alpha}{0.01}\right)\left(\frac{a_*(t_0)}{0.99}\right)^{1/4}. \tag{81}$$

(again, these scalings are valid when $\alpha \ll a_*(t_0)$).[8]

These equations suggest that, for small α, the value of $|\theta|$ never becomes large, so we would generically expect higher-dimensional interactions to remain unimportant. To see this more quantitatively, Fig. 7 shows the maximum value of $|\theta|$ attained during the evolution of the two-level 211/322 system, for different values of α and f. This has the expected behavior, increasing with decreasing f for $f \gtrsim f_{\rm BC}$, and reaching a constant value for smaller f (at a given α).

As discussed in Sec. IV, we expect that, for small f and $\alpha \gtrsim 0.2$, levels other than 211 and 322 will grow. At these parameters, the $|\theta_{\rm max}|$ values in Fig. 7 represent a lower

[8]Although Eq. (81) is valid at equilibrium, we noted in Sec. IV B 4 that ε_{211} can "overshoot" its equilibrium value as it evolves toward equilibrium. We have determined numerically that the overshoot estimate of Eq. (E5), or the approximation of Eq. (58), accurately predicts $\varepsilon^{\rm max}_{211}$ for $\alpha \gtrsim 0.05$ with an error less than 1%, deep in the self-interaction regime. Quantitatively, we found numerically that there is a thin band around the dashed boundary line of Fig. 7 [see Eq. (E7)], with a width of less than an order of magnitude in f, where both the quasiequilibrium and the overshoot estimates underpredict $\varepsilon^{\rm max}_{211}$ by $\gtrsim$5%. A significant discrepancy arises only in the region where $|\theta_{\rm max}|$ reaches its largest value and is ~20%. These translate to a ~2.5% and ~10% discrepancy in the analytically predicted $|\theta_{\rm max}|$, according to the scaling of Eq. (79).

bound (since the initial 211 overshoot value is still set by 211/322 dynamics). For the 411 level, which we expect to be the first to grow after the 211/322 quasiequilibrium (Sec. IV C 2), the maximum occupation reached is only around twice the maximum occupation number of 211. Consequently, the more spread-out wave function of 411 means that it does not attain a larger $|\theta|$ value. However, a more careful analysis would be required to determine $|\theta_{\max}|$ once other levels grow.

B. Bosenova

As well as higher-dimensional interactions becoming important, another possible issue arising at large occupation numbers is that the cloud may undergo a sudden collapse due to attractive self-interactions, known as a "bosenova" [6]. Here, we estimate the occupation number threshold for a bosenova to occur, using a variational approach.

The wave function for the hydrogenic 211 level is

$$\psi_{211} = \frac{\sqrt{N_{211}}}{2\sqrt{6}} a_0^{-5/2} r e^{-r/(2a_0)} Y_{11}(\theta,\phi) \tag{82}$$

where $a_0 \equiv 1/(\alpha\mu)$ is the Bohr radius. As our variational ansatz, we will take a wave function of this form, but with a modified radius,

$$\psi = \frac{\sqrt{N}}{2\sqrt{6}} R^{-5/2} r e^{-r/(2R)} Y_{11}(\theta,\phi) \tag{83}$$

For convenience, we will define a dimension-2 wave function $\tilde{\psi} = \sqrt{\mu}\psi$. Then, the nonrelativistic action for $\tilde{\psi}$ interacting with a gravitational field, sourced both by the central BH and by itself, is given by

$$\mathcal{S} \simeq \int \mathrm{d}^3 r \mathrm{d}t \frac{i}{2\mu}(\tilde{\psi}^*\partial_t\tilde{\psi} - \tilde{\psi}\partial_t\tilde{\psi}^*) - \frac{1}{2\mu^2}|\nabla\tilde{\psi}|^2 - \Phi|\tilde{\psi}|^2 + \frac{\lambda}{16\mu^4}|\tilde{\psi}|^4 - \frac{1}{8\pi G}|\nabla\Phi|^2 - \rho_{\mathrm{BH}}\Phi \tag{84}$$

The gravitational potential Φ obeys the Poisson equation,

$$\nabla^2\Phi = 4\pi G(\rho_{\mathrm{BH}} + |\tilde{\psi}|^2) \tag{85}$$

where we take $\rho_{\mathrm{BH}} = M\delta^3(\mathbf{r})$ and M is the mass of the BH. Using this potential, and integrating the action of Eq. (84) over space, we obtain an effective potential for R. Ignoring self-gravity of ψ, this is

$$V(\tilde{R}) = \frac{\alpha^4 {M_{\mathrm{pl}}}^2 \varepsilon}{\mu}\left(\frac{1}{8\tilde{R}^2} - \frac{1}{4\tilde{R}} - \frac{3\alpha^3\varepsilon {M_{\mathrm{pl}}}^2}{16384\pi\tilde{R}^3 f^2}\right), \tag{86}$$

where $\tilde{R} \equiv R/a_0$. The first two terms correspond to kinetic and gravitational energy, and set the radius of small-amplitude hydrogenic levels—the last terms arises from attractive self-interactions. The extrema of the potential $V(\tilde{R})$ are at

$$\tilde{R}^{\pm}_{\mathrm{extrema}} = \frac{1}{2} \pm \sqrt{\frac{1}{4} - \frac{9\alpha^3\varepsilon {M_{\mathrm{pl}}}^2}{4096\pi f^2}}. \tag{87}$$

If we decrease f, at some point these extrema will coincide, and the potential will no longer have a stable minimum. This leads to a bosenova, with the cloud collapsing. The critical occupation number for this to occur is

$$\varepsilon_{\mathrm{crit}} = \frac{1024\pi f^2}{9\alpha^3 {M_{\mathrm{pl}}}^2} \tag{88}$$

Incorporating the effects of self-gravity, this becomes

$$\varepsilon_{\mathrm{crit}} = \frac{32}{711\alpha^2}\sqrt{75840\pi\left(\frac{f}{M_{\mathrm{Pl}}}\right)^2 + 225\alpha^2} - \frac{160}{237\alpha} \tag{89}$$

which reduces to Eq. (88) for small f, i.e., for small clouds.

Given this, we can ask whether the 211 occupation number reaches $\varepsilon_{\mathrm{crit}}$ during its perturbative evolution. If it does not, then our assumption of perturbative evolution can be self-consistent. Figure 8 shows the maximum value of $\varepsilon_{211}/\varepsilon_{211}^{\mathrm{crit}}$ attained during the evolution of the two-level 211/322 system. For α small enough that other levels do not grow ($\alpha \lesssim 0.2$), we can see that this ratio is always $\lesssim 0.3$, so we do not expect a bosenova to occur. This is in

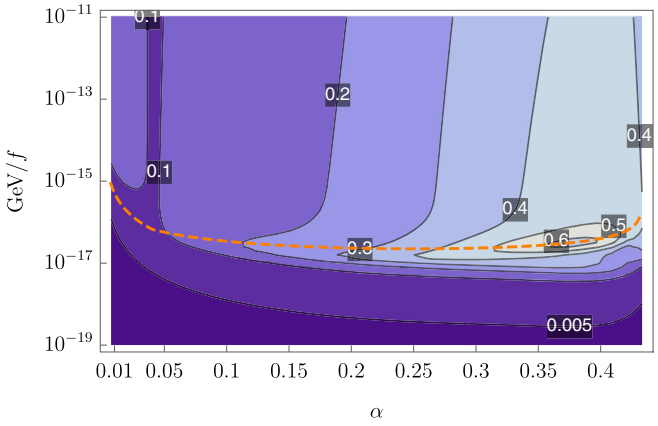


FIG. 8. Maximum value of $\varepsilon_{211}/\varepsilon_{211}^{\mathrm{crit}}$ attained during the evolution of the two-level 211/322 system, for a BH with initial spin $a_* = 0.99$. $\varepsilon_{211}^{\mathrm{crit}}$ is the critical occupation number above which a rapid collapse of the cloud (a bosenova) is expected to occur (Sec. V B). The dashed orange line indicates the boundary between the moderate and large self-coupling regimes (corresponding to f_{BC} as defined in Sec. IV). ε_{211} is computed by numerically solving the evolution equations for the 211 and 322 occupation numbers. The plot is roughly independent of the BH mass, within the range of astrophysical BHs.

contrast to the conclusions of much of the existing literature. As emphasized previously, other papers neglect the perturbative processes that lead to energy exchange between hydrogenic levels, causing the cloud to saturate to a quasiequilibrium configuration before its amplitude becomes large enough for a bosenova.

For $\alpha \gtrsim 0.2$, we expect that levels other than 211 and 322 will grow. This means that the $\varepsilon/\varepsilon_{\rm crit}$ values in Fig. 8 represent a lower bound. As we discussed in the previous subsection, the more spread-out wave function of the 411 level means that it is unlikely to get closer to the critical occupation number than the 211 level; we leave an analysis of the situation once other levels have grown to future work.

1. Subleading effects

As discussed in Appendix F, superradiance extracts mass from the BH in addition to angular momentum. As such, the cloud can actually grow to be somewhat larger than we have assumed so far. The modified equations for purely gravitational superradiance can be found in Appendix F. In deriving Fig. 7 and 8 we have included the correction coming from the change of the BH mass or, equivalently, from the time-dependence of α. As expected, we find that this correction can become quite large near the superradiance boundary, as the final spin is slightly modified (see Eq. (F10). However, for strong self-interactions, where the bosenova might be relevant, there is practically no significant correction, as the cloud does not grow appreciably and thus does not extract a significant amount of spin or mass from the BH.

One might also ask how the inclusion of another level, say 322, changes the above picture. Assuming that its fractional occupation number is small compared to our primary level (e.g., 211), we can treat such a level as a small perturbation and check whether our results are consistent. In what follows we will neglect self-gravity for clarity or, equivalently, we will work in the small f (large self-interactions) limit, where Eq. (89) coincides with Eq. (88). We add a contribution from 322 to our variational ansatz

$$\tilde{\psi} \supset \frac{M_{c2}^{1/2}}{4a_0^{3/2}} \frac{4}{81\sqrt{30}} \left(\frac{r}{a_0}\right)^2 \exp\left(-\frac{r}{3a_0}\right) Y_2^2(\theta,\phi) \tag{90}$$

where M_{c2} is the mass of the 322 cloud. Note that we treat 322 as rigid, i.e., we do not allow its radius to change. Following the same procedure as before, we get an effective potential for 211 with an additional attractive term, stemming from its interaction with 322

$$V(\tilde{R}) = \frac{\alpha^4 {M_{\rm pl}}^2 \varepsilon}{\mu}\left(\frac{1}{8\tilde{R}^2} - \frac{1}{4\tilde{R}} - \frac{3\alpha^3 \varepsilon {M_{\rm pl}}^2}{16384\pi\tilde{R}^3 f^2} - \frac{27\tilde{R}^4\alpha^3\varepsilon_2 {M_{\rm pl}}^2}{2\pi(3+2\tilde{R})^9 f^2}\right) \tag{91}$$

where ε_2 is the fractional occupation number of 322. Expanding around the critical values as $\tilde{R} = \frac{1}{2} + \sqrt{\varepsilon_2}\delta\tilde{R}$ and $\varepsilon = \varepsilon_{\rm crit} + \varepsilon_2\delta\varepsilon$, we find the correction $\delta\varepsilon = 21/16384$, giving

$$\frac{\varepsilon_2\delta\varepsilon}{\varepsilon} = \frac{21}{16384}\frac{\varepsilon_2}{\varepsilon} \ll 1. \tag{92}$$

The result is indeed small and, thus, it does not change our conclusions about the bosenova. In particular, the correction to $\varepsilon_{\rm crit}$ is *positive*. Since the interaction is attractive, as seen from the potential in Eq. (91), the 322 cloud attracts the 211 one and, since it resides at a larger radius, it effectively dilutes it.

In Fig. 8, we compared the ε_{211} value attained during the perturbative level evolution to $\varepsilon_{\rm crit}$. However, the rates of the different processes involved in the evolution were calculated for the unperturbed hydrogenic wave functions. Consequently, we should ask whether self-interaction-induced perturbations to the wave functions make a significant difference to the rates, and so the occupation numbers attained. From Eq. (87), we can see that if $\varepsilon_{211}/\varepsilon_{\rm crit}$ is always small, then the corrections to the wave functions will always be small, and our calculations should be self-consistent. Since $\varepsilon_{211}/\varepsilon_{\rm crit}$ only becomes large for larger α, where (as discussed previously) our perturbative evolution calculations are already incomplete, we leave a full analysis to future work.

In plotting $|\theta_{\rm max}|$, we have used the field defined using Eq. (83), that is, by taking into account the corrected radius of Eq. (87). This amounts to multiplying Eq. (79) by a factor of $(\tilde{R}^+_{\rm ext})^{-3/2}$ [Eq. (87)], giving

$$|\theta| \approx \alpha^{5/2}\sqrt{\varepsilon_{211}}\left(\frac{1}{\tilde{R}^+_{\rm ext}}\right)^{3/2}\left(\frac{M_{\rm pl}}{f}\right)\sqrt{\frac{1}{8\pi}}e^{-1}. \tag{93}$$

We have determined numerically that the radius change is at most 15% and introduces at most a 25% change in the region where $|\theta|$ grows to be the largest possible, driving to a value of $\sim$0.5, whereas the change is much smaller everywhere else.

Another possible issue with our variational analysis is that the evolution is not adiabatic during the last few e-folds before 211 reaches its maximum occupation number. As a result, the cloud might not trace the minimum of the potential of Eq. (86) but rather oscillate around it, in the manner of an "excited state." In this case, the cloud could overcome the barrier at $\tilde{R}^-$ [Eq. (87)] and collapse. We note that oscillations of the radius of the peak seem consistent with the results of Ref. [47]. The minimum of the potential would need to be fairly close to critical for this to be an issue, but we leave detailed investigation of this point to future work.

2. Comparison to simulations

While we expect our hydrogenic ansatz to be a good approximation, properly understanding the dynamics of a bosenova requires numerical simulations. In [27,28], the authors numerically simulate the evolution of a self-interacting scalar field around a high-spin Kerr BH, starting from a hydrogenic bound state profile with $\theta \sim \mathcal{O}(1)$. These simulations effectively operate in the large self-coupling regime, taking the cloud's mass to be very small compared to the BH. In simulations with $a_* = 0.99$ and $\alpha = 0.3$ [28], they find that a 211 bound state with initial amplitude such that $|\theta_{\rm max}| = 0.4$ does not undergo a bosenova, but one with $|\theta_{\rm max}| = 0.45$ does.

Comparing these to our variational calculations, we can convert the critical occupation number (88) to a field amplitude, giving the leading-α expression $|\theta^{\rm crit}_{\rm max}| = \frac{8\sqrt{2}}{3e}\alpha \simeq 0.42\frac{\alpha}{0.3}$. This is highly compatible with the threshold behavior observed in the simulations.

The simulations in [27,28] were evolved forward for $t \simeq 2000 r_g$. This is much shorter than the timescales for any of the perturbative processes studied in Sec. IV, including 211 superradiance, and the growth of 322 through self-interactions. A simulation would have to be run for much longer times to observe these effects. In particular, the fact that a bosenova was observed for the initial state $|\theta_{\rm max}| = 0.45$ is not evidence that a bosenova would occur around an astrophysical black hole. In the latter case, the true initial conditions are at an exponentially smaller amplitude, and according to our estimates, the maximum 211 amplitude reached during the evolution is $|\theta_{\rm max}| \simeq 0.3$ (Fig. 7), at which point interactions with 322 cut off its growth.

3. Repulsive self-interactions

In [31], it is claimed that if self-interactions are repulsive, they can completely suppress the growth of 322, by spreading out the 211 cloud and reducing the rate of the $211 \times 211 \to 322 \times {\rm BH}$ process. We can estimate the effect of repulsive self-interactions by looking at how they shift the 211 wave function radius in our variational ansatz. This gives

$$\tilde{R}_{\rm rep} = \frac{1}{2}\left(1 + \sqrt{1 + \frac{\varepsilon_{211}}{\varepsilon^{\rm crit}_{211}}}\right) \tag{94}$$

with $\varepsilon^{\rm crit}_{211}$ from Eq. (88). Since the perturbative evolution processes from Sec. IV all depend on λ^2, they are the same for attractive and repulsive self-interactions. Consequently, the maximum value of ε_{211} attained through perturbative evolution should be the same. As a result, we expect that, unless $\varepsilon_{211}/\varepsilon^{\rm crit}_{211}$ becomes large (which we cannot rule out for $\alpha \gtrsim 0.2$ and small f), the effects of repulsion should be small.

VI. BLACK HOLE SPIN-DOWN

One of the observational signatures of superradiance is the spin-down of initially fast-spinning BHs [5,6]. In the absence of nongravitational interactions, if a BH is born with spin high enough that a mode is superradiant, and the mode's growth time is much shorter than the lifetime of the BH, then a superradiant cloud will form around the BH. This spins down the BH to the point where the mode is stable, rather than growing. Consequently, observing a sufficiently old, sufficiently fast-spinning BH is good evidence against the existence of a light boson with such properties. Constraints of this kind have been placed on spin-0 [12,48] and spin-1 [20] particles from measurements of BH spins in x-ray binaries [35,49] (higher-spin particles have also been considered [50,51], though such models encounter theoretical issues, as we discuss in the conclusions).

In contrast, if self-interactions are large, then as discussed in Sec. IV, the occupation numbers in the quasiequilibrium state are suppressed. Consequently, the rate of energy and angular momentum extraction from the BH is suppressed, and the spin-down constraints described in the previous paragraph will not apply directly.

Instead, for small enough f, the time-averaged spin extraction rate will be approximately set by the equilibrium occupation number of the 211 level (at least in the case of 211 superradiance), as discussed in Sec. IV B 4. Since $\varepsilon^{\rm eq}_{211} \propto \alpha^{-3}\frac{f^2}{M_{\rm pl}^2}$ [Eq. (55a)], the time taken to fully spin down the BH (to the point where 211 superradiance is saturated) scales $\propto f^{-2}$. Consequently, as reviewed in Sec. IV B 5, there is some minimum f below which the BH is not significantly spun down in the time available.

This behavior is illustrated, for particular initial BH parameters, in Fig. 9. The figure shows how, for $f \lesssim f_{\rm BC}$ (Table II; $f_{\rm BC} \simeq 3 \times 10^{16}$ GeV for the left-hand panel, and $\simeq 2 \times 10^{17}$ GeV for the right-hand panel), spin-down to the $m = 1$ superradiance threshold takes longer as f is decreased, until it no longer occurs within the lifetime of the BH for $f \lesssim f_{\rm CD}$. The region of (μ, f) parameter space in which the BH is spun down to the $m = 1$ superradiance threshold is shown in the bottom-left panel of Fig. 3.

We have only performed a detailed analysis (at all f) of situations in which 211 is the first superradiant level to grow, and levels beyond 211 and 322 do not grow. From Sec. IV, this corresponds to $\alpha \lesssim 0.2$. Nevertheless, we can be confident that, when interactions are weak enough that superradiant growth of the 322 level is unaffected, the black hole is spun down as in the purely gravitational case. This is indicated in the bottom right of the lower panels in Fig. 3.

Applying this physics to observations of astrophysical BHs, Fig. 10 shows the regions in the (μ, f) plane for which sufficient spin-down occurs, so that spin measurements from BHs in x-ray binaries constrain an axion with that mass and coupling. For each black hole, the solid line of the

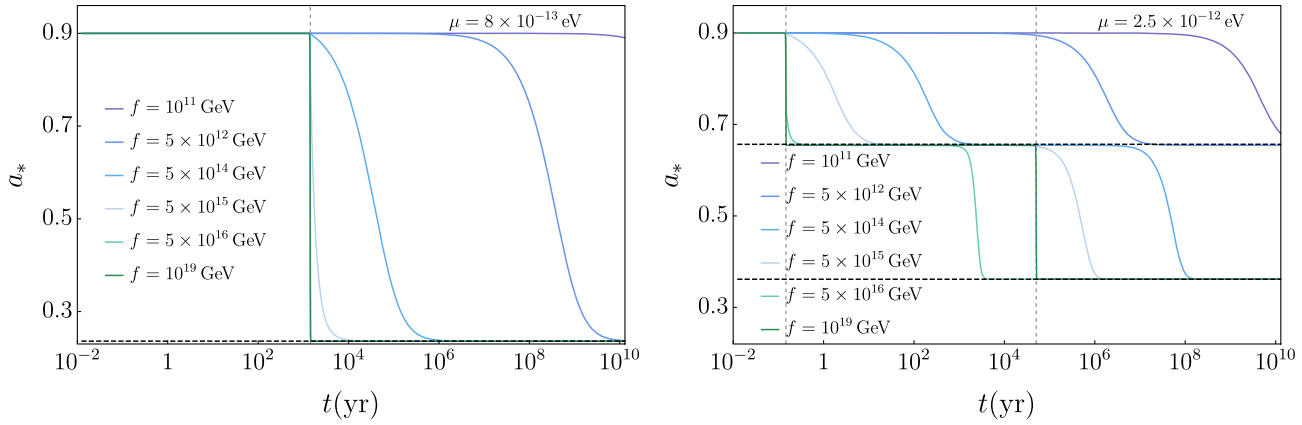


FIG. 9. Black hole spin-down as a function of time for $\mu = 8 \times 10^{-13}$ eV (left panel) and $\mu = 2.5 \times 10^{-12}$ eV (right panel) for a range of self-interactions strengths, and a 10 $M_\odot$ black hole. These axion masses correspond to $\alpha \simeq 0.06$ and $\alpha \simeq 0.19$ respectively. The dashed horizontal lines show the superradiance boundary for levels 211 (upper) and 322 (lower). The dashed vertical lines show the expected spindown time in the limit of no self-interactions for levels 211 (smaller t) and 322 (larger t).

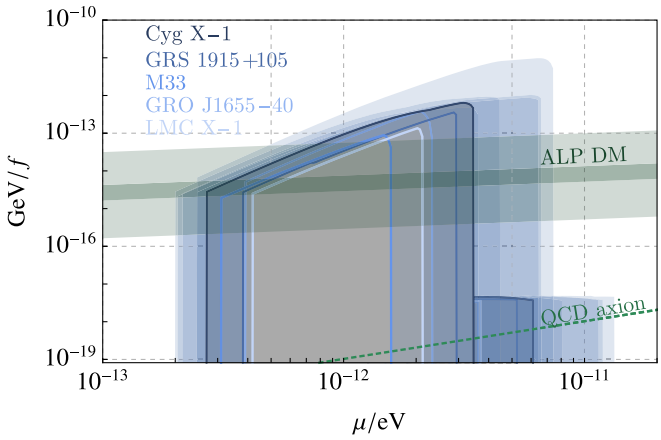


FIG. 10. Constraints on axion parameter space from black hole spin measurements in x-ray binaries. For each black hole, the region enclosed by the solid line of the corresponding color (see key at top left) is the intersection of the $m = 1$ spin-down regions for different BH parameters (mass, spin, lifetime, binary period, and mass of the binary companion) within the observational error intervals. This corresponds to the parameter space region in which we can be confident that spin-down occurs, so is constrained by observations of that BH. The light shaded regions of each color are the unions of the spin-down regions for different BH parameters and could be constrained by improved measurement and analysis of these BHs. Higher axion masses could potentially be constrained using higher-m levels; we include only the analog of the small and moderate self-coupling regimes A and B (for which self-interactions do not affect the extraction of angular momentum to the level with the largest SR rate) for $m = 2$, where the analysis in this work applies. The "ALP DM" band corresponds to the range of quartic couplings that allow the observed DM abundance to be produced by the misalignment mechanism. The darker middle band corresponds to $\mathcal{O}(1)$ values of the initial misalignment angle $[\theta \in (1, \pi - 1)]$, while the lighter bands above and below correspond to "tuned" initial values $[\theta \in (10^{-1}, \pi - 10^{-6})]$.

corresponding color indicates the region in which spin-down would occur with high confidence, given the uncertainties on the measured BH parameters. The larger shaded regions are those in which spin-down may occur, given BH parameter values within the confidence intervals; these represent the regions of parameter space which may be constrained by future, better observations of these BHs. Given the uncertainties in our analyses when $\alpha \gtrsim 0.2$ and f is small, the constraints in those parts of parameter space should be treated as estimates requiring further study.

Figure 10 can be compared to Fig. 11 of [12]. The latter assumed that the dominant effect of quartic self-interactions was to cause periodic bosenova events when the cloud became too large; parametrically, when

$$N \gtrsim 16\pi \frac{\ell^4}{\alpha} \frac{f^2}{\mu^2} \tag{95}$$

for an $l, m = \ell$ superradiant level, as discussed in [6]. From the previous section, we know that, at small α and small f, the critical occupation number for a bosenova to occur has the same parametric scaling as the equilibrium 211 occupation number, but is numerically larger, $\epsilon_{211}^{\rm eq}/\epsilon_{211}^{\rm crit} \sim 0.1$ (Eq. (89) and Fig. 8). Consequently, we expect the time-averaged 211 occupation number in our picture to be parametrically the same as that assumed in [12]. Numerically, since [12] assumes that a bosenova completely destroys the cloud, which then takes $\mathcal{O}(100)$ e-folds to be rebuilt, our time-averaged 211 occupation number is actually slightly larger, for the same parameters, resulting in slightly stronger spin-down constraints.

The age (or accretion timescale) of the BH limits how small a particle mass μ can be constrained by spin-down measurements—if μ is too small, then superradiance is not fast enough to spin down the BH. A separate effect is that, for small μ, the cloud is more dilute, and can be disrupted

by tidal forces from the companion star [45]. These gravitational perturbations mix superradiant levels with decaying ones (e.g., 211 with $21-1$), which can inhibit their growth. We do not attempt a careful analysis of the effects on the evolution of the cloud, but adopt the conservative approach of not placing constraints when the companion is closer than the maximum radius for the resonant depletion processes identified in [45] (see Appendix I). This sets the small-μ boundary of the constrained region in Fig. 10. We are able to constrain axion masses a factor ∼2 lighter than the limits from [12], which included an unphysical dipole gravitational potential effect from the companion.

In most of this paper, we have taken our nominal BH mass to be $\mathcal{O}(10\ M_\odot)$. However, our analyses can be easily rescaled to different BH masses; the most important dimensionless parameter that changes is the ratio of the BH lifetime to the light-crossing time. Figure 11 shows the spin-down parameter space for a supermassive BH (SMBH), with $M = 10^7\ M_\odot$. This parameter space sits at smaller μ (due to the larger BH size) and larger f (due to the smaller $T_{\rm BH}\mu$ parameter) than for a stellar-mass BH. There do exist spin measurements for some SMBHs [52–54], and these could be used to place constraints on very-low-mass bosons (see e.g., [30,55,56]). However, the galactic center environments in which SMBHs live are rather complicated, and understanding environmental effects on the evolution of a superradiant cloud (e.g., due to the occasional infall of compact objects) would be necessary to place robust constraints. We leave such an analysis to future work, but include Fig. 11 as a guide to the kind of region that might be constrained by these measurements.

As well as spin measurements for BHs in x-ray binaries, there are also spin measurements for $\mathcal{O}(10\ M_\odot)$ BHs from gravitational wave observations of binary BH mergers at LIGO and Virgo [57–62]. The statistical uncertainty of these measurements is generally much greater than the estimated errors of x-ray binary spin measurements—for most of the binary BH mergers observed so far, the spins of the primary BHs could lie in an $\mathcal{O}(1)$ range, and are consistent with zero. However, there were two events in recent observing runs for which one of the primary BHs was measured to have high spin (significantly different from zero); GW190412 and GW190517 [37]. The inferred masses of these BHs were $\sim 30\ M_\odot$, which is significantly heavier than the BHs observed in x-ray binary systems. Consequently, if one assumes that the history of the system would have allowed a superradiant cloud to grow around the BH, one can constrain smaller boson masses, in the range $\mu \sim 1.3\times10^{-13}$ eV—2.7×10^{-13} eV [37].

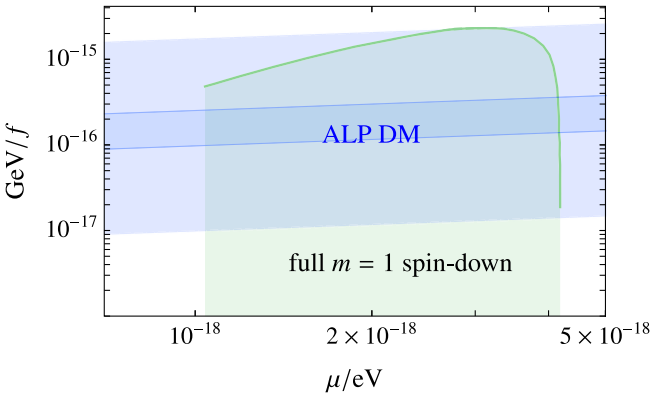


FIG. 11. Parameter space for which the 211 level of a supermassive BH ($M_{\rm BH} = 10^7\ M_\odot$), with initial spin $a_* = 0.9$, spins the BH down to saturation within an Eddington accretion timescale, $t_{\rm Edd} \simeq 4\times10^8$ yr. The "ALP DM" band is defined as in Fig. 10.

Given that we have no reliable information about the premerger history of these BHs, we do not include them in Fig. 10. However, with better understanding of such systems, gravitational wave observations of binary BH mergers could become a valuable tool for constraining (or providing evidence for) light bosons. In addition, while mergers other than the two mentioned above do not provide strong evidence regarding superradiance [36,37],[9] future data from many such mergers may provide statistical evidence for or against superradiant BH spin-down [15,64].

A. Axion models

Understanding the parameter space in which spin-down constraints apply is important in determining the consequences for motivated particle physics models. For the QCD axion, Fig. 10 confirms that, at least for 211 and 322 superradiance, self-interactions are small enough not to affect spin-down constraints.

Another motivated target model is an axion with a fixed (rather than temperature-dependent) potential. An initial "misalignment" axion field value in the early universe will lead to a dark matter density at late times, depending on the axion mass, the shape of the potential, and the initial field value. Consequently, while the mass and self-couplings of a generic axion can vary independently, imposing that the misalignment mechanism must generate the observed DM

[9]This is in contrast to some works which claim that earlier GW spin measurements can put constraints on BH superradiance (e.g., [56]). These claims appears to be based on a misinterpretation of the spin measurements presented by the LIGO collaboration. For example, the pre-merger spin of the primary BH in GW150914 is given as $0.32^{+0.47}_{-0.29}$, where the errors correspond to a 90% credible interval [63]. [56] appears to use the interpretation of spins below $0.32-0.29=0.03$ as being excluded at the 90% level, to place constraints on superradiant processes that would have reduced the spin to below this value. However, suppose (for example) that we had a uniform prior on $a_* \in [0,1]$, and that the measurement gave us no information about a_*. Then, [0.05, 0.95] would be a 90% interval, and spins <0.05 would be excluded at the 90% level, despite obtaining no new information; to set constraints a more complete analysis is needed.

density gives the "ALP DM" band in Fig. 10 (for a cosine potential $V \propto \cos(\varphi/f)$).

The darker central part of this band corresponds to masses and self-couplings for which a "generic," $\mathcal{O}(1)$ misalignment angle, $\theta_{\rm init} = a_{\rm initial}/f \in (1, \pi - 1)$, gives the correct dark matter density. For the same μ and $\theta_{\rm init}$, but larger f, we would obtain too large a dark matter density. However, this can be fixed by "tuning" the initial field value to be close to the bottom of the potential. Since $\rho_{\rm DM} \propto \mu^{1/2}\theta_{\rm init}^2 f^2$ for small $\theta_{\rm init}$, the tuning required is simply $\theta_{\rm init} \propto 1/f$. The lower edge of the band in Fig. 10 corresponds to $\theta_{\rm init} = 0.1$.

At smaller f, we have the opposite problem of not producing enough DM. For a cosine-type potential, this can be solved by tuning the initial field value to be close to the top of the potential, so that its transition to matterlike oscillations around the bottom of the potential is delayed. This "large-misalignment mechanism" [65] can lead to significant enhancements of dark matter density perturbations, resulting in a range of phenomenological signatures. In Fig. 10, the top edge of the band corresponds to $\theta_{\rm init} = \pi - 10^{-6}$ (see Appendix K for formulas), illustrating that, apart from the lower end of the μ range, BH spin-down constraints still apply to such models.

As well as affecting dark matter in the early universe, self-interactions could have effects at late times, leading to DM-DM scattering in halos. The associated relaxation rate is, parametrically [66,67],

$$\begin{aligned}\Gamma &\sim \frac{\rho_{\rm DM}^2}{f^4\mu^3 v^2} \\ &\sim 3\times 10^{-26}{\rm yr}^{-1}\left(\frac{\rho_{\rm DM}}{\rm GeV\,cm^{-3}}\right)^2\left(\frac{10^{11}\ {\rm GeV}}{f}\right)^4 \\ &\quad\times\left(\frac{10^{-12}\ {\rm eV}}{\mu}\right)^3\left(\frac{10^{-3}}{v}\right)^2 \end{aligned} \tag{96}$$

where v is the halo's virial velocity (this should be compared to the relaxation rate $\Gamma \sim \frac{\rho_{\rm DM}^2}{M_{\rm pl}^4\mu^3 v^6}$ for gravitational interactions [68–70]). Consequently, unless DM forms very dense structures, quartic self-interactions will not be significant in halos, for the parameter space we have been considering.

VII. GRAVITATIONAL WAVES

Gravitational waves emitted by the superradiant cloud are a unique signal of ultralight bosons, turning gravitational wave observatories into indirect particle detectors [5,6]. The superradiant cloud can grow to up to several percent of the black hole's mass, and sources gravitational waves through its oscillating stress-energy tensor. These are almost-monochromatic, coherent, and long-lasting. Such emission occurs in two parametrically different frequency ranges; higher-frequency "annihilation" signals, with $\omega \simeq 2\mu$, and lower-frequency "transitions," with $\omega = \omega_j - \omega_{j'}$ set by the frequency difference between different bound levels.

Conceptually, annihilation signals are sourced by the annihilation of two axions into a graviton. Consequently, they are emitted by any level populated by a single real scalar field. The timescale over which such emission lasts is parametrically longer than the superradiant growth time (Sec. IV B 2), making them promising for detection at gravitational wave observatories. Up to thousands of potential annihilation signals could be detectable, from black holes in the Milky Way, at Advanced LIGO and Virgo [12,15,38–40]. Such signals, and their detectability, have been studied in the context of continuous wave searches [12,15], stochastic searches [38,39,71], directed searches for clouds around products of binary mergers [15,72], and directed searches for clouds around BHs in x-ray binaries [73,74]. Searches with LIGO/Virgo data are ongoing; so far, no signals have been observed [40,75,76], though using this nonobservation to constrain superradiance relies on poorly measured black hole population properties, and may suffer from down-weighting of the signal [40]. Searches at space-based, lower-frequency gravitational wave detectors such as LISA will be sensitive to lighter axions [12,38,39], while heavier axions may be observable with future higher-frequency detectors [77,78].

Transition signals correspond to axions dropping into a more deeply bound level, emitting gravitational radiation at the frequency set by the level splitting. Attaining a significant emission rate requires both levels to have large occupation numbers simultaneously. For the case of purely gravitational superradiance, these circumstances only arise for higher-l levels and for short times, leading to limited observational prospects at current gravitational wave observatories [12].

More specifically, for a given $m < 3$, the fastest-growing superradiant level is also the most tightly bound one, so other modes with the same m have exponentially smaller occupation numbers. For $m \geq 3$, this is not always the case—for example, at large a_* and near-threshold α, the growth rate of 433 becomes smaller than that of 533 and higher levels. This can lead to multiple $m = 3$ levels having large occupation numbers simultaneously. Similar crossings happen for $m = 4$ and higher levels, as illustrated in Fig. 12.

These circumstances allow gravitational wave transition signals of nonnegligible amplitude to occur around astrophysical BHs. Even so, compared to annihilation signals, they offer less promising observational prospects. The total energy released, if the occupation number of the higher level transitions entirely to the lower one, is $E = \Delta\omega N \lesssim \alpha^2\mu N$, whereas annihilations can emit the entire energy stored in a cloud, $E \sim \mu N$. In addition, signal durations for transitions are typically of order a superradiance time, compared to the parametrically longer annihilation signals [12]. Nevertheless, transition signals could probe

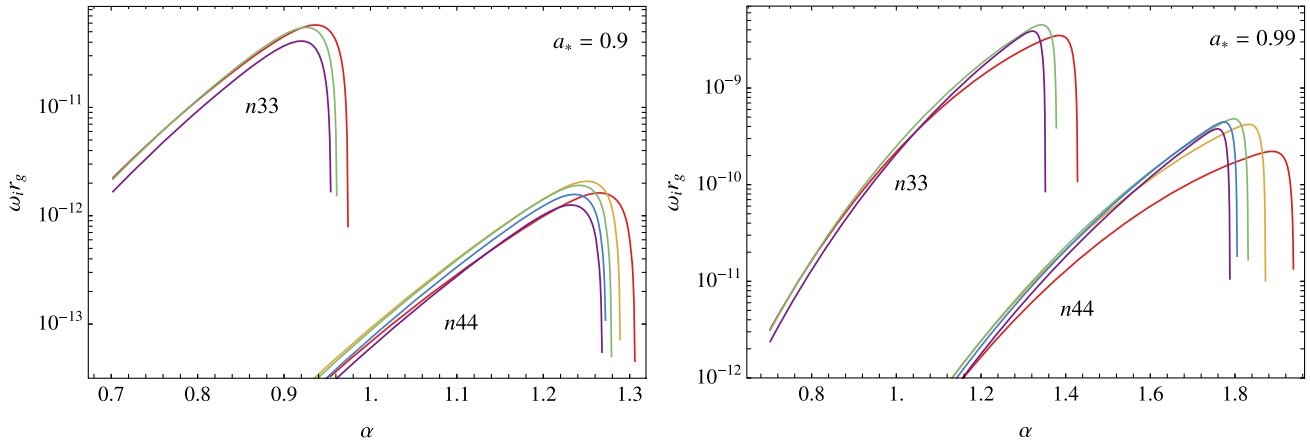


FIG. 12. Superradiance rates for the $n33$ and $n44$ hydrogenic bound states, computed numerically on the full Kerr background (using the continued fraction method of [79]). The left-hand plot shows rates for $a_* = 0.9$, and right-hand plot those for $a_* = 0.99$. The red curves correspond to the levels with smallest n; levels with larger n have cutoffs at progressively smaller α. These plots illustrate how, at some α parameters, different hydrogenic levels can have the same superradiance rates. As discussed in Sec. VII, this can give rise to gravitational wave transition signals.

interesting parts of parameter space, providing sensitivity to heavier axions than annihilation signals do (for a given BH mass).

Compared to the purely gravitational behavior summarized in the preceding paragraphs, the presence of self-interactions can have a significant effect on the gravitational wave signatures of superradiance. For annihilations, self-interactions suppress the potential signals due to two main effects: the gravitational wave power emitted is reduced due to the smaller cloud size, and the new energy loss mechanisms via scalar radiation reduce the total energy emitted in GWs. On the other hand, self-interactions provide a mechanism to populate multiple levels simultaneously, potentially increasing the parameter space for transition signals (though the cloud size and scalar radiation caveats still apply). In the rest of this section, we discuss annihilation and transition signals and their observational prospects in more detail. We focus on continuous wave searches for such signals, which are well-suited to louder signals from within our galaxy, and can provide a wealth of information about the detected signal properties. Stochastic searches to look for excess power in a narrow frequency range could potentially be performed more (computationally) cheaply and would also be interesting to study in future work.

A. Annihilations

In this subsection, we focus on the prospects for observing annihilation signals from the 211 level, for a range of self-couplings, at current gravitational wave observatories. We also comment briefly on other types of annihilation signals, including annihilation signals from complex scalar fields.

Figure 13 illustrates the effects of self-interactions on gravitational signatures of 211 superradiance, showing the peak signal amplitude, signal duration and sensitivity reach for different axion masses and self-couplings. To estimate the projected reach, we take the design strain sensitivity of Advanced LIGO [80], and assume all-sky semi-coherent continuous wave (CW) search strategies, with coherent integration times of 240 hours, and sensitivity depth $\mathcal{D}^c(f) \sim 50/\sqrt{\text{Hz}}$. The sensitivity depth is defined by $\mathcal{D}^c(f) \equiv \sqrt{S_h(f)}/h_0^c(f)$, where $\sqrt{S_h(f)}$ is the noise spectral density and $h_0^c(f)$ is the strain limit at the desired confidence level c. It allows comparisons of different searches, independently of the data used, and depends on the detailed search technique, coherent integration time, total integration time, etc. [81]. The latest searches with O2 data have used coherence times of up to $T_{\text{coh}} = 60$ hrs with $N_{\text{seg}} = 64$ segments in the first analysis stage [82], and have reached sensitivity depths of $\sim 30/\sqrt{\text{Hz}}$ [83] to $\sim 50/\sqrt{\text{Hz}}$ [82] for $c = 90\%$ exclusion limits. Since the CW searches assumes a constant signal amplitude over the entire integration time, while our signals may change on times shorter than the coherent search time, we conservatively penalize our reach by $\sqrt{\tau_{\text{sig}}/T_{\text{coh}}}$ (though the searches could be improved to take into account the time dependence of the signal, alleviating this penalty).

While the sensitivity reach is a useful quantity for a search targeting a specific BH, standard CW searches are 'blind', and look for signals from sources anywhere in the sky. Figure 14 shows the expected number of events in such a search at Advanced LIGO, given assumptions about the galactic BH population, for different self-couplings.[10]

[10]It should be noted that for small axion masses, where there may be multiple long-duration signals from galactic BHs, stochastic searches for excess power within a frequency range may be an advantageous approach. We leave a quantitative comparison of stochastic and CW searches to future work.

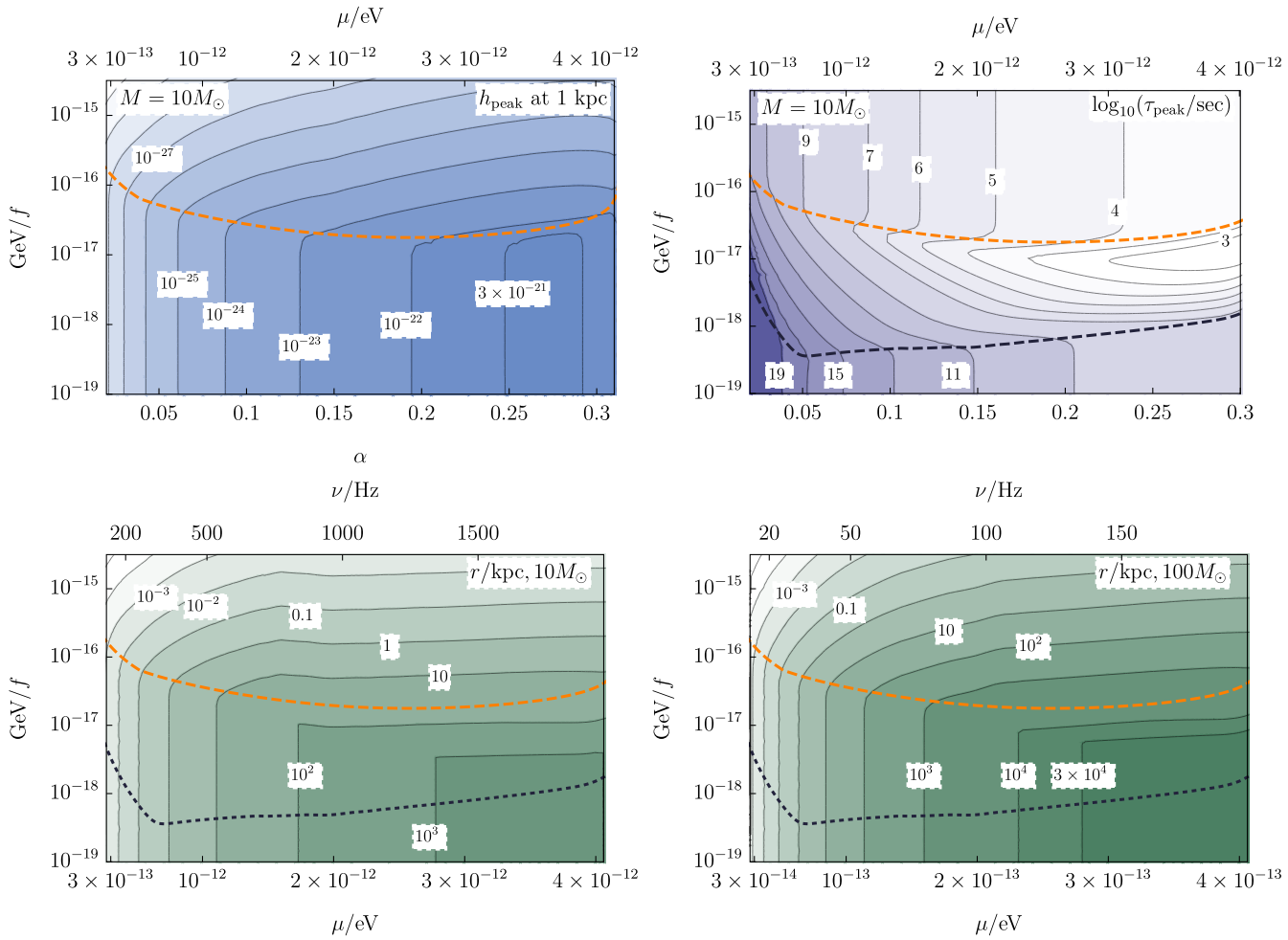


FIG. 13. Upper left: peak strain from $211 \times 211 \to$ GW annihilations for an observer at 1 kpc from a 10 $M_\odot$ BH, with initial spin 0.9. Upper right: typical duration $\tau_{\rm peak}$ of peak signal, $\log_{10}(\tau_{\rm peak}/\,{\rm sec})$. In the large self-interactions regime, we show the time-scale of the overshoot regime, corresponding to the peak signal strain. Lower left: sensitivity reach in kpc to a 10 solar mass BH, for continuous wave searches at Advanced LIGO design sensitivity [80]. Lower right: reach in kpc to a 100 solar mass BH. The dashed orange line indicates the boundary between the moderate and large self-coupling regimes (corresponding to $f_{\rm BC}$, Sec. IV), while the dotted black line indicates the boundary of the regime in which the 322 level grows appreciably ($f_{\rm AB}$).

We assume a power-law BH mass distribution, $dN/dM \propto M^{-2.35}$, with a minimum black hole mass of 5 $M_\odot$, and vary the maximum black hole mass from 20 to 45 $M_\odot$ [84]. For the BH spatial distribution, we take a combination of the disk and bulge distributions as in [40], with a total number of 10^8 BHs, born at a uniform rate throughout the age of the galaxy. We vary the BH spin distribution, with our extreme cases having 10% and 0.2% of BHs with initial spin $a_*(t_0) \geq 0.9$, respectively. The 10% figure is consistent with spin measurements from x-ray binaries [85,86], and 0.2% with models of rare high spin BHs associated with gamma ray bursts [87,88], making them reasonable upper and lower bounds.

The shaded bands in Fig 14 correspond to this range of BH population assumptions. While these unknowns do give rise to orders of magnitude uncertainty in the expected event rate, we can see that, for particle masses just below the spin-down threshold, even the pessimistic distributions give a promising number of events for purely gravitational superradiance. Conversely, the very large number of events (at design sensitivity) predicted by the optimistic distributions means that some of this parameter space is already ruled out by existing observations; axions with gravitational interactions and mass between $3\text{–}7 \times 10^{-13}$ eV would yield more than 10 signals in current LIGO data for all the BH mass and spin distributions considered here; masses between $2 \times 10^{-13}\text{–}2 \times 10^{-12}$ eV would yield 10 or more signals for the most optimistic spin distribution considered here [40]. An analysis of existing data taking into account the reduced event rates at larger self-interactions has not been performed and would be very valuable.

Once we incorporate self-interactions, there are three different parameter space regimes, with distinct behavior

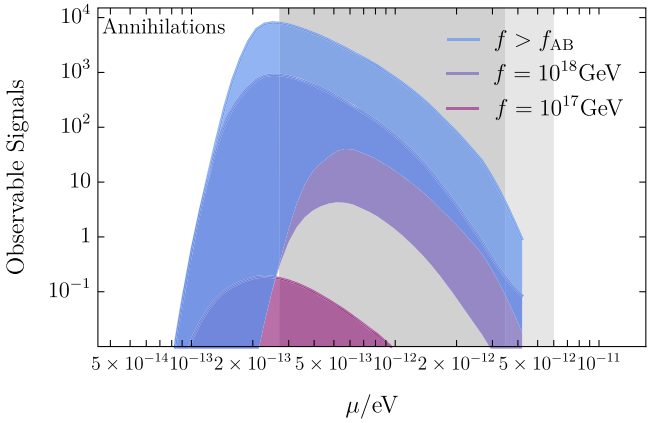


FIG. 14. Projections for the number of observable $211 \times 211 \rightarrow$ GW annihilation signals, using continuous wave searches at Advanced LIGO (with design sensitivity), for a range of self-interaction strengths (see text for details). The width of the bands results from varying the BH spin distribution and maximum BH mass as described in the text. The highest number of observable signals is in the small self-interactions regime, which includes gravitational superradiance and QCD axion parameter space. Increasing self-interactions reduces the number of signals expected. At high masses, the signal frequency falls above the band of typical CW searches ($\nu \gtrsim 2$ kHz). The darker (lighter) shaded regions are disfavored by black hole spin down for initially superradiating levels with $m = 1$ ($m = 2$) (see Sec. VI).

(as per Sec. IV). In the small self-coupling regime, $f > f_{\rm AB}$, the 322 level does not grow through self-interactions, and the dynamics proceeds as in the purely gravitational case. Consequently, the annihilation signal properties are independent of the self-coupling, and existing analyses of gravitational wave signals will apply without modification. This regime, which (for stellar mass BHs) includes $f \sim M_{\rm pl}$ as well as QCD axion self-couplings, can lead to as many as thousands of signals at LIGO/Virgo, as shown in Fig. 14.

In the moderate self-coupling regime, $f_{\rm AB} > f > f_{\rm BC}$, the growth of the 211 level is unaffected, but 322 grows earlier than it would otherwise have done. The main effect on the annihilation signal is through the addition of another energy loss process for the cloud, via $322 \times 322 \rightarrow 211 \times \infty$ emission. Consequently, while the peak emission amplitude is unaffected, the signal duration is reduced. This corresponds to the parameter space region between the orange and black dashed lines in the upper-right panel of Fig. 13. More specifically, when 211 is primarily depleted through gravitational waves, the signal strain as a function of time is given by,

$$h^{\rm GW,ann}(t) = \frac{h_{\rm peak}}{1 + t/\tau_{\rm ann}} \tag{97}$$

with $\tau_{\rm ann}$ defined in Eq. (39). However, due to the self-interaction processes, there is additional energy lost from the cloud, changing the time-evolution to that in Eq. (48), with

$$h^{\rm GW,ann}(t) \propto \sqrt{\tau_{\rm scalar}/t} \tag{98}$$

at late times, where $\tau_{\rm scalar} \propto (f/M_{\rm pl})^4$, Eq. (49). For f in the moderate self-coupling regime, $\tau_{\rm scalar}$ can be significantly less than $\tau_{\rm ann}$. Given the typical assumptions on black hole formation rates and distributions, the shortest signals that are likely to be observable in an all-sky continuous wave search have signal times on the order of 10^4 years or more [40].

Since, for moderate self-couplings, the peak signal strain is not affected, the sensitivity reach of gravitational wave detectors for signals observed around the optimum time is only moderately affected, as illustrated in the bottom panels of Fig. 13. One effect is that, especially for lighter black holes, the signal duration can become comparable to the typical coherent integration times used in continuous wave searches (e.g., [82]), which degrades the signal to noise.

For blind searches, the faster decrease of signal strain with time leads to less chance of seeing a signal, as illustrated in Fig. 14. The expected number of observable signals at $f \sim 10^{18}$ GeV, which is in the moderate self-interactions regime for $\mu \sim 10^{-12}$ eV, is around an order of magnitude lower than in the purely gravitational case. For larger and smaller μ, this value of f falls back into the weak self-interactions regime, so the difference is reduced. At $f \sim 10^{17}$ GeV, which is in the moderate self-interactions regime for the whole μ range, the signal durations are much shorter, and the expected number of observable signals is less than 1. As a result, such signals are unlikely to observed with current detectors, in a blind search. In addition, the faster time-evolution can lead to larger frequency drifts, which could degrade search sensitivity further (see Sec. VII C).

For strong self-couplings, $f > f_{\rm BC}$, the peak signal amplitude drops with increasing coupling as $(f/f_{\rm BC})^2$ (Fig. 13). In particular, this drop-off starts at larger f than for the suppression of BH spin-down, since $f_{\rm BC} > f_{\rm CD}$. Consequently, with current detectors, self-interactions strong enough to avoid BH spin-down constraints (Sec. VI) also render GW annihilation signals undetectable, for any plausible BH spin and mass distributions. For $f \lesssim f_{\rm BC}$, i.e., $f \lesssim 10^{16}$ GeV for stellar-mass BHs, the expected number of events in a blind search is $\lesssim 10^{-3}$, while for $f \lesssim 10^{15}$ GeV, where signal durations become comparable to those in the small self-interaction regime, signals beyond 10–100 pc are unlikely to be visible at Advanced LIGO sensitivities.

Nevertheless, it is possible that advanced future detectors, such as the Cosmic Explorer [89,90] or Einstein Telescope [91–94], may be able to probe this parameter space. The signal strain in the quasiequilibrium regime is a factor $\mathcal{O}(1$–$5)$ below the overshoot peak shown in the left

panel, but the quasiequilibrium regime lasts parametrically longer than in the moderate self-interaction regime, $\tau_{\rm sig} \propto (f_{\rm BC}/f)^2$ (see Fig. 18). If smaller strains come within reach of future detectors, the long-lasting signals would have an increased chance of being observed in the quasiequilibrium regime.

1. Additional annihilation channels

In addition to $211 \times 211 \to$ GW annihilations, as occur in the purely gravitational case, the presence of the 322 level allows $211 \times 322 \to$ GW and $322 \times 322 \to$ GW processes. These GWs will still have frequency $\omega \simeq 2\mu$, but due to the larger angular momentum of the 322 level, their rates are suppressed by higher powers of α, $P^{l,l'}_{\rm GW} \propto \alpha^{16+2(l+l')}$, where l and l' are the angular momentum numbers of the two levels. These powers are significantly smaller than the primary $211 \times 211 \to$ GW annihilation channel, and are further suppressed by the smaller occupation number of 322 at small α (Appendix D). For example, the $322 \times 322 \to$ GW process would lead to signals strains $\mathcal{O}(10^{-4})$ weaker than the primary signal at $\alpha \sim 0.3$. "Cross-annihilation" signals between two levels, $211 \times 322 \to$ GW, may be observable for the closest black holes; further study would require numerical GW power calculations which have not yet been performed for cross-annihilation signals.

2. Annihilation signals from complex fields

In this section, and throughout the rest of this paper, we have considered superradiance of a single, real spin-0 field. As has been pointed out in a number of papers [95–98], for the case of two scalar fields of degenerate masses (equivalently, a single complex scalar field), there are cloud configurations with a time-independent stress-energy tensor, which consequently do not emit any gravitational radiation. In complex field terms, these correspond to all-particle or all-antiparticle field configurations, whereas gravitational waves arise from particle-antiparticle annihilation. This has sometimes been interpreted [99] as indicating that annihilation radiation, of the type considered in this section, is not expected from superradiance of complex fields.

However, as per the discussion in Sec. II, the initial conditions for the growth of superradiant modes are either vacuum fluctuations, or whatever preexisting astrophysical fields are present. In the former case, we can view the growth of the particle and antiparticle field modes as effectively separate, and generically, they will obtain $\mathcal{O}(1)$-similar occupation numbers. For preexisting astrophysical fields, a generic expectation in many circumstances is for $\mathcal{O}(1)$-similar initial conditions for particle and antiparticle fields. Consequently, unless some mechanism drives us to an all-particle or all-antiparticle state, we expect that the particle and antiparticle fields generically attain roughly comparable occupation numbers. Compared to a real scalar field, this results in a total GW annihilation signal energy that is only $\mathcal{O}(1)$ smaller.

B. Transitions

For large enough self-interactions (regions B,C,D in Fig. 3), the 322 level grows earlier than it would have done otherwise, and both 211 and 322 can have significant occupation numbers at the same time. This gives rise to GW transition signals.

The transition quadrupole moment for the $322 \to 211 + \text{GW}$ process vanishes at leading order, so its rate is suppressed by a larger power of α than other gravitational transition processes (such as the $644 \to 544$ process considered in [6,12,15]). At leading order in α, the emitted power, as a function of polar angle θ, is

$$\frac{dP}{d\Omega} = \frac{GN_{322}N_{211}}{\pi r_g^4}\alpha^{14}\left(\frac{2^5}{3^6 5^8}(1-\cos^4\theta) + \frac{(27+28\cos(2\theta)+9\cos(4\theta))\sin^2\theta}{2^2 3^6 5^{10} 7^2}\right) \tag{99}$$

where the first term corresponds to l, $m = 2$, 1 emission, and the second to l, $m = 3$, 1. This gives a total emitted power of [6]

$$P = \frac{2^8 \times 5717}{3^5 5^{11} 7^3}\frac{GN_{322}N_{211}}{r_g^4}\alpha^{14}. \tag{100}$$

The emitted radiation is at a frequency $\omega = \omega_{322} - \omega_{211} \simeq \frac{5}{72}\alpha^2\mu$. In terms of the normalized occupation numbers, it contributes a term

$$\dot{\varepsilon}_{322} \simeq -5 \times 10^{-6}\alpha^{10}\varepsilon_{211}\varepsilon_{322} + \ldots \tag{101}$$

to the equations of motion.

Compared to the processes discussed in Sec. IV, which drive the evolution of the superradiant cloud, the effects of GW transitions are always subdominant. While this does reduce the peak signal amplitude, it also means that signal timescales can be longer compared to the transitions in the purely gravitational regime, which is helpful for detection.

Figure 15 shows projections for the peak signal strain, and sensitivity reach, for transition signals from a fairly light BH, $M_{\rm BH} = 3\ M_\odot$. The signal durations (for a given BH mass) are the same as those for annihilations (Fig. 13) in the region where 322 grows, $f < f_{\rm AB}$, as the two levels evolve together over time. Given the lower frequency compared to annihilations, the signal strains are typically larger (Fig. 15 left). However, transition signals only occur in the moderate and large self-interaction regimes, where much of the energy loss is through scalar radiation. Furthermore, for given BH mass, the frequency decreases $\propto \mu^3$ with decreasing μ, rapidly falling out of the sensitivity

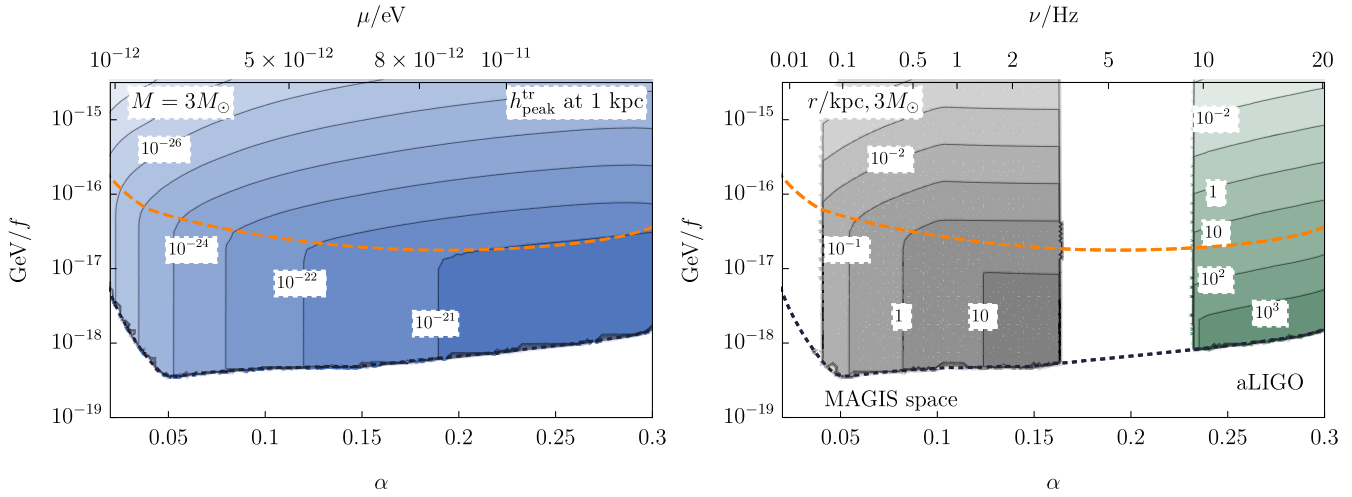


FIG. 15. Left panel: peak strain of the $322 \to 211 + \mathrm{GW}$ transition signal at 1 kpc from a BH of mass 3 $M_\odot$, as a function of the mass μ and self-coupling scale f of the scalar particle. Right panel: sensitivity reach for the detection of such signals, using the Advanced LIGO detector, or with the MAGIS proposal for a future space-based atom interferometer [41]. The dashed orange and dotted black lines are the f_{BC} and f_{AB} curves, respectively, as in Fig. 13.

band of current detectors such as Advanced LIGO. For heavier BHs, the frequency of transition signals would always be too low for ground-based GW detectors, due to overwhelming seismic and gravity-gradient noise.

For a narrow range of axion masses above 10^{-11} eV, current detectors could potentially probe signals in the moderate self-interaction regime (Fig. 15, right). Although the reach is poor at small f, there is a roughly order-of-magnitude range in f for which sensitivity to signals from the galactic center would be possible. The signal times in this region last on the order of minutes to hours, and the expected number of signals in a blind search is heavily dependent on the poorly measured black hole distribution in the "mass gap" below $5\,M_\odot$ [100–103] (although evidence for compact objects in this mass range is emerging [104–106]). Consequently, blind searches with current detectors are unlikely to lead to observable signals.

However, future space-based detectors such as LISA [107,108] and atom interferometer missions [41], could have promising sensitivity to such signals. For illustration we show the reach of the MAGIS proposal [41] in the right panel of Fig. 15, which can achieve a reach of 10 kpc for axions around $3\,M_\odot$ black holes, and up to 10^3 kpc for $100\,M_\odot$ black holes. Some of the more promising signals fall in the 0.1–10 Hz range, where future proposals such as DECIGO [109] could improve transition detection prospects.

C. Frequency drifts

While the frequency of gravitational wave annihilation signals is almost constant at $\nu_{\mathrm{ann}} \equiv 2\omega/(2\pi) \simeq 2\mu/(2\pi)$ (we will use frequency rather than angular frequency in this section, to match the GW literature), the potentially long signal durations mean that even very small frequency drifts can be measured. Moreover, the search algorithms employed in continuous wave detection analyses can be strongly affected by these small frequency drifts, so it is important to quantify them to determine the appropriate search strategy and sensitivity [110].

The self-energy of the cloud, from both gravity and self-interactions, affects the frequency of the bound axions, and therefore the frequency of the GWs emitted [12]. As the occupation numbers of the levels evolve, the self-energy contribution to the binding energy $\Delta\omega$ and thus the emitted frequency ν change over time.

The gravitational and self-interaction contributions to the energy of axions in level 211 are, respectively, (see Appendix G and Appendix B 1)

$$\Delta\omega_g \simeq -0.19\mu\alpha^3\varepsilon_{211} \tag{102}$$

$$\Delta\omega_\lambda \simeq -3.5\times10^{-5}\mu\alpha^5\varepsilon_{211}\left(\frac{M_{\mathrm{pl}}}{f}\right)^2, \tag{103}$$

where the energy is decreased (increased) in the presence of an attractive (repulsive) self-interaction. These corrections are always small compared to the axion mass, as well as the energy splitting between levels (for occupation numbers below the nonperturbative regime—see Sec. V).

As the cloud is growing through superradiance, the frequency changes relatively rapidly as $\propto \mu\alpha\dot{\alpha}$ on the order of the superradiance time due to the changing BH mass. However this period is short, and generally does not contribute much of the detectable signal. At late times, the cloud size is depleted over time, and the level's frequency drift is positive (assuming negligible or attractive self-interactions). This is in contrast to standard astrophysical sources of continuous gravitational radiation, such as

spinning neutron stars, and may provide a hint that a detected signal arises from superradiance. We describe the main contributions to these frequency drifts, at leading order in α, below. For a more complete discussion of frequency drifts we refer the reader to Appendix H.

At small self-interactions, the frequency drift is dominated by the depletion of the gravitational self-binding energy through annihilations, resulting in a frequency drift of order

$$\dot{\nu}_{\rm ann} \simeq 7 \times 10^{-15}\ \frac{\rm Hz}{\rm s}\left(\frac{\alpha}{0.1}\right)^{17}\left(\frac{\mu}{10^{-12}\ {\rm eV}}\right)^{2}, \quad (104)$$

to leading order in α. Throughout the small self-interaction regime $f > f_{\rm AB}$ (see also Fig. 3), the gravitational frequency drift dominates any contribution from the self-interactions.

As self-interactions increase, the frequency drift from the gravitational binding energy is increased due to the faster depletion of the cloud from axion emission,

$$\dot{\nu}_{\rm g} \simeq 10^{-10}\frac{\rm Hz}{\rm s}\left(\frac{10^{17}\,{\rm GeV}}{f}\right)^{4}\left(\frac{\mu}{10^{-12}\,{\rm eV}}\right)^{2}\left(\frac{\alpha}{0.1}\right)^{17}, \quad (105)$$

and there is an additional frequency drift from the change of self-interaction energy,

$$\dot{\nu}_{\lambda} \simeq 10^{-10}\frac{\rm Hz}{\rm s}\left(\frac{10^{17}\,{\rm GeV}}{f}\right)^{6}\left(\frac{\mu}{10^{-12}\,{\rm eV}}\right)^{2}\left(\frac{\alpha}{0.1}\right)^{19}. \quad (106)$$

The latter dominates when $f \lesssim 8.5 \times 10^{16}\ {\rm GeV}(\alpha/0.1)$. Finally, in the strong self-interactions regime $f < f_{\rm BC}$, the cloud reaches a long-lived quasiequilibrium configuration, and the dominant source of frequency drifts comes from the slow spindown of the BH.

Gravitational wave signals from $322 \to 211 + {\rm GW}$ transitions have frequency $\nu_{322} - \nu_{211}$, so the changing contributions to the 211 and 322 frequencies partially cancel, making frequency drifts a factor of a few smaller than for annihilations, and negative in most parts of the parameter space. Similarly to annihilations, for moderate self-couplings, self-interactions dominate the frequency drifts for $f \lesssim 10^{17}\ {\rm GeV}(\alpha/0.1)$.

At small α, the frequency drift can be small enough so as to be unobservable. Over a year, the minimum frequency change that can be measured is $\sim {\rm yr}^{-1} \sim 3 \times 10^{-8}$ Hz, so if the frequency drift is $\lesssim {\rm yr}^{-2} \simeq 10^{-15}\ {\rm Hz\,s^{-1}}$, it has no observational effect. At the other extreme, too large a frequency drift can be problematic for the search algorithms employed. Current LIGO/Virgo continuous wave searches cover a range of positive to negative frequency derivatives of e.g., 2×10^{-9} Hz/s through -1×10^{-8} Hz/s [83]. More sensitive searches, using longer coherent integration times, may require even smaller frequency drifts [75]. In the small coupling regime, the drift of the signal becomes larger than this threshold at $\alpha \sim 0.25$. In the moderate self-interactions regime, both annihilation and transition signals have drifts large compared to the current search range for $f \lesssim 5 \times 10^{16}\ {\rm GeV}(\alpha/0.1)^{17/4}$. However, as discussed above, the observational prospects for GW signals at such small f are not promising, with current-generation experiments.

VIII. AXION WAVES

As well as emitting gravitational radiation, the cloud also emits both relativistic (Sec. III A) and nonrelativistic (Sec. III B) scalar waves. If the scalar φ has nongravitational interactions[11] with the SM, such φ radiation could be detected in laboratory experiments. For an axionlike particle, a natural assumption is that interactions with the SM are suppressed by parametrically the same symmetry breaking scale f that sets the axion potential. If this is the case, then we have the unusual feature that, in the large self-coupling regime $f < f_{\rm BC}$, the signal does not decouple: while the power in axion radiation decreases as the quasiequilibrium size of the cloud decreases, this is compensated for by the increased interaction strength from the smaller f. In addition, the BH spin-down time increases with decreasing f, so such signals can last for very long times, increasing the chance of observing them. Consequently, axion waves could be a probe of the small-f regime, in which both GW and spin-down signatures are suppressed.

Quantitatively, if we take the 211 and 322 quasiequilibrium occupation numbers (55), then the emitted power is dominated by nonrelativistic $322 \times 322 \to 211 \times \infty$ radiation. At large distances r from the BH, this radiation has energy density

$$\begin{aligned}\rho_{\rm rad} &\sim \frac{\mu}{4\pi r^2}\frac{GM^2\gamma^{211\times\infty}_{322\times322}(\varepsilon^{\rm eq}_{322})^2\varepsilon^{\rm eq}_{211}}{v}\\ &\simeq 10^{-6}\ {\rm GeV/cm^3}\left(\frac{\alpha}{0.1}\right)^{6}\left(\frac{10\ {\rm kpc}}{r}\right)^{2}\\ &\quad\times\left(\frac{f}{10^{16}\ {\rm GeV}}\right)^{2}, \end{aligned} \quad (107)$$

where $v = \alpha/6$ is the velocity of the nonrelativistic axions emitted. The energy density $\rho_{\rm rad}$ depends only on α, and not on μ and $M_{\rm BH}$ independently. For given f, the emitted power is maximized when the superradiance rate is largest, at high a_* and α. The corresponding dimensionless amplitude θ of the axion waves is

[11] If the scalar φ's interactions with the SM are purely gravitational, then its interaction rate with matter is $\propto G^2 \sim 1/M_{\rm pl}{}^4$, whereas for gravitational radiation, the interaction rate is $\propto G \sim 1/M_{\rm pl}{}^2$. Consequently, such φ radiation would be practically undetectable.

$$\theta \simeq 10^{-19}\left(\frac{10^{-12}\text{ eV}}{\mu}\right)\left(\frac{\alpha}{0.1}\right)^3\left(\frac{10\text{ kpc}}{r}\right), \tag{108}$$

independent of f. This is in contrast to GW signals, for which the amplitude at Earth decreases as f^2 in the quasiequilibrium regime. Relativistic axion radiation from the $3 \to 1$ process (Sec. III A), and $2 \to 1$ cubic emission, also have f-independent θ, but are suppressed by higher powers of α, and are smaller than the nonrelativistic radiation for the parameter space we are interested in.

As we discussed in Sec. IV C, for $\alpha \gtrsim 0.2$ and small f we expect additional hydrogenic levels, other than 211 and 322, to be populated. While we have not performed a full analysis in this regime, a example of the possible effects can be seen from the 411 build-up studied in Sec. IV C 2, which for α not too far above 0.2 is expected to be the first additional level to grow. The 211, 322, and 411 levels form a new quasiequilibrium, with the 411 level having enhanced occupation number relative to those of the 211/322 equilibrium. Consequently, the rate of scalar radiation during this equilibrium is enhanced; numerically, we find that $\rho^{\text{rad}}_{3-\text{level}} \sim 25\rho^{\text{rad}}_{2-\text{level}}$ for $\alpha \simeq 0.3$. While this equilibrium will be disrupted in turn by the growth of further levels, this illustrates that, while the parametric behavior in f should remain the same, additional levels may change the numerical factors affecting the scalar radiation power. As discussed in Sec. V, if the growth of additional levels leads to large enough field amplitudes in the cloud, then higher-order processes or a nonperturbative collapse of the cloud may become possible, significantly altering the behavior.

Since the axion radiation is nonrelativistic and narrow-bandwidth, its effects on a laboratory system are similar to those of axion dark matter at the same mass. The masses of interest correspond to rather low frequencies, e.g., 10^{-12} eV $\simeq 2\pi \times 200$ Hz. For this parameter space, the axion-SM couplings most amenable to laboratory detection experiments are those to nuclear spins and to photons, which we discuss below.

Searches for axion DM via the axion-gluon coupling $\mathcal{L}_{\text{int}} \propto (\varphi/f)G_{\mu\nu}\tilde{G}^{\mu\nu}$ have promising sensitivity reach at low axion masses [111]. However, if an axionlike particle has the same $G\tilde{G}$ coupling, but a smaller mass than the QCD axion (or equivalently, a larger $G\tilde{G}$ coupling for the same mass), then it is strongly constrained by its behavior in dense environments such as the early universe and stellar cores [112,113]. For superradiance-sourced signals, $G\tilde{G}$ couplings significantly higher than the QCD axion value (for a given axion mass) are needed to have experimental sensitivity, and are affected by these constraints.

A. Nucleon spin coupling

The axion coupling to fermion spins is $\mathcal{L} \supset g_N(\partial_\mu\varphi)\bar{\psi}\gamma^\mu\gamma^5\psi$, where we generically expect $g_N \sim 1/f_a$. For a nonrelativistic fermion, this gives an axion-dependent term in the fermion Hamiltonian,

$$H \supset g_N\vec{\sigma}\cdot(\nabla\varphi + \dot{\varphi}\,\vec{v}) \tag{109}$$

where $\vec{\sigma}$ is the fermion's spin, and $\vec{v}$ is its velocity. We will focus on couplings to nucleons, which for low axion frequencies are easier to detect than couplings to electrons.

Since the $322 \times 322 \to 211 \times \infty$ axion radiation from the BH has $v \sim \alpha/6$ (Sec. III B), while the nucleon velocity changes associated to low-energy laboratory processes are much smaller, the "axion wind" term $H_{\text{wind}} = g_N\vec{\sigma}\cdot\nabla\varphi$ dominates. Due to the $\sim\alpha/6$ velocity being significantly larger than the virial velocity of DM in the galaxy, $\sim 10^{-3}$, and because of the coherent nature of the emitted radiation, an experiment searching for the axion wind coupling will have better sensitivity to BH-sourced radiation than it would for DM for an equivalent axion energy density.

The best-developed experimental proposal aiming to detect the axion wind coupling is CASPEr-Wind [111], which employs nuclear magnetic resonance (NMR) technologies. This uses a liquid xenon target, whose nuclear spins are polarized in a strong magnetic field. The axion wind coupling acts on the nuclei like an effective magnetic field, $H_{\text{wind}} = g_N\vec{\sigma}\cdot\nabla\varphi \equiv B_a\cdot\vec{\mu}_n$, where μ_n is the nuclear magnetic moment and B_a is the effective axion "magnetic field." If this effective magnetic field oscillates at close to the Larmor frequency of the nucleons in the external magnetic field, then the resulting spin precession of the nuclei is resonantly enhanced. This spin precession can then be picked up by a sensitive magnetometer.

In Appendix J, we review the sensitivity of such experimental setups to a monochromatic axion oscillation. If we are uncertain about the axion mass, and want to experimentally probe an $\mathcal{O}(1)$ axion mass range around an angular frequency ω_0, then a signal can be detected for

$$B_a^2 \gtrsim \text{few} \times \frac{\omega_0}{\mu_n^2 N_n T_{\text{tot}}}, \tag{110}$$

where T_{tot} is the total experimental running time, and N_n is the number of aligned spins in our spin-polarized sample.[12] This is a best-case sensitivity estimate, limited by the fundamental spin-projection noise of the sample—to achieve it, a well-shielded sample and a sufficiently sensitive magnetometer would be required. Experiments capable of sensing nuclear spin projection noise have been carried out [114], and such sensitivities are a goal for the CASPEr-Wind experimental program [111].

[12]This sensitivity estimate is for the detection of a single, monochromatic signal. As mentioned in Sec. VII, in situations where many galactic sources are emitting at any given time, it may be more effective to perform a "stochastic" search, looking for multiple unresolved signals within a given bandwidth. We leave analysis of such scenarios to future work.

A fully polarized liquid ^{129}Xe sample has $\sim 10^{22}$ spins/cm^3 [111], so the sensitivity limit for a relatively small target volume is

$$B_a \gtrsim 10^{-20}\ \mathrm{T}\sqrt{\frac{\nu}{\mathrm{kHz}}\frac{10^{22}}{N_n}\frac{\mathrm{yr}}{T_{\mathrm{tot}}}} \tag{111}$$

For comparison, an axion DM signal at the sensitivity threshold estimated in [111], for these parameters, has an effective magnetic field of $\sim$few $\times\ 10^{-20}$ T. The effective magnetic field from axion radiation emitted by a superradiant cloud is

$$B_a \simeq 3 \times 10^{-24}\ \mathrm{T} \times \mathcal{C}_N \left(\frac{\alpha}{0.1}\right)^4 \left(\frac{1\ \mathrm{kpc}}{r}\right), \tag{112}$$

for a high-spin BH, where $\mathcal{C}_N \equiv g_N f$. Consequently, some combination of larger experimental volumes (as planned for CASPEr-Wind phase II [111]), larger $\mathcal{C}_N$, larger α and a closer BH would enable laboratory experiments to be sensitive to axion waves.

This is illustrated in Fig. 16, which shows projected signal strengths for a selection of astrophysical BHs (both nominal and observed), along with sensitivity thresholds for different experimental configurations. While $\mathcal{C}_N \sim \mathcal{O}(1)$ is the 'natural' expectation in many models, larger values of $\mathcal{C}_N$ are possible. In particular, it is interesting to consider

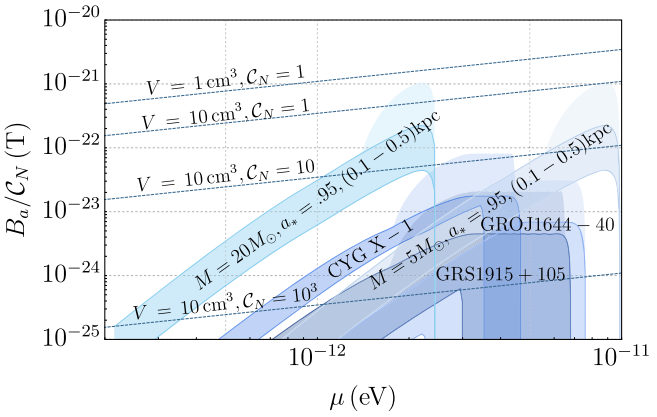


FIG. 16. Projected detectability of nonrelativistic axion radiation, assuming an axion-nucleon coupling. The signal strength is expressed in terms of the equivalent pseudo-magnetic field felt by nuclei. The blue dotted lines correspond to sensitivity estimates for NMR axion-wind detection experiments [111] with the indicated parameters. The bands correspond to signals from three astrophysical BHs and two nominal BHs with the indicated parameters. The widths of the bands correspond to the uncertainty on the BH parameters (for the nominal BHs, to the distance range indicated). The darker bands bounded by solid contours correspond to the signal emitted during two-level quasiequilibrium (Sec. IV). The lighter-shaded extensions above represent the enhanced signal from the three-level equilibrium with 411 (Sec. IV C 2), illustrating the potential range of signals.

how large a reach can be obtained in as-yet-unconstrained parameter space, below the existing astrophysical limits of $g_N \lesssim (\text{few} \times 10^8\ \mathrm{GeV})^{-1}$ [115–118]. While much of the axion mass range in Fig. 16 is excluded for large f by BH spin measurements (Fig. 10), these constraints do not apply for $f \lesssim 10^{12}$–10^{13} GeV, where the BH spin-down is too slow. The astrophysical bounds translate into $|\mathcal{C}_N| \lesssim 10^3 (f/10^{12}\ \mathrm{GeV})$; the $\mathcal{C}_N = 10^3$ line in Fig. 16 illustrates that such couplings can give good detection prospects for a wide range of BHs and axion masses.

To reflect the uncertain behavior of the superradiant cloud at $\alpha \gtrsim 0.2$, Fig. 16 displays the signal resulting from the radiation power during the three-level quasiequilibrium phrase, as a shaded area above the signal from the two-level equilibrium. The signal curves illustrate that, with larger-volume experiments, sensitivity to astrophysical BHs may be possible for $\mathcal{C}_N \sim \mathcal{O}(1)$. They also strongly motivate detailed numerical analyses of the high-α regime, where the strongest signals would arise.

Figure 17 displays the sensitivity reach to an optimal BH for a given axion mass. Again, we see that for larger experimental volumes, astrophysically relevant reaches—in particular, to the Galactic Center $\sim$8 kpc away—may be possible for fairly natural $\mathcal{C}_N$ values.

If we are interested in the signal from a specific, known BH, then the sensitivity reach is the most important parameter. However, as is the case for gravitational wave searches, many signals are expected to arise from as-yet-unobserved BHs, and could only be detected via a "blind",

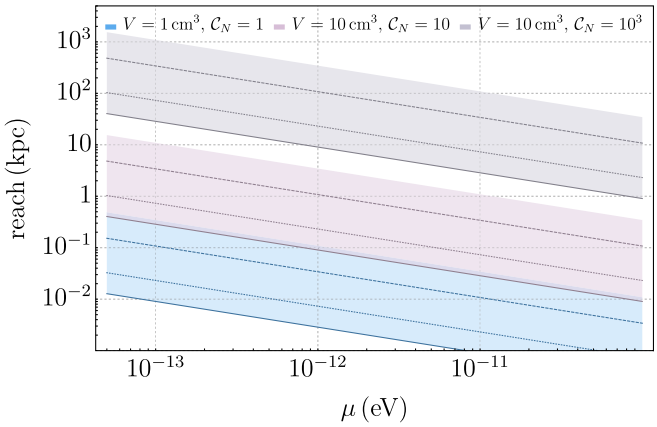


FIG. 17. Projected sensitivity reach (SNR = 1) for the detection of nonrelativistic axion waves from a BH-cloud system at large self-interactions, $f < f_{\mathrm{BC}}$, in a NMR-based axion-wind detection experiment. The bands show the reach to a BH-cloud system, ranging from a two-level quasiequilibrium with parameters $a_* = 0.9$ and $\alpha = 0.2$ (lower edge, solid), to that of a three-level quasiequilibrium system with $a_* = 0.99$ and $\alpha = \alpha^{\mathrm{optimal}}(0.99) \approx 0.41$ (upper edge). The reach for a BH-cloud system with $a_* = 0.9$ and $\alpha = \alpha^{\mathrm{optimal}}(0.9) \approx 0.28$ is also indicated for a two-level equilibrium (dotted line), and a three-level equilibrium (dashed line) system.

all-sky search. In this situation, another important factor is the typical duration of signals, which affects the probability that a given BH is still emitting today. Figure 18 shows the duration of the peak axion signal (which contributes most of the detectable SNR) from a nominal BH, as a function of axion mass and coupling. Lower f values lead to slower BH spin-down, and so to longer durations of quasi-equilibrium signal emission; this is relevant down to $f \sim 10^{11}$–10^{12} GeV, below which signals can last longer than the age of the universe.

Since, in the quasiequilibrium regime, the peak signal strength at Earth is independent of f for fixed $\mathcal{C}_N$, decreasing f down to $\sim 10^{11}$ GeV increases the expected number of events in a blind search. This is illustrated in Fig. 19. If, rather than fixing $\mathcal{C}_N$, we require that g_N is below the astrophysical bounds, then as shown in Fig. 20, there is a wide range of axion masses over which we might expect visible signals in an all-sky search (depending on the mass and spin distribution of astrophysical BHs). In both the Fig. 19 and Fig. 20 projections we assume the reach to the axion waves from the two-level equilibrium, not taking into account the possible enhancements in power from additional levels; on the other hand, the dynamics of additional levels could shorten the signal lifetime at large α values. In the blind search, an analysis similar to the techniques employed by Continuous Waves searches at LIGO/Virgo (Sec. VII A) would be required, to make use of the extremely long signal coherence times while at the same time taking into account the Doppler shifts from the many relative motions between the experiment and the unknown black hole positions.

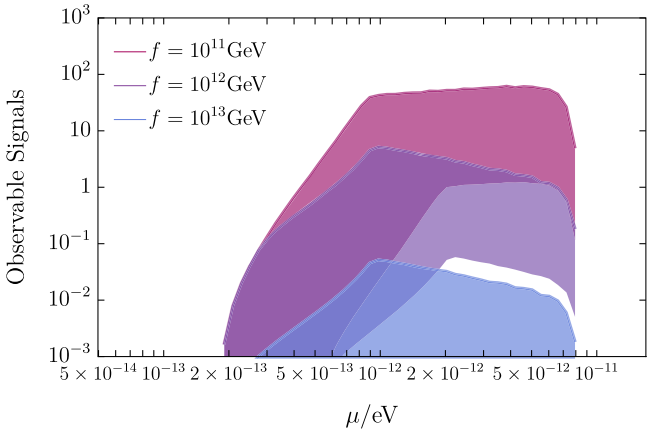


FIG. 19. Number of observable signals expected in an NMR axion wind experiment with $V = 10$ cm^3 and $\mathcal{C}_N = 100$, with different bands corresponding to different quartic coupling scales f. We require observable signals to have SNR ≥ 10, given the blind search strategy required for these events. The width of the bands results from varying the assumed BH spin distribution and maximum BH mass (see Sec. VII A). For a fixed $\mathcal{C}_N$, the number of observable signals increases for smaller f, due to longer signal durations, saturating at $f \sim 10^{11}$ GeV.

Unless $\mathcal{C}_N$ is extremely large, the effects of the axion field on spins in the vicinity of the black hole, and the effect of these spins on the axion field, are always small. The largest effective magnetic field obtained in the cloud is $\sim|\mathcal{C}_N|10^{-6}\mathrm{T}\frac{\mu}{10^{-12}\ \mathrm{eV}}$, which would not have any significant affect on accretion disk behavior. Similarly, the axion field sourced by a coherent nuclear spin density, if any exists in the accretion disk, is tiny compared to the fields of a

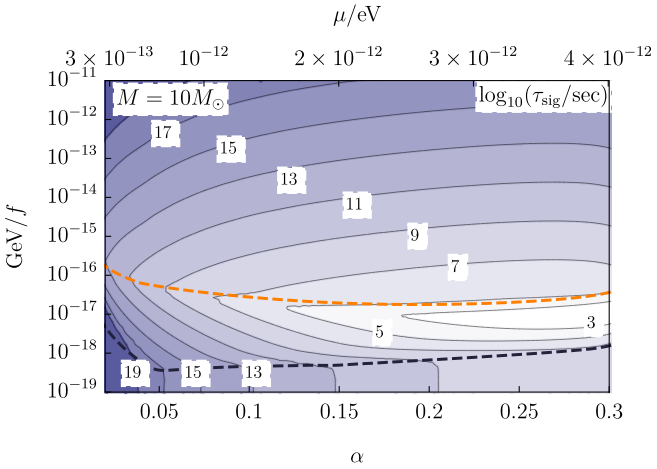


FIG. 18. Typical duration $\log_{10}(\tau_{\rm sig}/\ \mathrm{sec})$ the axion wave signal for a 10 $M_\odot$ BH with initial spin 0.9. In the large self-interactions regime, we show the time-scale corresponding to the quasiequilibrium evolution. For $f \lesssim 10^{12}$ GeV the signals can last longer than the age of the universe (note that 10 Gyr $\simeq 3 \times 10^{17}$ s). The dashed orange and dotted black lines are the $f_{\rm BC}$ and $f_{\rm AB}$ curves, respectively, as in Fig. 13.

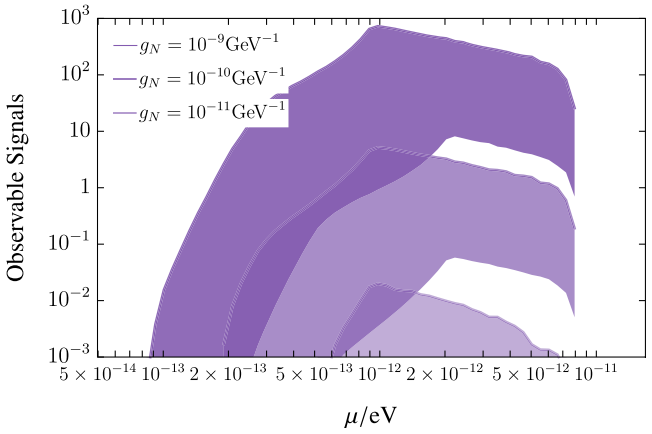


FIG. 20. Number of observable signals expected in an NMR axion wind experiment with $V = 10$ cm^3, for $f = 10^{12}$ GeV and different couplings to nuclear spins, as shown. We require observable signals to have SNR ≥ 10, given the blind search strategy required for these events. The width of the bands results from varying the assumed BH distribution as in Fig. 19. For a fixed self-interaction strength, the highest number of observable signals is for the largest coupling strength to nuclei.

superradiant cloud. For any reasonable nuclear spin response to small magnetic field perturbations, the effect of spin response on the dynamics of quasibound axion levels will be extremely small, so the growth of the cloud will not be affected. Similar considerations apply to the propagation of scalar waves through interstellar space; these will be undisturbed to a very good approximation.

B. Photon coupling

The axion coupling to photons is $\mathcal{L} \supset -\frac{g_{a\gamma\gamma}}{4}\varphi F_{\mu\nu}\tilde{F}^{\mu\nu} = g_{a\gamma\gamma}\varphi \boldsymbol{E}\cdot\boldsymbol{B}$. Generically, we expect the coupling constant to be $g_{a\gamma\gamma} = \mathcal{C}_\gamma \frac{\alpha_{\rm EM}}{2\pi f_a}$, where $\mathcal{C}_\gamma \sim \mathcal{O}(1)$ is related to the charged matter content of the UV theory [119].

An axion oscillation sources EM fields through the effective current density $J_a = g_{a\gamma\gamma}(\dot{\varphi}\boldsymbol{B} + \nabla\varphi\times\boldsymbol{E})$ (and the corresponding effective charge density $\rho_a = -g_{a\gamma\gamma}\nabla\varphi\cdot\boldsymbol{B}$). Axion DM, which is nonrelativistic, has $|\dot{\varphi}| \gg |\nabla\varphi|$, so detection experiments use strong magnetic fields to maximize J_a. Searches for low-frequency (∼kHz) axions have been proposed using static background magnetic fields [120,121], or GHz-frequency fields in superconducting cavities [122–125].[13]

If they can be realized in the future, quantum-limited meter-scale experiments could probe axion DM couplings as small as $g_{a\gamma\gamma} \sim 10^{-17}\ \mathrm{GeV}^{-1}$ at ∼kHz frequencies (unfortunately, this is still far from QCD axion sensitivity). With a monochromatic signal, as opposed to virialized axion DM, this would correspond to a sensitivity of $g_{a\gamma\gamma} \sim 10^{-18}\ \mathrm{GeV}^{-1}(\frac{\rho}{\rho_{\rm DM}})^{-1/2}$.[14] For nonrelativistic emission from a superradiant cloud, we would obtain a reach of

$$\frac{r}{\mathrm{kpc}} \approx (2\times 10^{-3})|\mathcal{C}_\gamma|\left(\frac{\mu}{10^{-12}\ \mathrm{eV}}\right)\left(\frac{\alpha}{0.1}\right)^3, \tag{113}$$

Consequently, signals from an superradiant cloud via the axion-photon coupling could only be seen for an exceptionally close, fast-spinning BH, and/or in models where $|\mathcal{C}_\gamma|$ is large.

At the small axion masses we are interested in, SN1987A observations constrain the axion-photon coupling to be $|g_{a\gamma\gamma}| \lesssim 5\times 10^{-12}\ \mathrm{GeV}^{-1}$ [129]. This translates to $|\mathcal{C}_\gamma| \lesssim 500(f/10^{11}\ \mathrm{GeV})$, which allows for somewhat smaller expected blind-search event rates than the nucleon-coupling case shown in Fig. 20.

Similarly to the case of nucleon couplings, the effects of astrophysical EM fields on the SR cloud will be tiny unless $|\mathcal{C}_\gamma| \gg 1$. In addition, the naive $\varphi \to \gamma\gamma$ decay rate, $\Gamma_{\varphi\to\gamma\gamma} \simeq \frac{g_{a\gamma\gamma}^2\mu^3}{64\pi}$, is much longer than the age of the universe for couplings of interest. However, in some circumstances it is possible for parametric resonance to greatly enhance the photon emission rate [130]. Parametrically, in the limit where $g_{a\gamma\gamma}$ is arbitrarily small, and taking L to be the approximate spatial extent of the axion profile, the total decay rate into a particular mode within the $\sim L^3$ volume is $\Gamma \sim g_{a\gamma\gamma}^2\varphi^2\mu^2 L$, where φ is the typical field amplitude. Consequently, the number of photons emitted into that mode, in the light-crossing time $\sim L$, is $\sim \Gamma L \sim g_{a\gamma\gamma}^2\varphi^2(\mu L)^2$. This tells us that for finite $g_{a\gamma\gamma}$, if $\Gamma L \gtrsim 1$, then stimulated emission will become important; for $\Gamma L \gg 1$, the emission rate will be exponentially enhanced.

This parametric argument agrees with the conclusions of [130], which analyses the growth of electromagnetic perturbations using Floquet theory, and finds that parametric resonance occurs if

$$|g_{a\gamma\gamma}\mu\varphi L| \gtrsim \text{few}. \tag{114}$$

Since $g_{a\gamma\gamma}\varphi = \mathcal{C}_\gamma\frac{\alpha}{2\pi}\theta$, the left-hand side (lhs) is maximized (for given $\mathcal{C}_\gamma$) by maximizing θ. For an axion of mass μ, this occurs at $f \simeq f_{\rm BC}$ (for the 211 level). Using Eq. (80), we find that for parametric resonance to occur, we need

$$|\mathcal{C}_\gamma| \gtrsim (9\times 10^2)\left(\frac{0.1}{\alpha}\right)^{3/4}, \tag{115}$$

for $a_*(t_0) = 0.99$. Consequently, if $|\mathcal{C}_\gamma| \ll 10^3$, then photon emission will be unimportant.

It should be noted that the above is a best-case estimate, which will only hold if the BH is in a sufficiently pristine environment. The plasma frequency in the interstellar medium is $\omega_p \sim 10^{-12}$–10^{-10} eV, which is comparable to the mass range for a superradiant axion around a stellar-mass BH. Moreover, one expects the plasma density in the vicinity of the BH to be greater, due to accretion [131]. Consequently, it is likely that plasma effects suppress the parametric resonance process, even at large $|\mathcal{C}_\gamma|$ [132].

IX. CONCLUSIONS

In this paper, we have investigated some of the most important consequences of scalar self-interactions for superradiance around astrophysical BHs. As we have showed, self-interactions can result in very rich and complicated dynamics, and there are a number of aspects which would benefit from further study. In particular, we have not systematically treated situations in which the initially fastest-growing level has $m \geq 2$. While we

[13]Experiments using optical-frequency fields have also been proposed [126–128], but these have significantly worse theoretical sensitivity.

[14]The ideal search strategy for monochromatic signals may be different from that for a virialized axion signal with non-negligible bandwidth. For static-field experiments such as those proposed in [121], an optimal search for monochromatic signals will overcouple the amplifier even more strongly to the pickup. However, for ∼kHz axion frequencies and practical temperatures, optimal axion DM experiments would already be strongly overcoupled (to the point of having almost $\mathcal{O}(1)$ fractional sensitivity bandwidth [121]), so there would not be a significant difference between the monochromatic and DM search strategies.

generally expect gravitational (and scalar) wave signatures to be dominated by cases where 211 grows first, BH spin-down constraints for higher-mass axions will depend on higher-m superradiance.

In addition, even for the 211 case, our calculations have been at the (semi-)analytic level, and may not be reliable for large enough α. In particular, we found that for $\alpha \gtrsim 0.2$ and small f, levels other than 211 and 322 might play an important role in the dynamics. One route to properly understanding the high-α regime might be to perform numerical simulations of the (self-interacting) field equations themselves, rather than of the occupation numbers of hydrogenic modes. Such approaches have been used to study purely gravitational superradiance in a number of papers [133–137]. As mentioned in Sec. V, numerical methods were applied to a self-interacting scalar field on the Kerr background by [27,28], but they did not evolve the system for long enough to observe the perturbative effects we have studied. Since the high-α regime is where observational signatures may be the strongest, and in which there is the possibility of phenomena such as bosenova, a fuller treatment would be valuable.

Our analyses focussed on the simplest form of self-interactions for a spin-0 particle; the lowest-order (renormalizable) potential terms. In more complicated hidden sector models, other forms of interactions, or extra hidden sector states, could affect the superradiance behavior. For example, [138] discusses a model in which the QCD axion couples to a hidden-sector photon, and there are hidden-sector fermions which interact with this photon. Such models illustrate that, while the minimal DM models we considered in Figs. 10 and 11 are often still subject to BH spin-down constraints, others may not be.

Beyond the spin-0 particle candidates we considered, superradiance of massive vectors is also of interest. Vector self-interactions are somewhat more complicated than those for scalars, since renormalizable interactions between vectors must take the form of Yang-Mills theory. For abelian theories, "light-by-light" scattering could lead to qualitatively similar dynamics to those discussed here, but has to be investigated in the context of a low cutoff and potential production of the charged particles which give rise to the vector self-interaction. Beyond self-interactions, a simple example of both theoretical and phenomenological interest is a light vector interacting with the SM via a kinetic mixing with the SM photon (though plasma dynamics may make the behavior around astrophysical black holes very complicated). A vector may also have interactions with other hidden sector states—for example, its mass may come from a Higgs mechanism, or it may mediate interactions between hidden sector matter. For the purely gravitational story to hold, such states must be sufficiently heavy, and/or sufficiently weakly coupled [20]. We leave investigations of such scenarios to future work.

Superradiance of spin-2 particles has also been investigated in the literature [22,23]. An issue with such models is that an effective field theory with a spin-2 particle of mass μ, along with the massless graviton (a "bigravity" theory), has a cutoff scale at or below $\Lambda_3 = (M_P \mu^2)^{1/3}$ [139,140]. Here, $M_P \sim \min(M_{\rm pl}, \Lambda)$ is an effective mass scale set by the mass scales $M_{\rm pl}$, which suppresses massless graviton interactions, and Λ, which suppresses massive spin-2 interactions. At the small masses μ we are interested in for BH superradiance, $\Lambda_3 \lesssim 10\ \text{eV}(\frac{\mu}{10^{-12}\ \text{eV}})^{2/3}$ is small compared to energy scales of interest. For example, the energy density in a fully occupied superradiant cloud is $\rho \sim (6\ \text{MeV})^4(\frac{\alpha}{0.2})^5(\frac{\mu}{10^{-12}\ \text{eV}})^2$. Consequently, it is unclear whether there are theories for which reliable calculations can be carried out in the regimes of interest.

Returning to the topic of spin-0 superradiance; as well as exploring the new observational signatures that may arise from self-interactions, our analyses clarify when self-interactions are small enough not to affect the usual gravitational dynamics of superradiance. As illustrated in Figs. 10 and 11, this is important for understanding when constraints and signatures from motivated models, such as the QCD axion or misalignment DM, can be trusted.

As we have demonstrated, adding a simple quartic interaction can dramatically change the dynamics of scalar superradiance. The additional interaction inevitably reduces the efficiency of black hole spindown as well as the strength and timescale of gravitational wave annihilation signals. Nevertheless, the new dynamics can lead to simultaneous population of multiple levels giving rise to gravitational wave transition signals, a narrow range of which may be observable at LIGO/Virgo. Given that the transition signals are at parametrically lower frequencies corresponding to the energy splitting between different levels, signals from scalars around stellar mass black holes generally fall below the LIGO/Virgo sensitivity band in frequency and present new targets for future mid-band detectors.

Perhaps the most novel signature is the emission of particles to infinity: a light, self-coupled axion can extract the energy of rotating black holes and populate our galaxy with axion waves, without the need for a cosmological abundance or a coupling to Standard Model matter. In the presence of such a coupling, these axion waves could be detected in the lab. While current experiments are not yet sensitive to this population of light axions, this mechanism further motivates the development of light axion direct detection experiments, as well as numerical work on self-interactions in superradiance to better characterize the signal from compact, semirelativistic axion clouds.

ACKNOWLEDGMENTS

We thank Asimina Arvanitaki, Savas Dimopoulos, Sergei Dubovsky, Peter Graham, Kurt Hinterbichler,

Junwu Huang, Ken Van Tilburg, and Sylvia Zhu for helpful discussions. We thank Perimeter Institute for warm hospitality during the completion of part of this work. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development & Innovation. M. B. is supported by the James Arthur Postdoctoral Fellowship. M. G. and R. L. are supported in part by the National Science Foundation under Grant No. PHYS-1720397, and the Gordon and Betty Moore Foundation Grant No. GBMF7946. O. S. is supported by the *Fonds de recherche du Québec Nature et Technologies* and by a DARE Fellowship from the Vice Provost for Graduate Education at Stanford University.

APPENDIX A: PARAMETRIC OSCILLATOR ANALYSIS

As discussed in Sec. III C, a useful way to analyse the growth of bound levels is to assume that we have some large-amplitude ψ_c, and to treat this as a parametric forcing in the Gross-Pitaevskii (GP) equation [Eq. (12)], i.e., to solve

$$(i\partial_t + \mathcal{M})\psi = \frac{-3\lambda}{24\mu^2}(\psi_c^2\psi^* + |\psi_c|^2\psi) \tag{A1}$$

(here, $\mathcal{M}$ represents the terms in the nonrelativistic Hamiltonian, including an absorbing term corresponding to the BH horizon). As compared to the forced oscillation analysis in Sec. III C, we ignore back-action for only two of the "legs" in diagrams such as Eq. (19), rather than for three of them.

To simplify our discussion, we will take $\psi_c \propto \psi_{211}$ [so we are interested in processes such as Eq. (19)]. It is helpful to extract the time dependence corresponding to the 211 oscillation, and write $\psi = \Psi e^{-i\tilde{\omega}_c t}$, where $\tilde{\omega}_c \equiv \tilde{\omega}_{211}$ (for simplicity, we will assume that ω_{211} is real, as it is when 211 has reached its saturation value). Then, if we take a harmonic ansatz, $\Psi = Ae^{-i\hat{\omega}t} + Be^{i\hat{\omega}^* t}$, the GP equation

$$(i\partial_t + \tilde{\omega}_c + \mathcal{M})\Psi = \tilde{\lambda}(\Psi_{211}^2\Psi^* + |\Psi_{211}|^2\Psi) \tag{A2}$$

(where $\tilde{\lambda} \equiv -\frac{3\lambda}{24\mu^2}$) implies that

$$(\hat{\omega} + \tilde{\omega}_c + \mathcal{M})A = \tilde{\lambda}(\Psi_{211}^2 B^* + |\Psi_{211}|^2 A) \tag{A3}$$

and

$$(-\hat{\omega}^* + \tilde{\omega}_c + \mathcal{M})B = \tilde{\lambda}(\Psi_{211}^2 A^* + |\Psi_{211}|^2 B) \tag{A4}$$

If we take the complex conjugate of Eq. (A4), then together with Eq. (A3), we have a linear eigenvalue problem that we can solve for $\hat{\omega}$. For $\lambda = 0$, the solutions correspond to usual hydrogenic (quasibound) states.

For non-Hermitian Hamiltonians, the eigenstates are generally nonorthogonal [141]. However, in our case, we can write $\mathcal{M} = \mathcal{M}_R + i\mathcal{M}_I$, and treat $\mathcal{M}_I$ as being diagonal in the basis of $\mathcal{M}_R$ eigenstates (that is, we ignore the detailed dynamics behind the absorption, since this is outside the regime of the nonrelativistic approximation). In this case, the ($\lambda = 0$) quasibound states Ψ_k are orthogonal [141], and we will assume the normalization $\int dV \Psi_k^* \Psi_j = \delta_{jk}$.

To linear order in λ, if we start with the unperturbed solution $A = \Psi_i$, $B = 0$, then we can write the perturbed solution as $A = \Psi_i + \sum_k \alpha_k \Psi_k$, $B = \sum_k \beta_k \Psi_k$ (expanding in the unperturbed basis). Using equations (A3) and (A4),

$$(\hat{\omega}_i - \hat{\omega}_k)\alpha_k = \tilde{\lambda}\int dV \Psi_k^* |\Psi_{211}|^2 \Psi_i \tag{A5}$$

$$(-\hat{\omega}_i^* - \hat{\omega}_k)\beta_k = \tilde{\lambda}\int dV \Psi_k^* \Psi_{211}^2 \Psi_i^* \tag{A6}$$

As well as these perturbations to the wave function, we are interested in finding the perturbation to the frequency $\hat{\omega}$ of the state. Writing $\hat{\omega} = \hat{\omega}_i + \delta\hat{\omega}$, we have

$$(\hat{\omega}_i + \tilde{\omega}_c + \mathcal{M})A = -\delta\hat{\omega}A + \tilde{\lambda}(\Psi_{211}^2 B^* + |\Psi_{211}|^2 A) \tag{A7}$$

If we take A to be normalized so that $\int dV \Psi_i^* A = 1$ even for nonzero λ, then this implies that

$$\delta\hat{\omega} = \tilde{\lambda}\int dV \Psi_i^* (\Psi_{211}^2 B^* + |\Psi_{211}|^2 A) \tag{A8}$$

$$= \tilde{\lambda}\int dV \Psi_i^* |\Psi_{211}|^2 \Psi_i \tag{A9}$$

$$- \tilde{\lambda}^2 \sum_k \frac{1}{\hat{\omega}_i + \hat{\omega}_k^*} \left| \int dV \Psi_k^* \Psi_{211}^2 \Psi_i^* \right|^2 \tag{A10}$$

$$+ \tilde{\lambda}^2 \sum_k \frac{1}{\hat{\omega}_i - \hat{\omega}_k} \left| \int dV \Psi_k^* |\Psi_{211}|^2 \Psi_i^* \right|^2 \tag{A11}$$

The second and fourth lines of this expression give behavior similar to standard perturbation theory. However, the $-1/(\hat{\omega}_i^* + \hat{\omega}_k)$ factor in the second line gives rise to qualitatively different effects. If the Ψ_i mode is decaying, but the Ψ_k mode is damped sufficiently strongly that $\mathrm{Im}(\hat{\omega}_i + \hat{\omega}_k^*) > 0$, then $\mathrm{Im}(\frac{-1}{\hat{\omega}_i + \hat{\omega}_k^*}) > 0$. Consequently, the "mixing" with the Ψ_k mode contributes a *growing* term to the perturbed Ψ_i mode. In our case, the 211 parametric forcing gives the 322 mode a "mixing" with the decaying 100 mode (and the $n00$ modes, etc), contributing a growing term for 322. The perturbations to the 322 wave function correspond to the forced oscillation discussed in Sec. III C. Using Eq. (A11), we obtain the same 322 growth rate as calculated from the forced-oscillation picture.

Similarly, mixing with superradiant (rather than decaying) modes contributes a negative imaginary part to $\delta\hat{\omega}$. This is again as we'd expect from the forced oscillation picture. Including a growing 211 occupation number, as is appropriate when 211 is still superradiant, leads to more complicated expressions. However, since the superradiant growth timescale is always much longer than the oscillation period of ψ_{211}, we can separate these timescales, with 322 growth at a particular time being driven by the 211 amplitude at that time.

This kind of perturbative analysis can be applied in the hydrogenic approximation (at leading order in α), or using numerical wave functions for the bound states, which will be more accurate at higher α. In Fig. 21, we plot the decay rate of the 100 level, relative to its leading-α power-law behavior (the $n00$ levels have very similar behavior). The lower panel of the figure also shows a numerical approximation to the rate of the $211 \times 211 \to 322 \times$ BH process, computed by numerically integrating the forced equation of motion in the Kerr background (for practical reasons, over a restricted range in α). The close correspondence between the behaviors of these two rates illustrates that, for the $211 \times 211 \to 322 \times$ BH process, the most significant high-α corrections come from short-distance effects that affect the flux across the horizon; at long distances from the BH, the forcing term, and the forced oscillation, are not strongly affected (at the α of interest).

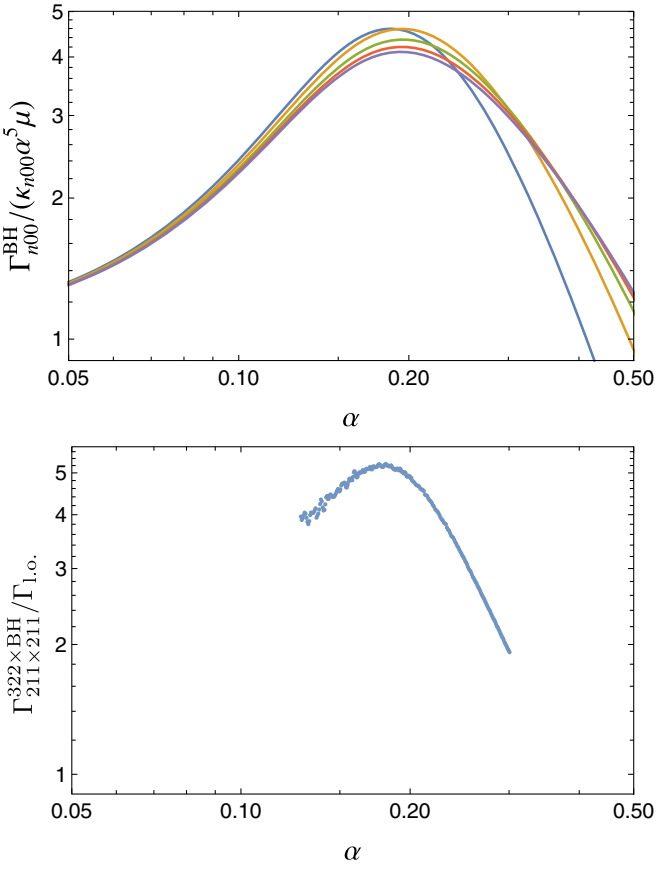


FIG. 21. Top panel: decay rates of the $n00$ hydrogenic levels, for $n = 1$ to 5, relative to their leading-order power-law behavior as a function of α (for a BH with $a_* = 0.9$). Bottom panel: rate of the $211 \times 211 \to 322 \times$ BH process, for a BH with $a_* = 0.9$, relative to its leading power-law behavior $\Gamma_{\rm l.o.}$ as a function of α (see Table V). As discussed in Sec. III C, the deviation of this rate from its leading-order form is mostly driven by the same short-distance effects that modify the $n00$ decay rates.

The parametric forcing analysis above is not specific to black hole superradiance. In the simplest case, if we had two oscillators, with an oscillating coupling between them,

$$\ddot{x} + \omega_0^2 x = f\cos(2\omega_c t)y \tag{A12}$$

$$\ddot{y} + \gamma\dot{y} + \omega_1^2 y = f\cos(2\omega_c t)x \tag{A13}$$

then the same kind of analysis would apply. In the absence of the damping term γ, if $\omega_0 + \omega_1$ is detuned from $2\omega_c$, then the system is not unstable to growth. Introducing γ leads to the exponential growth of x, as per above.

While the above analyses were at the level of classical equations, a similar analysis could be done in terms of quantum master equations. The most important physical difference is that, while the classical ground state is stationary, quantum fluctuations are amplified by the instability, so the ground states evolves into a probability mixture of coherent states. This is precisely analogous to the amplification of quantum fluctuations by superradiance, as discussed in Sec. II.

From Eq. (A6), our perturbative treatment breaks down when

$$\frac{\tilde{\lambda}\int dV \Psi_k^* \Psi_{211}^2 \Psi_i^*}{\hat{\omega}_i^* + \hat{\omega}_k} \gtrsim 1 \tag{A14}$$

In terms of the physical mode frequencies, $\hat{\omega}_i^* + \hat{\omega}_k = \omega_i^* + \omega_k - 2{\rm Re}\omega_{211}$. For generic hydrogenic modes, this is $\mathcal{O}(\alpha^2)\mu$, and in this case, the lhs of Eq. (A14) is parametrically $\sim\alpha^3 \frac{M_{\rm pl}^2}{f^2}\varepsilon_{211}$, similarly to the self-energy corrections [Eq. (H1a)]. As we discuss in Sec. V, the largest value that ε_{211} attains decreases as we decrease f, and the numerical value of this quantity is always small.

In special cases, the source term for the k oscillation can be almost on resonance, and the denominator can become smaller. We discuss a specific example in Sec. III C [Eq. (22)], where it is $\mathcal{O}(\alpha^4)\mu$ for the $211 \times 311 \to 322 \times$ BH process. However, in this case, the source term does not appear to grow large enough for there to be a problem, in most of the parameter space of interest.

It is also possible to treat emission to infinity, e.g., through the $322 \times 322 \to 211 \times \infty$ process discussed in Sec. III B, in terms of a parametric forcing, with loss to infinity acting as like a damping term.

APPENDIX B: PERTURBATIVE CALCULATIONS OF FREQUENCY SHIFTS AND RATES

In this Appendix, we will provide more detailed derivations of the leading-α rates for quartic self-interaction processes involving hydrogenic levels.

Up to corrections from self-gravity, the system obeys the classical equation of motion

$$(D^2-\mu^2)\varphi = -\frac{\lambda}{6}\varphi^3, \tag{B1}$$

where $D^2 = D_\nu D^\nu$ and D_ν is the covariant derivative of the Kerr geometry. Expending D^2 to first order in r_g/r, this becomes

$$\left(\frac{\partial^2}{\partial t^2} - \vec{\nabla}^2 + \mu^2\right)\varphi - \frac{2\alpha}{r}(\mu + \hat{K}) = \frac{\lambda}{6}\varphi^3. \tag{B2}$$

The term

$$\hat{K} = \frac{1}{r}\frac{\partial}{\partial r} - 2\left(\mu^2 + \frac{\partial^2}{\partial t^2}\right) - \frac{\hat{L}^2}{r^2} \tag{B3}$$

is parametrically suppressed relative to μ for nonrelativistic components of φ, and we drop $\hat{K}$ except for calculations of relativistic emissions. Here $\hat{L}^2$ denotes the total angular momentum operator: $\hat{L}^2 Y_l^m = l(l+1)Y_l^m$ for the Laplace spherical harmonics $Y_l^m(\theta,\varphi)$.

We seek a perturbative solution in the self-interaction parameter λ,

$$\varphi = \varphi^{(0)} + \lambda\varphi^{(1)} + \ldots \tag{B4}$$

At zeroth order,

$$\left(\frac{\partial^2}{\partial t^2} - \vec{\nabla}^2 + \mu^2 - \frac{2\alpha\mu}{r}\right)\varphi^{(0)} = 0. \tag{B5}$$

This equation admits nonrelativistic (quasi)bound states with hydrogenic waveforms and energies which we identify with the superradiant cloud:

$$\varphi^{(0)} \equiv \sum_{nlm}\varphi^{(0)}_{nlm} = \sum_{nlm}\sqrt{\frac{N_{nlm}}{2\mu}}\, e^{-i\omega_{nlm}t}\psi_{nlm} + \text{c.c.}, \tag{B6}$$

up to phases, where ψ_{nlm} are the normalized hydrogenic wave functions $\int |\psi_{nlm}|^2 d^3\vec{r} = 1$.

To avoid secular terms at the next perturbative order, we must also introduce a perturbation series for the normal frequencies:

$$\omega_{nlm} = \omega^{(0)}_{nlm} + \lambda\omega^{(1)}_{nlm} + \ldots, \tag{B7}$$

where

$$\omega^{(0)}_{nlm} = \omega_n + i\Gamma^{\text{SR}}_{nlm}, \tag{B8}$$

$$\omega_n \approx \mu\left(1 - \frac{\alpha^2}{2n^2}\right), \tag{B9}$$

and Γ^{SR}_{nlm} is the superradiance rate. We call the energy corrections $\Delta\omega_{nlm} \equiv \lambda\omega^{(1)}_{nlm}$

At first order in perturbation theory, this gives a driven massive Coulomb wave equation,

$$\left(\frac{\partial^2}{\partial t^2} - \vec{\nabla}^2 + \mu^2 - \frac{2\alpha\mu}{r}\right)\varphi^{(1)} = \frac{1}{6}(\varphi^{(0)})^3 + \sum_{nlm} 2\mu\omega^{(1)}_{nlm}\varphi^{(0)}_{nlm}. \tag{B10}$$

Plugging (B6) into (B10), and expanding the driving term as a sum of harmonic driving terms gives

$$(\varphi^{(0)})^3 \sim \sum_\Omega f(\vec{r},\Omega)e^{-i\Omega t} + \text{c.c.}, \quad \Omega > 0. \tag{B11}$$

Since $\varphi^{(0)} \sim a_0^{-3/2}\mu^{-1/2} \sim \alpha^{3/2}\mu$, the source $f(\vec{r}) \sim (\varphi^{(0)})^3$ scales as $\sim\alpha^{9/2}\mu^3$. The physical intuition behind that scaling is that a cloud with larger α has a smaller characteristic size a_0 and therefore larger densities, enhancing the rate of many-body processes.

The physical nature of the process associated to each summand depends on the value of Ω:

(1) $\Omega - \mu > 0$ corresponds to *free* radiation emitted in the continuum and travelling to infinity either with nonrelativistic or relativistic velocities,
(2) $\Omega - \mu < 0$ and $\Omega \neq \omega_n$ for all n is off-resonant driving of discrete *bound* modes, i.e., the production of off-shell particles trapped in the gravitational well.
(3) $\Omega = \omega_n$ for some n is resonant driving, which either corresponds to resonant (on-shell) production of particles inside the cloud, or to a correction to the frequencies (one-particle energies) and waveforms (one-particle states) of the zeroth-order normal modes.

For clarity, we focus on the source

$$\varphi^{(0)}(\vec{r},t) = \sqrt{\frac{N_{211}}{2\mu}}e^{-i\omega_2 t}\psi_{211}(\vec{r}) + \sqrt{\frac{N_{322}}{2\mu}}e^{-i\omega_3 t}\psi_{322}(\vec{r}) + \text{c.c.} \tag{B12}$$

for the remainder of this Appendix. The source (B12) represents the only two levels of the cloud relevant to the

intra-cloud dynamics at small enough α, as argued in Sec. IV C and Appendix C 2 a.

1. Frequency corrections

The source term includes components at the frequency ω_2 of the 211 bound state,

$$\frac{1}{6}(\varphi^{(0)})^3 \supset \frac{e^{-i\omega_2 t}}{(2\mu)^{3/2}}\left(\frac{1}{2}N_{211}^{3/2}(\psi_{211})^2\psi_{211}^* + N_{322}\sqrt{N_{211}}\psi_{322}\psi_{322}^*\psi_{211}\right). \tag{B13}$$

This source contains components in resonance with the normal mode $\psi_{211}e^{-i\omega_2 t}$ which would drive $\varphi^{(1)}$ to very large amplitudes, preventing a perturbative treatment. The frequency correction $\omega_{211}^{(1)}$ is therefore determined by demanding that those resonant components be exactly cancelled:

$$\omega_{211}^{(1)} = -\frac{1}{4\mu^2}\int\left(\frac{N_{211}}{2}|\psi_{211}|^4 + N_{322}|\psi_{322}|^2|\psi_{211}|^2\right)d^3\vec{r}. \tag{B14}$$

The two terms in Eq. (B14) correspond to self-energy corrections of the level 211 from its interaction with itself and with 322, respectively. The integral can be computed analytically by using the explicit form of the hydrogenic waveforms ψ_{211} and ψ_{322}.

Since bound state wave functions scale as $\psi_{nlm} \propto 1/a_0^{-3/2} \propto (\alpha\mu)^{3/2}$ and only depend on $\vec{r}$ through r/a_0, the frequency correction scales with α as $\omega_{nlm}^{(1)} \sim \alpha^3\mu$. A denser cloud gives larger frequency corrections.

We calculated the integral of Eq. (B14) and the equivalent for $\omega_{322}^{(1)}$ and we found the corrections:

$$\Delta\omega_{211} \simeq -\lambda\alpha^3\mu(1.2\times10^{-4}N_{211} + 3.5\times10^{-5}N_{322}) = -\alpha^5\mu\left(\frac{M_{\rm pl}}{f}\right)^2(1.2\times10^{-4}\varepsilon_{211} + 3.5\times10^{-5}\varepsilon_{322}) \tag{B15}$$

$$\Delta\omega_{322} \simeq -\lambda\alpha^3\mu(3.5\times10^{-5}N_{211} + 1.4\times10^{-5}N_{322}) = -\alpha^5\mu\left(\frac{M_{\rm pl}}{f}\right)^2(3.5\times10^{-5}\varepsilon_{211} + 1.4\times10^{-5}\varepsilon_{322}) \tag{B16}$$

2. $l=0$ damped-driven oscillation

When $\Omega-\mu<\mu$ and $\Omega\neq\omega_n$ for any n, the source generates a forced bound oscillation which is damped by the BH. For example, when the cloud consists of particles in the 211 and 322 levels (B12), the frequency of the forced oscillation is $\omega_{\rm ind} = 2\omega_2 - \omega_3 = \mu(1-7\alpha^2/36) < \mu$, so the oscillation is bound.

The bound state $\varphi^{(1)} \supset e^{-i\omega_{\rm ind}t}\Psi^{(1)}(\vec{r}) + \text{c.c.}$ satisfies the time-independent equation for the complex field $\Psi^{(1)}$,

$$\left(k_{\rm ind}^2 - \vec{\nabla}^2 - \frac{2\alpha\mu}{r}\right)\Psi^{(1)}(\vec{r}) = \frac{1}{2}\frac{N_{211}\sqrt{N_{322}}}{(2\mu)^{3/2}}(\psi_{211})^2\psi_{322}^*, \tag{B17}$$

where $k_{\rm ind}^2 = \mu^2 - \omega_{\rm ind}^2 \approx (7/18)\alpha^2\mu^2$.

We expand $\Psi^{(1)}(\vec{r})$ in the complete basis of the hydrogenic differential operator $-\nabla^2 - 2\alpha\mu/r$,

$$\Psi^{(1)}(\vec{r}) = \sum_{nlm}c_{nlm}\psi_{nlm} + \sum_{lm}\int dk c(k)\psi_{klm}, \tag{B18}$$

where the eigenfunctions nlm of the discrete spectrum satisfy

$$\left(-\nabla^2 - \frac{2\alpha\mu}{r}\right)\psi_{nlm} = -k_n^2\psi_{nlm}, \qquad k_n^2 = \frac{\alpha^2\mu^2}{n^2}, \tag{B19}$$

with n a positive integer, and eigenfunctions klm of the continuous spectrum obey

$$\left(-\nabla^2 - \frac{2\alpha\mu}{r}\right)\psi_{klm} = k^2\psi_{klm}, \qquad \frac{k}{\alpha\mu}\in(0,+\infty). \tag{B20}$$

Moreover, the eigenfunctions obey orthonormality conditions:

$$\int d^3\vec{r}\psi_{nlm}^*\psi_{n'l'm'} = \delta_{n,n'}\delta_{m,m'}\delta_{l,l'}, \tag{B21a}$$

$$\int d^3\vec{r}\psi_{klm}^*\psi_{k'l'm'} = 2\pi\delta(k'-k)\delta_{m,m'}\delta_{l,l'}, \tag{B21b}$$

$$\int d^3\vec{r}\psi_{klm}^*\psi_{nl'm'} = 0. \tag{B21c}$$

Explicitly, the states of the discrete spectrum are the usual bound hydrogenic wave functions,

$$\psi_{nlm}(r,\theta,\varphi) = R_{nl}(r)Y_l^m(\theta,\varphi), \tag{B22}$$

with the radial part

$$R_{nl}(r) = \sqrt{\left(\frac{2}{na_0}\right)^3\frac{(n-l-1)!}{2n(n+l)!}} \times \exp\left[\frac{-r}{na_0}\right]\left(\frac{2r}{na_0}\right)^l L_{n-l-1}^{2l+1}\left[\frac{2r}{na_0}\right], \tag{B23}$$

where $L^{2l+1}_{n-l-1}(x)$ is the generalized Laguerre polynomial of degree $n-l-1$.

The states of the continuous spectrum are stationary Coulomb waves [142],[15]

$$\psi_{klm}(r,\theta,\varphi) = R_{kl}(r)Y^m_l(\theta,\phi) \tag{B24}$$

with the radial part

$$R_{kl}(r) = \frac{2ke^{\pi/(2ka_0)}|\Gamma(l+1-i/(ka_0))|}{(2l+1)!} \times (2kr)^l e^{-ikr}\,{}_1F_1(i/(ka_0)+l+1,2l+2,2ikr), \tag{B25}$$

where ${}_1F_1$ is the confluent hypergeometric function of the first kind.

To obtain the coefficients c_{nlm}, we put (B18) in (B17) and integrate both sides against $\psi^*_{n'l'm'}$. We can then use the Hermiticity of $(-\nabla^2 - 2\alpha\mu/r)$ (which in this case amounts to integrating by parts, so that $-\nabla^2 - 2\alpha\mu/r$ acts on $\psi^*_{n'l'm'}$), along with (B19) and (B21) to find

$$c_{nlm} = \frac{1}{k^2_{\rm ind} - k^2_n} \times \int \frac{1}{2}\frac{N_{211}\sqrt{N_{322}}}{(2\mu)^{3/2}}(\psi_{211})^2\psi^*_{322}\psi^*_{nlm}d^3\vec{r}. \tag{B26}$$

Similarly, the values of the transform $c(k)$ are obtained by integrating both sides of (B17) against $\psi_{k'l'm'}$. The analogue procedure then yields

$$c(k) = \frac{1}{2\pi}\frac{1}{k^2_{\rm ind}+k^2} \times \int \frac{1}{2}\frac{N_{211}\sqrt{N_{322}}}{(2\mu)^{3/2}}(\psi_{211})^2\psi^*_{322}\psi^*_{klm}d^3\vec{r}. \tag{B27}$$

It is appropriate to do these integrals in units of the Bohr radius $a_0 = (\alpha\mu)^{-1}$ to reconstitute the dependence on α. The prefactors of $k_{\rm ind}$, k_n and k are naturally in units of a_0^{-1}, while bound state wave functions are in units of $a_0^{-3/2}$ and continuum wave functions are in units of a_0^{-1}. The c_{nlm}'s then have dimension $a_0^{-1}\mu^{-3/2}$ and $c(k)$ has units of $a_0^{-1/2}\mu^{3/2}$. The amplitude of the induced oscillation $\varphi^{(1)}$ therefore has units of $a_0^{-5/2}\mu^{-3/2} = \alpha^{5/2}\mu$.

These overlap integrals are nonvanishing for $l = 0, 2, 4$ and $m = 0$. For $l > 0$ however, the angular momentum barrier suppresses the field amplitude at the horizon, and therefore the corresponding rates of absorption are smaller. This in turn leads to a smaller induced growth rate, as discussed previously. We therefore focus on $l = m = 0$ and ignore the $l = 2, 4$ terms.

For $l = 0$ states, the power absorbed at the horizon in terms of complex field $\Psi^{(1)}$ goes as the square of the norm at the origin:

$$P_{\rm abs} \approx 4\alpha^2\left(1+\sqrt{1-a_*^2}\right)\lambda^2|\Psi(0)^{(1)}|^2. \tag{B28}$$

In terms of particles in the cloud carrying energy $\approx\mu$, this contributes

$$\dot{N}_{\rm c} \supset -\Gamma_{\rm damp}N^2_{211}N_{322}, \tag{B29}$$

with

$$\Gamma_{\rm damp} = \frac{P_{\rm abs}}{\mu N^2_{211}N_{322}}. \tag{B30}$$

3. Nonrelativistic emission

Generally, the source term $(\varphi^{(0)})^3$ will generate some driving terms oscillating at the frequency $\omega_r^{\rm NRE} = \omega_n + \omega_{n'} - \omega_{n''}$. When $\omega_r^{\rm NRE} > \mu$, the driven oscillation is free. These free emissions are nonrelativistic because $\omega_{\rm r}^{\rm NRE} \approx \mu + \mathcal{O}(\alpha^2)$ for the constituents of the superradiant cloud. The superscript NRE ("nonrelativistic emissions") will be suppressed will be suppressed for the remainder of this section.

Generically, we seek to solve

$$\left(\frac{\partial^2}{\partial t^2} - \vec{\nabla}^2 + \mu^2 - \frac{2\alpha\mu}{r}\right)\varphi_r = e^{-i\omega_{\rm r}t}f(\vec{r}) + \text{c.c.}, \tag{B31}$$

where $e^{-i\omega_{\rm r}t}f(\vec{r})$ is a localized source of radiation with harmonic time-dependence, and $\varphi_r \subset \varphi^{(1)}$ is the radiation part of the field. The time-averaged differential power per solid angle that such a source emits in the radiation zone at infinity in the direction $(\theta_{\vec{k}}, \varphi_{\vec{k}})$ is

$$\frac{d\langle P\rangle}{d\Omega}(\theta_{\vec{k}},\varphi_{\vec{k}}) = 2\frac{\omega_r|\vec{k}|}{(4\pi)^2}\lambda^2|\tilde{f}(\vec{k})|^2, \tag{B32}$$

where $\vec{k} = (\sqrt{(\omega_r)^2-\mu^2})\hat{r}$ is the momentum at spatial infinity, $\hat{r}$ is a radial unit vector pointing in the direction (θ_k,φ_k) and $\tilde{f}(\vec{k})$ is the "Coulomb" transform

$$\tilde{f}(\vec{k}) = \sum_{lm}Y^m_l(\theta_k,\varphi_k)\int d^3\vec{r}(4\pi)(-i)^l f(\vec{r})\frac{\psi^*_{klm}(\vec{r})}{2k}. \tag{B33}$$

This is analogous to the usual Fourier transform that one would compute for the emission rate in flat spacetime, with

[15]this is appropriate in the hydrogenic approximation, where we take into account the Newtonian $1/r$ gravitational potential. Corrections from the full Kerr potential will be higher order in α.

the regular spherical Bessel functions having been replaced by the appropriate regular Coulomb waves.

For nonrelativistic emissions, $k \sim \mathcal{O}(a_0^{-1})$. It was noted earlier that $f(\vec{r}) \sim \alpha^{9/2}\mu^3$. Furthermore, since it is a product of hydrogenic wave functions, $f(\vec{r})$ depends on $\vec{r}$ only through the combination $\vec{r}/a_0$. On the other hand, $\psi_{klm}(\vec{r})/2k$ is dimensionless and depends on $\vec{r}$ only through the combination $kr \sim r/a_0$, for nonrelativistic k. Therefore, all the dependence of (B33) on α can be extracted by evaluating the intgeral in units of the Bohr radius $a_0 = (\alpha\mu)^{-1}$. Thus $\tilde{f}(\vec{k}) \sim \alpha^{3/2}$ and $d\langle P\rangle/d\Omega \sim \lambda^2\alpha^4\mu^2$.

The total radiated power is determined by integrating (B32) over solid angles:

$$P_r^{\mathrm{NRE}} = \int d\Omega \frac{d\langle P\rangle}{d\Omega} \tag{B34}$$

In terms of particles in the cloud, and particles radiated to infinity with energy $\omega_r \approx \mu$, we have

$$\dot{N}_{\mathrm{c}} \supset -\Gamma_{\mathrm{r}}^{\mathrm{NRE}} N_{nlm} N_{n'l'm'} N_{n"l"m"}, \tag{B35}$$

with the rate

$$\Gamma_r^{\mathrm{NRE}} = \frac{P_r^{\mathrm{NRE}}}{\mu N_{nlm} N_{n'l'm'} N_{n"l"m"}}. \tag{B36}$$

A particularly important process is $211\times211 \to 322\times\infty$. This is sourced by

$$\frac{1}{6}(\varphi^{(0)})^3 \supset e^{-i(2\omega_3-\omega_2)t} \times \frac{1}{2}\frac{N_{322}\sqrt{N_{211}}}{(2\mu)^{3/2}}(\psi_{322})^2\psi_{211}^* + \mathrm{c.c.} \tag{B37}$$

By substituting

$$f(\vec{r}) \to \frac{1}{2}\frac{N_{322}\sqrt{N_{211}}}{(2\mu)^{3/2}}(\psi_{322})^2\psi_{211}^* \tag{B38}$$

in the above, we obtain the rate in Table II.

4. Relativistic emission

The source term $(\varphi^{(0)})^3$ will also contain terms oscillating at the frequency $\omega_r^{RE} = \omega_n + \omega_{n''} + \omega_{n'''}$. When $\omega_r^{RE} > \mu$, the driven oscillation is free. These free emissions are relativistic because $\omega_r \approx 3\mu + \mathcal{O}(\alpha^2)$ for the constituents of the superradiant cloud. Cubic self-interactions would also generate relativistic emissions through $(\varphi^{(0)})^2$ in the equations of motion. In this case $\omega_r^{\mathrm{RE}} \approx 2\mu + \alpha^2$.

For the remainder of this section, the superscript RE ("relativistic emissions") will be suppressed. As is the case for nonrelativistic emissions, the radiated power is controlled by the integral (B33) which projects the source onto the Coulomb scattering state with outgoing momentum $\vec{k}$. The source $f(\vec{r})$ is a product of hydrogenic wave functions,

$$f(\vec{r}) \simeq \frac{-\lambda/6}{(2\mu)^{3/2}}N_{211}^{3/2}\psi_{211}^3 = \frac{\lambda}{768\pi\sqrt{70}}\alpha^{3/2}a_0^{-3}(r/a_0)^3 e^{-3r/(2a_0)}Y_{33}N_{211}^{3/2} \tag{B39}$$

For nonrelativistic emission, $k \sim a_0^{-1}$ so we need to use the full form of the Coulomb scattering state. In contrast, for relativistic emission, $k \sim \mu \sim \alpha^{-1}a_0^{-1}$, so $ka_0 \sim \alpha^{-1}$ is a large parameter. As a result, we can expand the radial part of the Coulomb wave function around its flat-space, spherical Bessel function form.

It turns out that the contributions to $\tilde{f}(\vec{k})$ from the spherical Bessel function, and from the leading-α correction, are at the same order in α. This effectively occurs due to the contribution from the spherical Bessel function suffering a "cancellation," making it higher-order in α than a naive guess based on the behavior of $f(\vec{r})$ near the origin would have indicated. The integral against the leading-α correction term does not suffer this kind of cancellation, making the contributions from both of the same order. This is why our result for the emitted power [Eq. (15)] has the same α dependence as that derived in [6] using a flat-space approximation, but has a larger constant factor ([6] also treats emission as light-like, taking $\omega_r^2 = k^2$ rather than $\omega^2 = k^2 + \mu^2$).

Higher-order corrections, and effects from working in the full Kerr metric instead of just a $1/r$ potential, all contribute to the emitted power at higher order in α.

APPENDIX C: MIXING BEYOND 211 AND 322

1. Selection rules for mixing with damped states

As explained in Appendix A, in the presence of a quartic self-coupling λ, one can view a background SR cloud $\varphi_{\mathrm{c}}(\vec{r},t) \sim e^{-i\mu t}\psi_{211}(t) + e^{-i\mu t}\psi_{322}(t) + \mathrm{c.c}$ as providing a time-dependent mixing potential $V_{\mathrm{mixing}} \sim \lambda e^{-i2\mu t}(\psi_{211}(t)^2 + \psi_{211}(t)\psi_{322}(t) + \psi_{322}(t)^2)$ between states. In particular, if the mixing matrix element $\langle\psi_{n'l'm'}|V_{\mathrm{mixing}}|\psi_{nlm}^*\rangle$ between a superradiant state ψ_{nlm} and a *decaying* state $\psi_{n'l'm'}$ is nonvanishing, then a forced oscillation $\propto \psi_{n'l'm'}$ is sustained and a growth instability is induced for ψ_{nlm}.

We are therefore interested in the selection rules when $V_{\mathrm{mixing}} \sim \psi_{211}^2 \sim Y_2^2$, $V_{\mathrm{mixing}} \sim \psi_{211}\psi_{322} \sim Y_3^3$, and $V_{\mathrm{mixing}} \sim \psi_{322}^2 \sim Y_4^4$. In each case, $V_{\mathrm{mixing}} \sim Y_{l''}^{m''}$ can be viewed as an element of an irreducible tensor operator representation of the rotation group with angular momentum numbers

(l'', m'').[16] Considering further that, $\psi^*_{nlm} \propto Y_l^{-m}$, then by the Wigner-Eckart theorem $\langle \psi_{n'l'm'} | V_{\rm mixing} | \psi^*_{nlm} \rangle \propto (l, l'', -m, m'' | l' m')$, where $(j_1, j_2, m_1, m_2 | J, M) \equiv \langle j_1, j_2, m_1, m_2 | j_1, j_2, J, M \rangle$ is the Clebsch-Gordon (CG) coefficient for the addition of two irreducible angular momentum representations j_1 and j_2. Furthermore, since the parity of a spherical harmonic Y_l^m is $(-1)^l$, inserting parity transformations inside the matrix element yields $\langle \psi_{n'l'm'} | Y_{l''}^{m''} | \psi^*_{nlm} \rangle = (-1)^{l+l'+l''} \langle \psi_{n'l'm'} | Y_{l''}^{m''} | \psi^*_{nlm} \rangle$.

From this we get the selection rules for an induced growth instability to develop:

(1) Mixing with a damped state: $m' \leq 0$,
(2) CG coefficient: $m'' = m' + m$,
(3) CG coefficient: $|l - l''| \leq l' \leq l + l''$,
(4) Invariance under parity: $l + l' + l'' =$ even.

The first rule assumes that the spin in the BH is such that $m \geq 1$ states are SR.

2. Dependence of rates on the quantum numbers

a. Dependence of rates on overtone number n

The sources components ψ_{211} and ψ_{322} are peaked within a few Bohr radii, while hydrogenic wave functions in general are peaked further and further away from the origin as the quantum numbers are taken to be larger and larger. Thus, the interaction of a level $n\ell m$ with a combination of 211 and 322 will depend on the behavior of R_{nl} near the a_0:

$$R_{nl}(r \sim a_0) \sim \left(\frac{2}{(n_r + l + 1)a_0}\right)^{3/2} \left(\frac{1}{2(n_r + l + 1)}\right)^{1/2} \times \left(\frac{(n_r + 2l + 1)!}{n_r!}\right)^{1/2} \frac{1}{(2l+1)!} \left(\frac{2r}{(n_r + l + 1)a_0}\right)^l, \tag{C1}$$

where $n_r = n - l - 1$ is the radial quantum number. If $n_r \to \infty$, while l is held fixed,

$$R_{nl}(r \sim a_0) \sim \left(\frac{1}{n_r a_0}\right)^{3/2} \sim \left(\frac{1}{n a_0}\right)^{3/2}. \tag{C2}$$

Thus, any overlap integral with R_{nl} decreases as $\sim n_r^{-3/2} \sim n^{-3/2}$. This is simply saying that as n_r is taken larger, the characteristic volume of the driving wave function ψ_{nlm} gets larger as $\sim (n a_0)^3$, and so the driving is uniformly diluted by that same factor. A forced oscillation with a ψ_{nl} component as a source term therefore suffers the same suppression.

Rates (whether emission rates or rates of absorption into the BH) depend on the square of the forced oscillation and therefore behave as $\propto n^{-3}$ in the limit of large n. This means that ratios of emissions and absorption processes become independent of n.

The discussion in Sec. IV C relied on the behavior of various ratios of rates at large n. To assess how fast the relevant ratios converge to the expected scaling in n, we plot them for the first $200n$ (Figs. 22, 23, 24, 25, 26).

b. Mixing with $l' = 0$ damped states

The analysis of levels that can grow from 211 and 322 mixing with an $l' = 0$ forced oscillation is done in the main text (IV C).

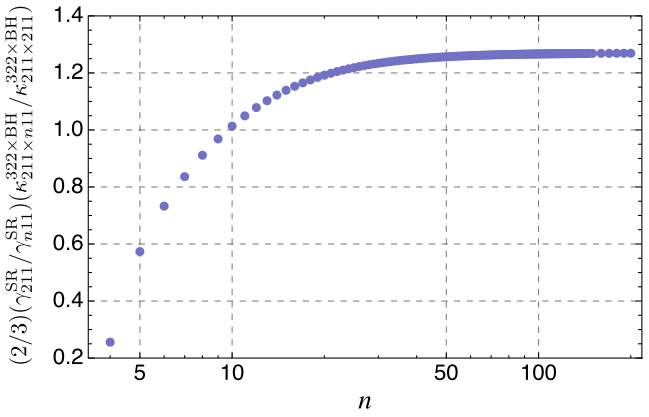


FIG. 22. Behavior of the first term in the ratio in (65) as $n \to \infty$. As discussed in the paragraph below (65), and as expected from C 2 a, the ratio rapidly becomes independent of n and is >1 for $n \gtrsim 10$.

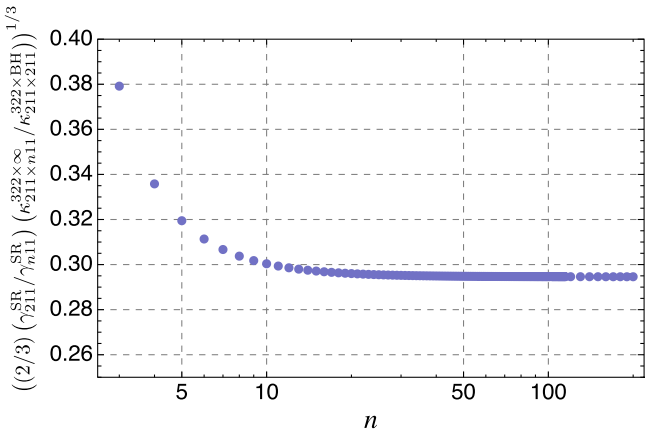


FIG. 23. Behavior of the ratio in (66) as $n \to \infty$. As expected from C 2 a, the ratio rapidly becomes independent of n.

[16] Strictly speaking, since the Kerr metric breaks spherical symmetry, l is not a good quantum number (though m is, since we still have axial symmetry). However, since the metric terms that break spherical symmetry are suppressed at large r/r_g, they lead to effects that are suppressed by more powers of α in the hydrogenic limit.

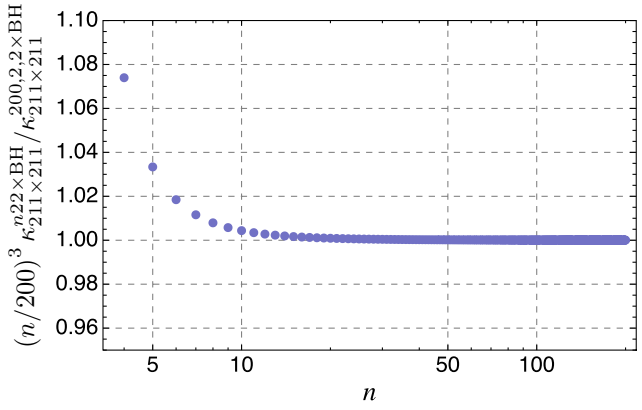


FIG. 24. Behavior of the growth ratio $211\times211\to n22\times\mathrm{BH}$ normalized to its value at $n=200$. As stated in (69), the ratio scales as n^{-3}.

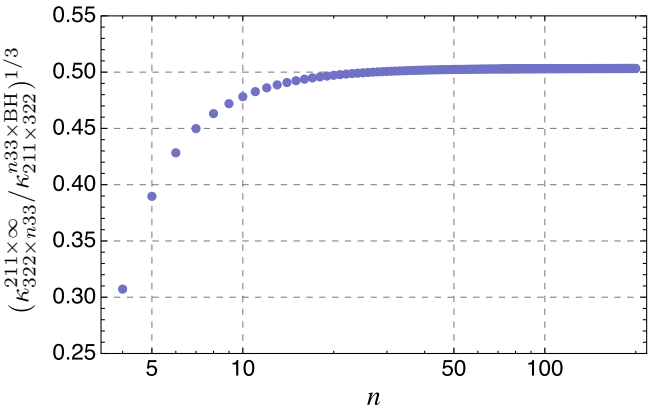


FIG. 25. Behavior of the ratio in (72) as $n\to\infty$. As expected from C 2 a, the ratio rapidly becomes independent of n.

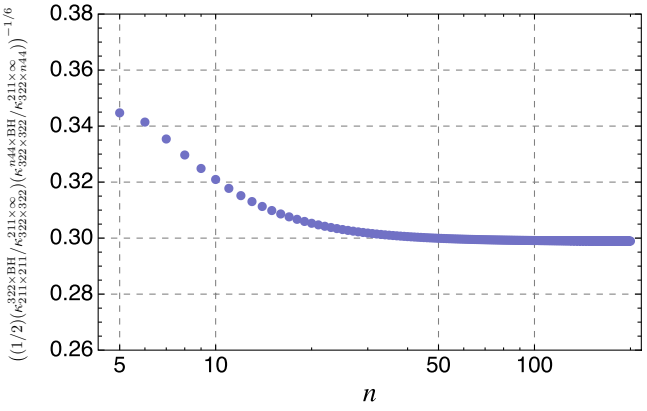


FIG. 26. Behavior of the ratio in (74). As expected from C 2 a, the ratio rapidly becomes independent of n.

c. Mixing with $l'>0$ damped states

We give an exhaustive list of the possible processes involving mixing with $l'=1$ and $l'=2$ damped states.

For $l'=1$,

$$\gamma^{655\times\mathrm{BH}(1,-1)}_{322\times322}(\varepsilon^{\mathrm{eq}}_{322})^2\sim10^5\left(\frac{\alpha}{0.3}\right)^{15}\left(\frac{M_\odot}{M}\right)\ \mathrm{Myr}^{-1},\tag{C3a}$$

$$\gamma^{654\times\mathrm{BH}(1,0)}_{322\times322}(\varepsilon^{\mathrm{eq}}_{322})^2\sim10^4\left(\frac{\alpha}{0.3}\right)^{16}\left(\frac{M_\odot}{M}\right)\ \mathrm{Myr}^{-1},\tag{C3b}$$

$$\gamma^{543\times\mathrm{BH}(1,0)}_{211\times322}\varepsilon^{\mathrm{eq}}_{211}\varepsilon^{\mathrm{eq}}_{322}\sim10^6\left(\frac{\alpha}{0.3}\right)^{13}\left(\frac{M_\odot}{M}\right)\ \mathrm{Myr}^{-1},\tag{C3c}$$

$$\gamma^{432\times\mathrm{BH}(1,0)}_{211\times211}(\varepsilon^{\mathrm{eq}}_{211})^2\sim10^7\left(\frac{\alpha}{0.3}\right)^{10}\left(\frac{M_\odot}{M}\right)\ \mathrm{Myr}^{-1}.\tag{C3d}$$

For $l'=2$,

$$\gamma^{766\times\mathrm{BH}(2,-2)}_{322\times322}(\varepsilon^{\mathrm{eq}}_{322})^2\sim10^{-2}\left(\frac{\alpha}{0.3}\right)^{19}\left(\frac{M_\odot}{M}\right)\ \mathrm{Myr}^{-1},\tag{C4a}$$

$$\gamma^{765\times\mathrm{BH}(2,-1)}_{322\times322}(\varepsilon^{\mathrm{eq}}_{322})^2\sim10^{-3}\left(\frac{\alpha}{0.3}\right)^{19}\left(\frac{M_\odot}{M}\right)\ \mathrm{Myr}^{-1},\tag{C4b}$$

$$\gamma^{764\times\mathrm{BH}(2,0)}_{322\times322}(\varepsilon^{\mathrm{eq}}_{322})^2\sim10^{-4}\left(\frac{\alpha}{0.3}\right)^{20}\left(\frac{M_\odot}{M}\right)\ \mathrm{Myr}^{-1},\tag{C4c}$$

$$\gamma^{653\times\mathrm{BH}(2,0)}_{211\times322}\varepsilon^{\mathrm{eq}}_{322}\varepsilon^{\mathrm{eq}}_{211}\sim10^{-3}\left(\frac{\alpha}{0.3}\right)^{17}\left(\frac{M_\odot}{M}\right)\ \mathrm{Myr}^{-1},\tag{C4d}$$

$$\gamma^{542\times\mathrm{BH}(2,0)}_{211\times211}(\varepsilon^{\mathrm{eq}}_{211})^2\sim10^{-4}\left(\frac{\alpha}{0.3}\right)^{14}\left(\frac{M_\odot}{M}\right)\ \mathrm{Myr}^{-1}.\tag{C4e}$$

Clearly, rates for processes involving $l'\geq2$ are too small to be relevant on astrophysical timescales. Rates from mixing with $l'=1$ states however can become quite large for $\alpha=\mathcal{O}(0.1)$, but, similarly to processes with $l'=0$, they should be compared to depletion processes of the form $nlm\times322\to211\times\infty$.

First,

$$\dot{\varepsilon}_{655}=\gamma^{655\times\mathrm{BH}(1,-1)}_{322\times322}\left(1-\frac{\gamma^{211\times\infty}_{655\times322}}{\gamma^{655\times\mathrm{BH}(1,-1)}_{322\times322}}\frac{\varepsilon_{211}}{\varepsilon_{322}}\right)\varepsilon^2_{322}\varepsilon_{655}.\tag{C5}$$

The depletion term dominates as long as

$$\alpha\tilde{r}_+^{1/9}\lesssim\left(\frac{\kappa^{211\times\infty}_{655\times322}}{\kappa^{655\times\mathrm{BH}(1,-1)}_{322\times322}}\frac{2\kappa^{211\times\infty}_{322\times322}}{\kappa^{322\times\mathrm{BH}(0,0)}_{211\times211}}\right)^{1/9}\approx0.7.\tag{C6}$$

Next,

$$\dot{\varepsilon}_{654}=\gamma^{654\times\mathrm{BH}(1,0)}_{322\times322}\left(1-\frac{\gamma^{211\times\infty}_{654\times322}}{\gamma^{654\times\mathrm{BH}(1,0)}_{322\times322}}\frac{\varepsilon_{211}}{\varepsilon_{322}}\right)\varepsilon^2_{322}\varepsilon_{654}.\tag{C7}$$

The depletion term dominates as long as

$$\alpha\tilde{r}_+^{1/10} \lesssim \left(\frac{\kappa^{211\times\infty}_{654\times322}}{\kappa^{654\times\mathrm{BH}(1,0)}_{322\times322}}\frac{2\kappa^{211\times\infty}_{322\times322}}{\kappa^{322\times\mathrm{BH}(0,0)}_{211\times211}}\right)^{1/10} \approx 0.7. \quad \text{(C8)}$$

Next,

$$\dot{\varepsilon}_{543} = \gamma^{543\times\mathrm{BH}(1,0)}_{211\times322}\left(1-\frac{\gamma^{211\times\infty}_{543\times322}}{\gamma^{543\times\mathrm{BH}(1,-1)}_{211\times322}}\right)\varepsilon_{322}\varepsilon_{211}\varepsilon_{543}. \quad \text{(C9)}$$

The depletion term dominates as long as

$$\alpha\tilde{r}_+^{1/7} \lesssim \left(\frac{\kappa^{211\times\infty}_{543\times322}}{\kappa^{543\times\mathrm{BH}(1,0)}_{211\times322}}\right)^{1/7} \approx 1. \quad \text{(C10)}$$

Finally,

$$\dot{\varepsilon}_{432} = \gamma^{432\times\mathrm{BH}(1,0)}_{211\times211}\left(1-\frac{\gamma^{211\times\infty}_{432\times322}}{\gamma^{432\times\mathrm{BH}(1,0)}_{211\times211}}\frac{\varepsilon_{322}}{\varepsilon_{211}}\right)\varepsilon_{322}\varepsilon_{211}\varepsilon_{432}. \quad \text{(C11)}$$

The depletion term dominates as long as

$$\alpha \lesssim \left(\frac{\kappa^{211\times\infty}_{432\times322}}{\kappa^{432\times\mathrm{BH}(1,0)}_{211\times211}}\frac{1}{2}\frac{\kappa^{322\times\mathrm{BH}(0,0)}_{211\times211}}{\kappa^{211\times\infty}_{322\times322}}\right)^{1/4} \approx 1. \quad \text{(C12)}$$

Since 211 SR stops for $\alpha \geq 0.5$, we conclude that the net growth rate of all four levels (and their radial overtones) is negative over the whole range of relevant parameter space.

APPENDIX D: EQUILIBRIUM RATIO FOR MODERATE SELF-INTERACTIONS

We derive a more precise formula for the value of the time-independent equilibrium ratio by the system of equations (46). In terms of the γ rates,

$$\dot{\varepsilon}_{211} = \gamma^{211\times\infty}_{322\times322}\varepsilon_{211}\varepsilon^2_{322} - 2\gamma^{322\times\mathrm{BH}}_{211\times211}\varepsilon^2_{211}\varepsilon_{322}, \quad \text{(D1a)}$$

$$\dot{\varepsilon}_{322} = -2\gamma^{211\times\infty}_{322\times322}\varepsilon_{211}\varepsilon^2_{322} + \gamma^{322\times\mathrm{BH}}_{211\times211}\varepsilon^2_{211}\varepsilon_{322}. \quad \text{(D1b)}$$

Therefore,

$$\frac{1}{\varepsilon_{211}\varepsilon_{322}}\frac{d}{dt}\left(\frac{\varepsilon_{322}}{\varepsilon_{211}}\right) = \gamma^{322\times\mathrm{BH}}_{211\times211} + 2(\gamma^{322\times\mathrm{BH}}_{211\times211} - \gamma^{211\times\infty}_{322\times322})\left(\frac{\varepsilon_{322}}{\varepsilon_{211}}\right) - \gamma^{211\times\infty}_{322\times322}\left(\frac{\varepsilon_{322}}{\varepsilon_{211}}\right)^2. \quad \text{(D2)}$$

The zeros of the right-hand side are

$$\eta^B = \frac{1}{\gamma^{211\times\infty}_{322\times322}}\left(\gamma^{322\times\mathrm{BH}}_{211\times211} - \gamma^{211\times\infty}_{322\times322} \pm \sqrt{(\gamma^{322\times\mathrm{BH}}_{211\times211})^2 - \gamma^{322\times\mathrm{BH}}_{211\times211}\gamma^{211\times\infty}_{322\times322} + (\gamma^{211\times\infty}_{322\times322})^2}\right). \quad \text{(D3)}$$

Since the right-hand side is an inverted parabola, the "+" solution is dynamically stable (attractive), while the "−" solution is unstable. Parametrically, $\gamma^{322\times\mathrm{BH}}_{211\times211} \propto \alpha^{11}$ and $\gamma^{211\times\infty}_{322\times322} \propto \alpha^8$. Therefore at small α, $\gamma^{322\times\mathrm{BH}}_{211\times211} < \gamma^{211\times\infty}_{322\times322}$, and so the "−" root is negative. Moreover, the "+" root is

$$(\eta^B)_{\mathrm{small}\,\alpha} \approx \frac{1}{2}\frac{\gamma^{322\times\mathrm{BH}}_{211\times211}}{\gamma^{211\times\infty}_{322\times322}} = \frac{\varepsilon^{\mathrm{eq}}_{322}}{\varepsilon^{\mathrm{eq}}_{211}}. \quad \text{(D4)}$$

TABLE IV. Rates for gravitational processes involved in the evolution of the scalar cloud.

Process	Rate [γ/μ, Eq. (32)]
$\Gamma^{\mathrm{SR}}_{211}$	$4\times10^{-2}\alpha^8(a_* - 2\alpha(1+\sqrt{1-a_*^2}))$
$\Gamma^{\mathrm{SR}}_{322}$	$8\times10^{-5}\alpha^{12}(a_* - \alpha(1+\sqrt{1-a_*^2}))$
$\Gamma^{\mathrm{SR}}_{433}$	$2\times10^{-8}\alpha^{16}(a_* - \frac{2}{3}\alpha(1+\sqrt{1-a_*^2}))$
$\Gamma^{\mathrm{SR}}_{544}$	$2\times10^{-12}\alpha^{20}(a_* - \frac{1}{2}\alpha(1+\sqrt{1-a_*^2}))$
$\Gamma^{\mathrm{GW,ann}}_{211}$	$1\times10^{-2}\alpha^{14}$
$\Gamma^{\mathrm{GW,ann}}_{322}$	$3\times10^{-8}\alpha^{18}$
$\Gamma^{\mathrm{GW,tr}}_{322\to211}$	$5\times10^{-6}\alpha^{10}$

TABLE V. Rates for quartic processes involving nonrelativistic bound states.

Process	Rate [γ/μ, Eq. (32)]
$\Gamma^{322\times\mathrm{BH}}_{211\times211}$	$4.3\times10^{-7}\alpha^{11}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$
$\Gamma^{422\times\mathrm{BH}}_{211\times211}$	$1.5\times10^{-7}\alpha^{11}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$
$\Gamma^{322\times\mathrm{BH}}_{211\times411}$	$2.5\times10^{-8}\alpha^{11}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$
$\Gamma^{322\times\mathrm{BH}}_{411\times411}$	$9.8\times10^{-11}\alpha^{11}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$
$\Gamma^{433\times\mathrm{BH}}_{211\times322}$	$9.1\times10^{-8}\alpha^{11}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$
$\Gamma^{544\times\mathrm{BH}}_{322\times322}$	$1.9\times10^{-9}\alpha^{11}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$
$\Gamma^{544\times\mathrm{BH}}_{211\times433}$	$1.1\times10^{-9}\alpha^{11}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$
$\Gamma^{655\times\mathrm{BH}}_{322\times433}$	$2.8\times10^{-10}\alpha^{11}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$
$\Gamma^{655\times\mathrm{BH}}_{211\times544}$	$3.6\times10^{-12}\alpha^{11}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$
$\Gamma^{766\times\mathrm{BH}}_{433\times433}$	$2.1\times10^{-10}\alpha^{11}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$
$\Gamma^{877\times\mathrm{BH}}_{433\times544}$	$5.2\times10^{-12}\alpha^{11}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$
$\Gamma^{988\times\mathrm{BH}}_{544\times544}$	$1.6\times10^{-12}\alpha^{11}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$
$\Gamma^{1099\times\mathrm{BH}}_{544\times655}$	$5.6\times10^{-13}\alpha^{11}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$
$\Gamma^{433\times200}_{211\times422}$	$1.1\times10^{-9}\alpha^{7}(\frac{M_{\mathrm{pl}}}{f})^4(1+\sqrt{1-a_*^2})$

TABLE VI. Rates for quartic processes leading to nonrelativistic emission.

Process	Rate [γ/μ, Eq. (32)]
$\Gamma^{100\times\infty}_{211\times211}$	$1.3\times10^{-7}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{100\times\infty}_{211\times322}$	$8.5\times10^{-9}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{100\times\infty}_{322\times322}$	$1.1\times10^{-10}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{211\times\infty}_{322\times411}$	$3.8\times10^{-9}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{211\times\infty}_{322\times322}$	$1.1\times10^{-8}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{211\times\infty}_{322\times433}$	$2.6\times10^{-9}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{211\times\infty}_{433\times433}$	$9.2\times10^{-11}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{211\times\infty}_{322\times544}$	$6.1\times10^{-11}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{211\times\infty}_{433\times544}$	$1.9\times10^{-11}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{211\times\infty}_{544\times544}$	$4.2\times10^{-13}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{322\times\infty}_{544\times544}$	$4.4\times10^{-11}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{322\times\infty}_{433\times544}$	$7.8\times10^{-10}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{2l-1\times\infty}_{322\times322}$	$2.3\times10^{-10}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{211\times\infty}_{655\times322}$	$7.3\times10^{-13}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{211\times\infty}_{655\times433}$	$4.6\times10^{-13}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{211\times\infty}_{655\times544}$	$6.9\times10^{-14}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{211\times\infty}_{655\times655}$	$1.1\times10^{-15}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{322\times\infty}_{655\times433}$	$3.7\times10^{-11}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{322\times\infty}_{655\times544}$	$1.6\times10^{-11}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{322\times\infty}_{655\times655}$	$6.2\times10^{-13}\alpha^8(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{433\times\infty}_{766\times766}$	$5.6\times10^{-13}\alpha^8(\frac{M_{\rm pl}}{f})^4$

TABLE VII. Rates for self-interaction induced relativistic emission processes.

Process	Rate [γ/μ, Eq. (32)]
$\Gamma^{2\to1}_{211}$(cubic)	$1.9\times10^{-4}\alpha^{14}\lvert C\rvert^2(\frac{M_{\rm pl}}{f})^2$
$\Gamma^{3\to1}_{211}$	$5\times10^{-9}\alpha^{21}(\frac{M_{\rm pl}}{f})^4$
$\Gamma^{3\to1}_{322}$	$6\times10^{-14}\alpha^{27}(\frac{M_{\rm pl}}{f})^4$

APPENDIX E: BOUNDARY OF THE REGIME OF EARLY EQUILIBRIUM

We derive a more precise formula for the value of $f/M_{\rm pl}$ such that the SR growth of 211 is halted before $\mathcal{O}(1)$ of the spin is extracted. At early times, if we neglect the dependence of $\gamma^{\rm SR}_{211}$ on the BH spin a_*,

$$\varepsilon_{211}(t)\approx\frac{1}{GM^2}e^{\gamma^{\rm SR}_{211}t}. \tag{E1}$$

We use this into

$$\dot{\varepsilon}_{322}=\gamma^{322\times\rm BH}_{211\times211}\varepsilon^2_{211}\varepsilon_{322}, \tag{E2}$$

where we neglect the dependence of $\gamma^{322\times\rm BH}_{211\times211}$ on a_*. Therefore

$$\varepsilon_{322}(t)\approx\frac{1}{GM^2}\exp\left[\frac{\gamma^{322\times\rm BH}_{211\times211}}{2\gamma^{\rm SR}_{211}}\frac{1}{G^2M^4}(e^{2\gamma^{\rm SR}_{211}t}-1)\right]. \tag{E3}$$

The condition for SR to be impeded is that

$$\gamma^{\rm SR}_{211}\simeq2\gamma^{322\times\rm BH}_{211\times211}\varepsilon_{211}(t)\varepsilon_{322}(t). \tag{E4}$$

Using the approximations (E1) and (E3), one finds that (E4) is satisfied at the time $t_{\rm eq}$ such that

$$\gamma^{\rm SR}_{211}t_{\rm eq}\approx\frac{1+4\beta\log\beta-2\beta W(\frac{1}{2}\beta e^{1/2\beta})}{4\beta}, \tag{E5}$$

where

$$\beta\equiv G^2M^4\frac{\gamma^{\rm SR}_{211}}{2\gamma^{322\times\rm BH}_{211\times211}}, \tag{E6}$$

and $W(z)$ is the product logarithm (sometimes called the Lambert W function).

When $\gamma^{\rm SR}_{211}t\simeq\log(GM^2\Delta a_*)$, then SR has happened completely. So, in order for (E4) to be obtained before SR has run its course, we must have

$$\frac{1+4\beta\log\beta-2\beta W(\frac{1}{2}\beta e^{1/2\beta})}{4\beta}\lesssim\log(GM^2\varepsilon^{\rm max}_{211}). \tag{E7}$$

(E7) implicitly defines $f_{\rm thresh}$. Note that since $M\gg M_{\rm pl}$, $\beta\gg1$ for much of parameter space. One can then approximate $W(z)$ with the leading terms of its expansion around a large argument: $W(z)\to\log z-\log(\log z)$ as $z\to+\infty$. In this approximation, the left-hand side of (E7) becomes $\approx\log\sqrt{2\beta\log\beta}$, and the condition for SR to be halted early simplifies to

$$2\gamma^{\rm SR}_{211}\log(GM^2)\lesssim\gamma^{322\times\rm BH}_{211\times211}(\varepsilon^{\rm max}_{211})^2. \tag{E8}$$

APPENDIX F: CLOUD MASS

Here we calculate the mass of the cloud in the case $f\to\infty$, i.e., in the purely gravitational case. We will do the computation for the 211 level for clarity, but it is straightforward to generalize the formalism to any $n\ell m$ level. To simplify notation, we drop the level subscripts for the rest of our discussion here. The cloud parameters are referring to 211, unless stated otherwise.

Since the BH loses $<0.1\%$ of its mass due to SR, we usually treat its mass to be constant, or, equivalent, that α is just a parameter. In the case of self-interactions, in particular, the cloud tends to grow to a smaller occupation

number, which strengthens this assumption. A further simplification comes from setting $\omega \simeq \mu$. By noting that $\varepsilon = -\dot{a}_*$ in this regime, we get that $\varepsilon_{\max} = \Delta a_*$. The final a_* can be found by setting the SR rate equal to zero. Eventually, the maximum occupation number one gets is

$$\varepsilon_{\max} = a_*(0) - \frac{4\alpha}{1+4\alpha^2}. \tag{F1}$$

In general, the equations we need to solve are

$$\dot{N} = \gamma_{\rm SR} N \tag{F2a}$$

$$\dot{M} = -\omega_{211}\gamma_{\rm SR} N \tag{F2b}$$

$$\dot{J} = -\gamma_{\rm SR} N \tag{F2c}$$

$$a_* \equiv \frac{J}{GM^2} \tag{F2d}$$

where

$$\omega_{211} = \mu\left(1 - \frac{\alpha^2}{8}\right). \tag{F2e}$$

We define the ε with respect to the initial BH mass M^i, i.e., $\varepsilon = N/G(M^i)^2$ and Eqs. (F2) become:

$$\dot{\varepsilon} = \gamma_{\rm SR}\varepsilon \tag{F3}$$

$$\dot{\alpha} = -\alpha_i^2\left(1 - \frac{\alpha^2}{8}\right)\dot{\varepsilon} \tag{F4}$$

$$\dot{a}_* = -\frac{\dot{\alpha}}{\alpha}\left[2a_* - \frac{1}{\alpha(1-\alpha^2/8)}\right] \tag{F5}$$

where $\alpha_i = \alpha(t=0)$, given by the initial BH mass. The usual treatment is to expand these equations for small α, which is equivalent to neglecting terms of order $\mathcal{O}(\alpha)$. This reduces Eq. (F4) to $\dot{\alpha} = 0$. However, the expansion in Eq. (F5) has to be taken more carefully because the denominator is also small in this limit. By substituting Eq. (F4) it becomes evident that the first term is of order $\mathcal{O}(\alpha)$, whereas the second is independent of α. Therefore, we can neglect the former, which gives the standard result $\dot{a}_* = -\dot{\varepsilon}$.

Equation (F4) has the following solution

$$\alpha = -2\sqrt{2}\tanh\left[\frac{1}{4}\left(\sqrt{2}\alpha_i^2\varepsilon - 4\,\mathrm{arctanh}\left[\frac{\alpha_i}{2\sqrt{2}}\right]\right)\right]. \tag{F6}$$

Now, Eq. (F5) can also be solved analytically. The result is

$$a_* = \frac{1}{\alpha^2}\left[a_0\alpha_i^2 - 2\sqrt{2}\,\mathrm{arctanh}\left(\frac{2\sqrt{2}(\alpha-\alpha_i)}{-8+\alpha\alpha_i}\right)\right]. \tag{F7}$$

The final spin of the BH is that which saturates the SR condition $\omega - m\Omega_H$ is

$$a_*^{\rm fin} = -\frac{8(-8\alpha_{\rm fin} + \alpha_{\rm fin}^3)}{16 + 64\alpha_{\rm fin}^2 - 16\alpha_{\rm fin}^4 + \alpha_{\rm fin}^6}, \tag{F8}$$

where the "fin" superscript denotes final quantities, after the 211 cloud has been saturated and the BH has spun down.

Now we can use Eqs. (F6), (F7) and (F8) to numerically solve for $\varepsilon_{\max}$, the final occupation number of the cloud. The mass of the cloud is then $M_{\rm c} = \varepsilon_{\max} G(M^i)^2\omega$.

By neglecting the α^2 term in Eq. (F2e), i.e., by approximating $\omega \simeq \mu$, we can get a simpler analytic result for the final BH mass. In this case, the equivalents of Eqs. (F6), (F7) are

$$\alpha = \alpha_i(1 - \alpha_i\varepsilon) \tag{F9}$$

$$a_* = \frac{a_0 - \varepsilon}{(1-\alpha_i\varepsilon)^2} \tag{F10}$$

which can be used along with Eq. (F8), truncated to $\mathcal{O}(\alpha^2)$, to give the final occupation number of the cloud. We find that

$$\varepsilon_{\max} = \frac{1 - 8\alpha_i^2 + 8\alpha_i^3 a_0 - \sqrt{1 - 16\alpha_i^2 + 32a_0\alpha_i^3 - 16a_0^2\alpha_i^4}}{8(-\alpha_i^3 + a_0\alpha_i^4)} \tag{F11}$$

where $a_0 = a_*(t=0)$.

In Fig. 27 we plot the ratio of the final cloud mass over the initial BH mass. We solve numerically Eq. (F8) with

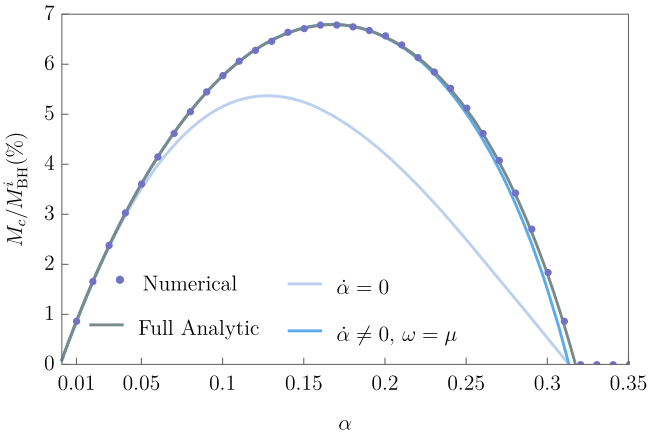


FIG. 27. Ratio of the final mass of the cloud to the initial BH mass. We plot points from the numerical evolution of Eqs. (F3)–(F5), the full analytic result of Eqs. (F6)–(F8), as well as the $\dot{\alpha} = 0$ approximation of Eq. (F1) and the $\omega_{211} \simeq \mu, \dot{\alpha} \neq 0$ approximation of (F11). The cloud can grow to have a mass of up to 7% of the initial BH mass. This plot assumes an initial spin $a_*(t_0) = 0.9$.

respect to $\epsilon_{\max}$ and compare it to the numerical evolution of Eqs. (F3)–(F5). We also plot the results of Eq. (F1) and Eq. (F11) for comparison. We find that the mass of the cloud can grow up to 7% of the initial BH mass.

APPENDIX G: SELF-GRAVITY ENERGY CORRECTIONS

The Poisson equation for the gravitational potential sourced by the cloud is

$$\nabla^2\Phi_{\rm SG} = 4\pi G\mu|\psi|^2 \tag{G1}$$

where ψ is the wave function of the cloud, i.e.,

$$\psi(\mathbf{r}) = \sum_{nlm}\sqrt{N_{nlm}}\psi_{nlm} \tag{G2}$$

where N_{nlm} are the occupation numbers of the levels and ψ_{nlm} the hydrogenic wave functions. Treating $\Phi_{\rm SG}$ as a small perturbation, the energy correction of the (n,l,m) level is

$$\begin{aligned}\Delta\omega_{nlm} &= \langle nlm|\mu\Phi_{\rm SG}|nlm\rangle \\ &= -G\mu^2\int|\psi_{nlm}(\mathbf{r})|^2\int\frac{|\psi(\mathbf{r}')|^2}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}^3\mathbf{r}'\mathrm{d}^3\mathbf{r}\end{aligned} \tag{G3}$$

Expanding $1/|\mathbf{r}-\mathbf{r}'|$ in spherical harmonics we get

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = 4\pi\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\frac{1}{2l'+1}\frac{r_<^{l'}}{r_>^{l'+1}}Y_{l'}^{m'*}(\theta',\phi')Y_{l'}^{m'}(\theta,\phi), \tag{G4}$$

where $r_{<(>)}$ is the smallest (largest) of r and r'. We can perform the integration over θ and ϕ, since $\psi_{nlm}\propto Y_l^m$. By the selection rules of the spherical harmonics we can write

$$Y_l^m Y_l^{m*} = \sum_{k=0}^{l}c_{k,lm}Y_{2k}^0, \qquad c_{k,lm} = \int|Y_l^m|^2Y_{2k}^{0*}\mathrm{d}\Omega. \tag{G5}$$

Therefore, the integral over θ and ϕ selects $m'=0$ and $l'=2k$, giving

$$\begin{aligned}\Delta\omega_{nlm} = -4\pi G\mu^2\sum_{k=0}^{l}\frac{c_{k,lm}}{4k+1}\int R_{nl}(r)r^2 \\ \times\int|\psi(\mathbf{r}')|^2\frac{r_<^{2k}}{r_>^{2k+1}}Y_{2k}^{0*}(\theta',\phi')\mathrm{d}^3\mathbf{r}'\mathrm{d}r\end{aligned} \tag{G6}$$

where R_{nl} are the hydrogenic radial wave functions. We will now make the simplifying assumption that the ψ given by Eq. (G2) is a sum of levels such that $(n,l,m)=(l+1,l,l)$, which is the case treated in this work. Since $|\psi|^2$ is integrated against Y_{2k}^0, only the terms consisting of products of complex conjugates will survive. Thus, we can substitute the integrand as follows:

$$\begin{aligned}&|\psi(\mathbf{r}')|Y_{2k}^{0*}(\theta',\phi') \\ &\to\sum_{l'=0}|N_{l'+1,l',l'}^{1/2}R_{l'+1,l'}(r')|^2|Y_{l'}^{l'}(\theta',\phi')|^2Y_{2k}^{0*}(\theta',\phi').\end{aligned} \tag{G7}$$

Then the integral over θ' and ϕ' is just $c_{k,l'l'}$, as defined in Eq. (G5). Note that this integral is nonzero only for $l'>k$. Thus, we can rewrite the sum $\sum_{l'=0}\to\sum_{l'=k}$. The coefficients $c_{k,l'l'}$ have a simple analytic form

$$c_{k,l'l'} = (-1)^k\frac{2l'+1}{\sqrt{4\pi}}\sqrt{4k+1}\frac{(2l')!(2k)!(l'+k)!}{(k!)^2(2l'+2k+1)!(l'-k)!}, \quad \text{for } k<l'. \tag{G8}$$

The energy corrections are then

$$\Delta\omega_{nlm} = -4\pi G\mu^2\sum_{k=0}^{l}\frac{c_{k,lm}}{4k+1}\sum_{l'=k}N_{l'+1,l',l'}c_{k,l',l'}I_{nl}^{kl'} \tag{G9}$$

where the last quantity is the radial integral given by

$$I_{nl}^{kl'} = \int R_{nl}^2(r)\int R_{l'+1,l'}^2(r')\frac{r_<^{2k}}{r_>^{2k+1}}r'^2r^2\mathrm{d}r'\mathrm{d}r, \tag{G10}$$

which can be calculated analytically. Assuming a simultaneous occupation of just 211 and 322, the corrections are

$$\Delta\omega_{211}\simeq-\frac{\alpha^3\mu}{GM^2}(0.19N_{211}+0.11N_{322}) \tag{G11}$$

$$\Delta\omega_{322}\simeq-\frac{\alpha^3\mu}{GM^2}(0.11N_{211}+0.09N_{322}) \tag{G12}$$

APPENDIX H: FREQUENCY DRIFTS

The corrections to the energy of the 211 and 322 levels from self-interactions and self-gravity were calculated in Appendixes B 1 and G respectively. The angular frequency of a particle occupying 211 or 322 is

$$\begin{aligned}\omega_{211} = \mu\left(1-\frac{\alpha^2}{8}\right)-\mu\alpha^5\left(\frac{M_{\rm pl}}{f}\right)^2(\kappa_1^\lambda\varepsilon_{211}+\kappa_2^\lambda\varepsilon_{322}) \\ -\mu\alpha^3(\kappa_1^{\rm gr}\varepsilon_{211}+\kappa_2^{\rm gr}\varepsilon_{322}),\end{aligned} \tag{H1a}$$

$$\begin{aligned}\omega_{322} = \mu\left(1-\frac{\alpha^2}{36}\right)-\mu\alpha^5\left(\frac{M_{\rm pl}}{f}\right)^2(\kappa_3^\lambda\varepsilon_{211}+\kappa_4^\lambda\varepsilon_{322}) \\ -\mu\alpha^3(\kappa_3^{\rm gr}\varepsilon_{211}+\kappa_4^{\rm gr}\varepsilon_{322}),\end{aligned} \tag{H1b}$$

where $\alpha = \mu M M_{\rm pl}^{-2}$ and $\kappa_1^\lambda = 1.2 \times 10^{-4}$, $\kappa_2^\lambda = 3.5 \times 10^{-5}$, $\kappa_1^{\rm gr} = 0.19$, $\kappa_2^{\rm gr} = 0.11$, $\kappa_3^\lambda = 3.5 \times 10^{-5}$, $\kappa_4^\lambda = 1.4 \times 10^{-5}$, $\kappa_3^{\rm gr} = 0.11$ and $\kappa_4^{\rm gr} = 0.09$ are numerical coefficients.

In what follows, we define the frequency ν as

$$\nu \equiv \frac{\omega}{2\pi}. \tag{H2}$$

So the frequency drifts $\dot{\nu}$ are given by,

$$\begin{aligned}\dot{\nu}_{211} = -\frac{\mu\alpha^2}{2\pi}\Bigg[&\frac{1}{4}\frac{\dot{\alpha}}{\alpha} + \alpha^2\left(\frac{M_{\rm pl}}{f}\right)^2 \\ &\times [\alpha(\kappa_1^\lambda \dot{\varepsilon}_{211} + \kappa_2^\lambda \dot{\varepsilon}_{322}) + 5(\kappa_1^\lambda \varepsilon_{211} + \kappa_2^\lambda \varepsilon_{322})\dot{\alpha}] \\ &+ +[\alpha(\kappa_1^{\rm gr} \dot{\varepsilon}_{211} + \kappa_2^{\rm gr} \dot{\varepsilon}_{322}) + 3(\kappa_1^{\rm gr} \varepsilon_{211} + \kappa_2^{\rm gr} \varepsilon_{322})\dot{\alpha}]\Bigg]\end{aligned} \tag{H3a}$$

$$\begin{aligned}\dot{\nu}_{322} = -\frac{\mu\alpha^2}{2\pi}\Bigg[&\frac{1}{18}\frac{\dot{\alpha}}{\alpha} + \alpha^2\left(\frac{M_{\rm pl}}{f}\right)^2 \\ &\times [\alpha(\kappa_3^\lambda \dot{\varepsilon}_{211} + \kappa_4^\lambda \dot{\varepsilon}_{322}) + 5(\kappa_3^\lambda \varepsilon_{211} + \kappa_4^\lambda \varepsilon_{322})\dot{\alpha}] \\ &+ +[\alpha(\kappa_3^{\rm gr} \dot{\varepsilon}_{211} + \kappa_4^{\rm gr} \dot{\varepsilon}_{322}) + 3(\kappa_3^{\rm gr} \varepsilon_{211} + \kappa_4^{\rm gr} \varepsilon_{322})\dot{\alpha}]\Bigg]\end{aligned} \tag{H3b}$$

to leading order in α for every term.

The mass of the BH evolves according to (37), which can be written equivalently as an equation for α as

$$\dot{\alpha} \simeq -\alpha^2(\gamma_{211}^{\rm SR}\varepsilon_{211} + \gamma_{322}^{\rm SR}\varepsilon_{322} - \gamma_{211\times211}^{322\times{\rm BH}}\varepsilon_{211}^2\varepsilon_{322}). \tag{H4}$$

As a result, the last terms in the second and third row of Eqs. (H3a) and (H3b) are parametrically suppressed by an additional power of α and ε_i compared to the respective first term and thus will be neglected in what follows. In addition, all drifts are given to leading order in α and are the maximum possible for each individual regime.

In what follows we calculate the frequency drifts of the GWs coming from annihilations of two 211 particles and from transitions from 322 to 211. These are given by the relations $\dot{\nu}_{\rm ann} \equiv 2\dot{\nu}_{211}$ and $\dot{\nu}_{\rm tr} \equiv \dot{\nu}_{322} - \dot{\nu}_{211}$ We separate the sources of frequency drifts in the following categories:

(1) Due to the change of the mass of the BH, given by the first terms of (H3a) and (H3b), denoted as ν^α.
(2) Due to the change in the self-interaction energy, given by the second term of (H3a) and (H3b), denoted as ν^λ.
(3) Due to the change in the self-gravitational energy, given by the third term of (H3a) and (H3b), denoted as $\nu^{\rm gr}$.

In the regime of small self-interactions we treat the depletion due to gravitational radiation (annihilations and transitions) separately for points 2 and 3 above, and we denote by the superscript "GW".

We also note that there is an additional source of frequency drift coming from the change of the radial velocity of the BH to the observer, but for isolated black holes it is $\dot{\nu}_{\rm Doppler} < 10^{-19}$ Hz/s [40], which is negligible.

For reference, LIGO/Virgo continuous wave searches currently cover a range of positive to negative frequency derivatives of [83]

$$2 \times 10^{-9}\ {\rm Hz/s} \quad {\rm through} - 1 \times 10^{-8}\ {\rm Hz/s}. \tag{H5}$$

All drift calculations carried out here are to leading approximation in α [which is accurate only for $\alpha \ll a_*(0)$] but the formalism includes in principle all higher-order corrections. At higher α the calculations can be carried out numerically using the full expressions and the numerical rates, but at $\alpha \gtrsim 0.2$ the approximation of the two-level system essentially breaks down. We have verified that, for our purposes, the leading order approximation gives accurate results.

1. Small self-coupling

Here we revisit the frequency drifts from purely gravitational interactions, i.e., $f \to \infty$ as described in [12], which corresponds to region (A) of Fig. 3. There is a clear separation of times when different levels grow, so whenever a higher level gets populated, the lower ones have already fallen back into the BH, as their SR rates have become negative. In what follows, we will consider only 211, from which comes the stronger signal.

The interesting region for signatures is when the BH has spun down, the level has saturated and slowly gets depleted by radiating GWs. The only source of a frequency drift then comes from the gravitational self-energy of the cloud, given by the last line of Eqs. (H3a) and (H3b). In particular, the last term is exactly zero, since $\dot{\alpha} = 0$.

The 211 cloud obeys the equation $\dot{\varepsilon}_{211} = -2\gamma_{211\times211}^{\rm GW}\alpha^{14}\varepsilon_{211}^2$. The maximum drift comes about when $\varepsilon_{211} = \varepsilon_{211}^{\rm max} \simeq \Delta a_*$ (for a better estimate, see Appendix F), when SR shuts just off. For $a_*(0) = 0.9$ we find the drifts to be

$$\dot{\nu}_{\rm ann}^{\lambda,{\rm GW}} \simeq 4\times10^{-22}\,\frac{\rm Hz}{\rm sec}\left(\frac{\alpha}{0.075}\right)^{19}\left(\frac{\mu}{10^{-12}\,{\rm eV}}\right)^2\left(\frac{10^{19}\,{\rm GeV}}{f}\right)^2 \tag{H6}$$

$$\dot{\nu}_{\rm ann}^{\rm gr,GW} \simeq 8\times10^{-17}\,\frac{\rm Hz}{\rm sec}\left(\frac{\alpha}{0.075}\right)^{17}\left(\frac{\mu}{10^{-12}\,{\rm eV}}\right)^2 \tag{H7}$$

In the small self-interactions regime, the drift coming from self-interactions is always subdominant to that of self-gravity in the parameter space of interest.

The drift can become larger than the range LIGO/Virgo cover [Eq. (H5)] only for α around 0.27, taking higher order α contributions into account.

2. Moderate self-coupling

Here we are interested in the region where both levels are occupied and they drift away slowly, which corresponds to region (B) of Fig. 3. In this regime 211 reaches its maximum occupation Δa_* and we can use Eq. (47) to relate the ε_{322} to ε_{211}. Note that even though the BH has spun down due to the growth of 211, $\dot{\alpha} \neq 0$, since particles fall back into the BH, as described by the last term of Eq. (H4). The resulting frequency drifts are as follows:

Due to the change of the BH mass:

$$\dot{\nu}^{\alpha}_{\rm ann} \simeq -10^{-11}\ \frac{\rm Hz}{\rm sec}\left(\frac{10^{17}\ {\rm GeV}}{f}\right)^4\left(\frac{\mu}{10^{-12}\ {\rm eV}}\right)^2\left(\frac{\alpha}{0.075}\right)^{17} \tag{H8a}$$

$$\dot{\nu}^{\alpha}_{\rm tr} \simeq 3\times 10^{-12}\ \frac{\rm Hz}{\rm sec}\left(\frac{10^{17}\ {\rm GeV}}{f}\right)^4\left(\frac{\mu}{10^{-12}\ {\rm eV}}\right)^2\left(\frac{\alpha}{0.075}\right)^{17} \tag{H8b}$$

The negative sign in Eq. (H8) comes from the fact that the SR rates are zero, so the BH is actually gaining mass by the depletion of 211, from the last term of Eq. (H4).

Due to self-interactions:

$$\dot{\nu}^{\lambda}_{\rm ann} \simeq 6\times 10^{-13}\ \frac{\rm Hz}{\rm sec}\left(\frac{10^{17}\ {\rm GeV}}{f}\right)^6\left(\frac{\mu}{10^{-12}\ {\rm eV}}\right)^2\left(\frac{\alpha}{0.075}\right)^{19} \tag{H9a}$$

$$\dot{\nu}^{\lambda}_{\rm tr} \simeq -2\times 10^{-13}\ \frac{\rm Hz}{\rm sec}\left(\frac{10^{17}\ {\rm GeV}}{f}\right)^6\left(\frac{\mu}{10^{-12}\ {\rm eV}}\right)^2\left(\frac{\alpha}{0.075}\right)^{19} \tag{H9b}$$

Due to self-gravity:

$$\dot{\nu}^{\rm gr}_{\rm ann} \simeq 10^{-11}\ \frac{\rm Hz}{\rm sec}\left(\frac{10^{17}\ {\rm GeV}}{f}\right)^4\left(\frac{\mu}{10^{-12}\ {\rm eV}}\right)^2\left(\frac{\alpha}{0.075}\right)^{17} \tag{H10a}$$

$$\dot{\nu}^{\rm gr}_{\rm tr} \simeq -2\times 10^{-12}\ \frac{\rm Hz}{\rm sec}\left(\frac{10^{17}\ {\rm GeV}}{f}\right)^4\left(\frac{\mu}{10^{-12}\ {\rm eV}}\right)^2\left(\frac{\alpha}{0.075}\right)^{17} \tag{H10b}$$

These are calculated for $a_*(0) = 0.9$. Note that α scalings of Eqs. (H8) and (H10) are the same, which comes from the fact that SR has shut off and the scalings in both $\dot{\alpha}$ and $\dot{\varepsilon}_i$ of Eqs. (H4) and of (H3) are set by the same term, i.e., $\gamma^{322\times{\rm BH}}_{211\times 211}\varepsilon^2_{211}\varepsilon_{322}$. This is why the numerical coefficients of both the annihilation and transition drifts are very close. In particular, for the annihilation drift we find more precisely that

$$\dot{\nu}^{\alpha}_{\rm ann} + \dot{\nu}^{\rm gr}_{\rm ann} \simeq 1.4\times 10^{-12}\ \frac{\rm Hz}{\rm sec}\left(\frac{10^{17}\ {\rm GeV}}{f}\right)^4\left(\frac{\mu}{10^{-12}\ {\rm eV}}\right)^2\left(\frac{\alpha}{0.075}\right)^{17} \tag{H11}$$

Within the moderate self-interactions regime, we find that self-interactions are the dominant source of frequency drift for $f \lesssim 8.5\times 10^{16}(\alpha/0.1)$ GeV. The drift can become larger than the range LIGO/Virgo cover [Eq. (H5)] for $f \lesssim 5.6\times 10^{16}(\alpha/0.1)^{17/4}$ GeV. In Fig. 28 we plot the full annihilation frequency drift stemming from Eq. (H3) in this regime.

Analogously, for transitions, self-interactions are the dominant source of frequency drift for $f \lesssim 10^{17}(\alpha/0.1)$ GeV. The drift can become larger than the range LIGO/Virgo cover [Eq. (H5)] for $f \lesssim 4\times 10^{16}(\alpha/0.1)^{17/4}$ GeV. In Fig. 29 we plot the full annihilation frequency drift stemming from Eqs. (H3a) and (H3b), in this regime.

3. Large self-coupling

We are interested in the part of the evolution where the levels have reached their equilibrium values, given by Eqs. (55a) and (55b), which corresponds to region (C) in Fig. 3. These are slowly drifting because of the slow spin-down of the BH and the change of its mass. Neglecting the SR of ε_{322}, which is subdominant, the spin evolves according to

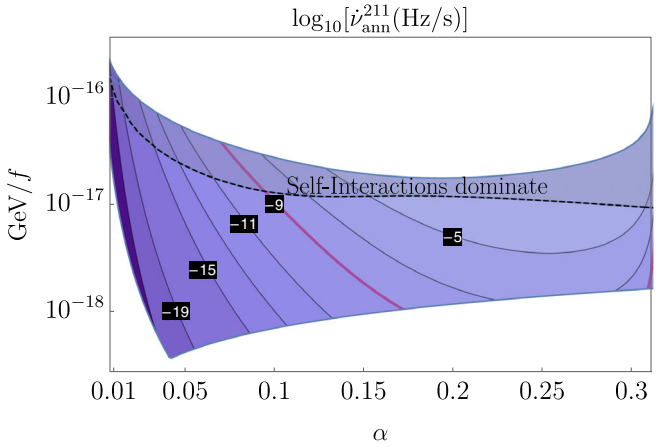


FIG. 28. Frequency drift contours for annihilations of axions to GWs, given by twice the quantity in Eq. (H3a), in the moderate self-coupling regime. The gray shaded region above the dashed black contour is where the drift due to self-interactions [second line of Eq. (H3a)] dominates. The red contour corresponds to the largest positive drift covered by LIGO/Virgo continuous searches, taken here to be 2×10^{-9} Hz/s [83].

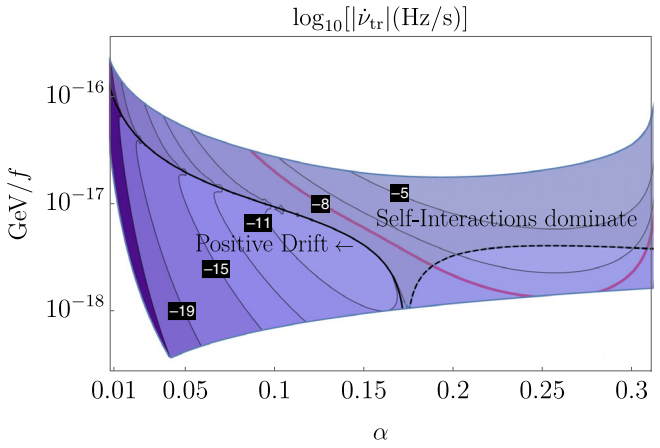


FIG. 29. Frequency drift contours for GWs sourced by axion transitions from 322 to 211, given by the difference of Eq. (H3a) and Eq. (H3b), in the moderate self-coupling regime. The gray shaded region above the black contour (solid and dashed) is where the drift due to self-interactions [second line of Eqs. (H3)] dominates. The frequency drift is negative to the right (i.e., to the large-α side) of the solid black line. Note that here we are plotting the absolute value of the frequency drift. The red contour corresponds to the largest *negative* drift covered by LIGO/Virgo continuous searches, taken here to be -1×10^{-8} Hz/s [83].

$$\dot{a}_* = -\gamma^{\rm SR}_{211}\varepsilon^{\rm eq}_{211}, \tag{H12}$$

and its mass changes according to Eq. (H4). By plugging in the equilibrium values of Eq. (55) we get

$$\dot{\alpha} = -\frac{2\alpha^2\gamma^{\rm SR}_{211}}{3\sqrt{3}}\frac{\sqrt{\gamma^{\rm SR}_{211}\gamma^{211\times\infty}_{322\times322}}}{\gamma^{322\times\rm BH}_{211\times211}}. \tag{H13}$$

Then, the equilibrium values evolve according to

$$\dot{\varepsilon} = \dot{\varepsilon}^{\rm eq} = \frac{\partial\varepsilon^{\rm eq}}{\partial a_*}\dot{a}_* + \frac{\partial\varepsilon^{\rm eq}}{\partial\alpha}\dot{\alpha}. \tag{H14}$$

The second term of Eq. (H14) gives a subdominant contribution and is further suppressed by another power of α compared to the first term. The signal is maximum at the beginning when $a_* \simeq a_*(0)$. The resulting drifts are given below.

Due to the change of the BH mass:

$$\dot{\nu}^{\alpha}_{\rm ann} \simeq 2\times10^{-13}\,\frac{\rm Hz}{\rm sec}\left(\frac{f}{10^{15}\,\rm GeV}\right)^2\left(\frac{\mu}{10^{-12}\,\rm eV}\right)^2\left(\frac{\alpha}{0.075}\right)^8 \tag{H15a}$$

$$\dot{\nu}^{\alpha}_{\rm tr} \simeq -6\times10^{-14}\,\frac{\rm Hz}{\rm sec}\left(\frac{f}{10^{15}\,\rm GeV}\right)^2\left(\frac{\mu}{10^{-12}\,\rm eV}\right)^2\left(\frac{\alpha}{0.075}\right)^8 \tag{H15b}$$

Due to self-interactions:

$$\dot{\nu}^{\lambda}_{\rm ann} \simeq 3\times10^{-13}\,\frac{\rm Hz}{\rm sec}\left(\frac{f}{10^{15}\,\rm GeV}\right)^2\left(\frac{\mu}{10^{-12}\,\rm eV}\right)^2\left(\frac{\alpha}{0.075}\right)^7 \tag{H16a}$$

$$\dot{\nu}^{\lambda}_{\rm tr} \simeq -10^{-13}\,\frac{\rm Hz}{\rm sec}\left(\frac{f}{10^{15}\,\rm GeV}\right)^2\left(\frac{\mu}{10^{-12}\,\rm eV}\right)^2\left(\frac{\alpha}{0.075}\right)^7 \tag{H16b}$$

Due to the self-gravity:

$$\dot{\nu}^{\rm gr}_{\rm ann} \simeq 5\times10^{-16}\,\frac{\rm Hz}{\rm sec}\left(\frac{f}{10^{15}\,\rm GeV}\right)^4\left(\frac{\mu}{10^{-12}\,\rm eV}\right)^2\left(\frac{\alpha}{0.075}\right)^5 \tag{H17a}$$

$$\dot{\nu}^{\rm gr}_{\rm tr} \simeq -10^{-16}\,\frac{\rm Hz}{\rm sec}\left(\frac{f}{10^{15}\,\rm GeV}\right)^4\left(\frac{\mu}{10^{-12}\,\rm eV}\right)^2\left(\frac{\alpha}{0.075}\right)^5 \tag{H17b}$$

These are calculated for $a_*(0) = 0.9$ as well.

In the large self-interactions regime, for $\alpha \gtrsim 0.1$ the change of the mass of the BH is the dominant source of frequency drift for annihilations. For $\alpha \lesssim 0.1$ self-interactions are dominant. The drift can become larger than the range LIGO/Virgo cover [Eq. (H5)] for $f \gtrsim 3\times10^{16}(\alpha/0.1)^{-4}$ GeV, which is relevant above $\alpha \simeq 0.1$.

Analogously for transitions, for $\alpha \gtrsim 0.13$, the change of the mass of the BH dominates and for $\alpha \lesssim 0.13$ self-interactions are dominant. The drift can become larger than the range LIGO/Virgo cover [Eq. (H5)] for $f \gtrsim 2.5\times10^{16}(\alpha/0.15)^{-4}$ GeV, which is relevant above $\alpha \simeq 0.13$.

APPENDIX I: PERTURBATIONS FROM BH COMPANION

When the primary BH has a companion, the perturbation in the gravitational potential induces mixing of different levels. In particular, SR levels can mix with non-SR ones, resulting in the depletion of the cloud. According to [45], the perturbation $\delta V_{\rm gr}$ mixes the levels ψ_i and ψ_j according to

$$\langle\psi_j|\delta V_{\rm c}|\psi_i\rangle = -\frac{\alpha M_{\rm c}}{M}\sum_{l\geq2}\sum_{|m|\leq l}\frac{4\pi}{2l+1}\frac{Y^{m*}_l(\theta_c,\phi_c)}{R_c^{l+1}}I_{\tilde{r}}I_\Omega \tag{I1}$$

where the subscript $M_{\rm c}$ is the mass of the companion, θ_c, ϕ_c its angular coordinates and $R_{\rm c}$ its distance from the primary BH of mass M, whereas the constant $\alpha = G\mu M$. We have also defined

$$I_r \equiv \int_0^\infty \mathrm{d}r r^{2+l} R_{n_j l_j}(r) R_{n_i l_i}(r), \tag{I2}$$

$$I_\Omega \equiv \int \mathrm{d}\Omega Y_{l_j}^{m_j *}(\theta,\phi) Y_{l_i}^{m_i}(\theta,\phi) Y_l^m(\theta,\phi) \tag{I3}$$

where R_{nl} is the radial part of the hydrogenic wave function and Y_m^l are the spherical harmonics.

Note that the first sum in Eq. (I1) starts from $l=2$, which demonstrates the fact that the first nonzero correction from gravity comes from the quadrupole term, as expected from the equivalence principle.[17]

We are interested in the mixing of the 211 level with non-SR levels of the BH, which can lead, in principle, to the depletion of our cloud. The dominant contribution comes from $n=2, l=1, m=-1$, and it is largest when the companion lies on the plane perpendicular to the spin of the primary BH, i.e., when $\theta_c = \pi/2$.

The horizon flux becomes *positive*, i.e., more axions fall back into the BH than are extracted due to SR [12], when

$$\left|\frac{\Gamma^j_{\mathrm{dump}}}{\Gamma_i}\right|^{1/2} \left|\frac{\langle\psi_j|\delta V_c|\psi_i\rangle}{\Delta\omega_{ji}}\right| > 1 \tag{I4}$$

where "dump" denotes the non-SR level that mixes with the SR one, Γ are the superradiance rates, and $\Delta\omega_{ji}$ is the difference of the energies between the two levels, which are given by [45]

$$\omega_{nlm} = \mu\left(1 - \frac{\alpha^2}{2n^2} - \frac{\alpha^4}{8n^4} + \frac{(2l-3n+1)\alpha^4}{n^4(l+1/2)} + \frac{2a_* m\alpha^5}{n^3 l(l+1/2)(l+1)}\right) \tag{I5}$$

The physical quantities measured for BH binaries are the BH masses, their spins and the orbital period. We assume that the companion is far away (which is where Eq. (I1) is valid), so we relate the distance to the orbital period using Kepler's 3rd Law: $R^3/T^2 = G(M+M_c)/(4\pi^2)$, where T is the orbital period. Then, the condition (I4) becomes parametrically:

$$\frac{M_c}{M}\frac{144\pi^2\sqrt{3}}{a_*\alpha^7(1+\frac{M_c}{M})(\mu T)^2} \gtrsim 1 \tag{I6}$$

where we have omitted an $\mathcal{O}(1)$ factor in the α region of interest.

The cloud may also be depleted by resonances that can occur when the period of the companion hits the energy difference between two levels, as shown in [45]. To estimate when this happens, we can compare the period to the energy splitting of the two mixing levels. As the companion spirals closer to the primary BH, its orbital period increases. When it crosses the value $\Delta\omega_{ji}^{-1}$, we expect that the cloud will be significantly depleted. A more careful analysis can be found in [45]. The condition, therefore, is

$$\frac{1}{6} a_* \alpha^5 \mu T \simeq 1. \tag{I7}$$

In deriving the BH spin bounds in Sec. VI, we take into account both Eqs. (I6) and (I7).

APPENDIX J: AXION WIND SENSITIVITY PROJECTIONS

As discussed in Sec. VIII A, given an axion coupling to nucleon spins, an axion oscillation $\varphi(t) = \varphi_0 \cos\omega t$ will act on nuclei as an effective magnetic field $B_a(t) = B_a \cos\omega t$. For nuclei which are spin-polarized in an external magnetic field, with Larmor frequency $\omega_0 \simeq \omega$, a transverse B_a will induce a transverse magnetic moment

$$\mu_a \simeq \mu_n^2 N_n B_a \frac{\omega_0}{\omega^2 - \omega_0^2 + i\omega\gamma} \tag{J1}$$

where μ_n is the nuclear magnetic moment, N_n is the total number of nuclei, and γ is the damping rate (in terms of the spin coherence time T_2, $\gamma = 2/T_2$ [114]).

In the absence of an axion forcing, the fluctuation spectrum for the transverse magnetic momentum is

$$S_{\mu\mu} \simeq \frac{\mu_n^2 N_n}{\gamma}\frac{1}{1+T_2^2(\omega-\omega_0)^2} \tag{J2}$$

which is related to the response function [Eq. (J1)] by the fluctuation-dissipation relation [114].

If we read out the transverse magnetic moment using a sufficiently sensitive magnetometer (e.g., a SQUID [111,114]), then it is possible to detect fluctuations as small as the quantum fluctuations from Eq. (J2). With a sensor that is bounded by the standard quantum limit [143,144], this is possible over a bandwidth $\sim 1/T_2$. Consequently, for an integration time of $T \gtrsim T_2$, we need

$$\frac{\mu_a^2}{S_{\mu\mu}/T} \simeq \frac{1}{2}\mu_n^2 N_n B_a^2 T T_2 \gtrsim \text{few} \tag{J3}$$

in order to reliably detect an axion signal.

To cover an $\mathcal{O}(1)$ axion mass range, we need to operate in $\sim\omega_0 T_2$ different resonant configurations (we will not be careful about constant factors). Consequently, if our total experimental time is T_{tot}, the time we spend in each configuration is $T \sim T_{\mathrm{tot}}/(\omega_0 T_2)$, and our sensitivity limit is[18]

[17] In [12] it was incorrectly assumed that the leading order contribution came from a dipole term. See [45] for an explanation.

[18] If $T \ll T_2$, then Eq. (J3) will not apply, since the response signal will not have time to ring up fully (equivalently, we cannot resolve the bandwidth of the response function). For the experimental parameters of interest, we will not be in this regime.

$$B_a^2 \gtrsim \text{few} \times \frac{\omega_0}{\mu_n^2 N_n T_{\text{tot}}} \tag{J4}$$

Note that, while this naive form does not depend on T_2, the signal amplitude from Eq. (J1) is $\propto T_2$; consequently, achieving a sensitive enough magnetometer may be easier for larger T_2. As discussed in Sec. VIII A, the CASPEr-Wind project aims to achieve spin-noise-limited sensitivities at frequencies in the kHz–30 kHz range [111].

APPENDIX K: DARK MATTER ABUNDANCE

In Sec. VI A, we reviewed models in which an axion dark matter abundance is generated via the early universe misalignment mechanism. For attractive potentials, if the initial value of the axion field is tuned close to the top of its potential, then the generated dark matter abundance can be enhanced through the "large-misalignment mechanism" [65]. In this Appendix, we give formulas for the DM density obtained in this way.

For a general cosine potential of the form $V(\varphi) = m^2 f^2 [1 - \cos(\varphi/f)]$, the enhanced final density for a large initial misalignment is given by

$$\frac{\rho}{\rho_{\pi/2}} \simeq 0.2[t_\mu^{\text{osc}} + 4 \log t_\mu^{\text{osc}}]^2 \tag{K1}$$

$$t_\mu^{\text{osc}} \equiv \log\left[\frac{1}{\pi - |\theta_0|} \frac{2^{1/4}\pi^{1/2}}{\Gamma(5/4)}\right] \tag{K2}$$

where $\rho_{\pi/2}$ is the final density when the initial amplitude of the field is $\theta_0 = \phi_0/f = \pi/2$, and t_μ^{osc} marks the onset of the oscillation in units of μ.

Fixing the final density to be the observed DM abundance today, we arrive at the relation [65]

$$\frac{f_{\text{DM}}}{m_{\text{pl}}} \simeq \frac{3^{1/2}}{2^{5/4} C_{\pi/2}^{1/2}} \left(\frac{\rho}{\rho_{\pi/2}}\right)^{-1/2} \left(\frac{H_{\text{eq}}}{\mu}\right)^{1/4} \tag{K3}$$

where $C_{\pi/2} \simeq 1.15$, H_{eq} is the Hubble parameter at matter-radiation equality and m_{pl} is the reduced Planck mass. We plot Eq. (K3) for different initial misalignments in Fig. 30, as a function of $\alpha = GM\mu$, for a 10 $M_\odot$ BH.

Somewhat separately, we can compare the energy density in a superradiant cloud to the DM energy density.

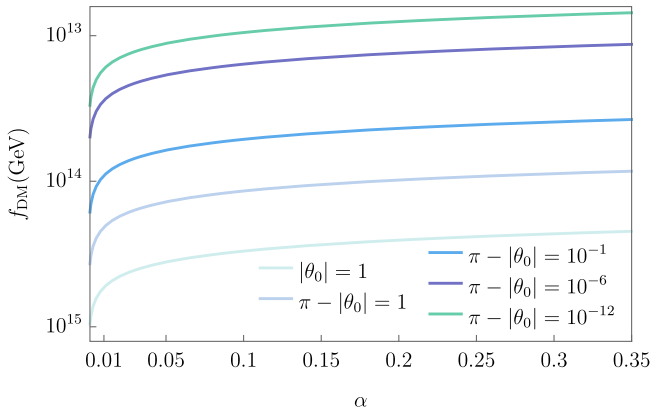


FIG. 30. The decay constant f of Eq. (K3) that gives the observed DM abundance today, as a function of $\alpha = GM\mu$ for a 10 $M_\odot$ BH. Note that the vertical axis is reversed. We are assuming a general cosine potential of the form $V(\varphi) = \mu^2 f^2 (1 - \cos(\varphi/f))$ and plot for different large initial misalignments from the top of the potential, following the results of [65]. We also plot for usual misalignment values of $|\theta_0| = 1, \pi - 1$.

The energy density of the cloud is $\rho_c \sim \theta^2 f^2 \mu^2$, up to an $\mathcal{O}(1)$ prefactor, which can be found to be

$$\begin{aligned}\rho_c &= \frac{1}{2}\dot{\varphi}^2 + \frac{1}{2}(\nabla\varphi)^2 + \frac{1}{2}\mu^2 + \frac{1}{4!}\frac{\mu^2}{f^2}\varphi^4 \\ &\sim \left(1 + \frac{\alpha^2}{2(\tilde{R}^+)^2} + \frac{\theta^2}{4!}\right)\theta^2 (f\mu)^2\end{aligned} \tag{K4}$$

where $\tilde{R}^+$ is given by Eq. (87). We estimate it to be

$$\begin{aligned}\rho_c &\sim 2 \times 10^{28} \frac{\text{GeV}}{\text{cm}^3} \left(\frac{\mu}{10^{-12}\ \text{eV}}\right)^2 \left(\frac{f}{10^{16}\ \text{GeV}}\right)^2 \left(\frac{\theta}{0.04}\right)^2 \\ &\sim 2 \times 10^{28} \frac{\text{GeV}}{\text{cm}^3} \left(\frac{M}{10\ M_\odot}\right)^{-2} \left(\frac{\alpha}{0.07}\right)^2 \\ &\quad \times \times \left(\frac{f}{10^{16}\ \text{GeV}}\right)^2 \left(\frac{\theta}{0.04}\right)^2\end{aligned} \tag{K5}$$

Even for the smallest f and α we show in our plots, this density is far larger than astrophysical DM densities. For example, in the SMBH parameter space shown in Fig. 11, $\rho_c \gtrsim 10^{14}$ GeV cm^{-3}.

[1] R. Penrose and R. M. Floyd, Nat. Phys. Sci. **229**, 177 (1971).
[2] Y. B. Zeldovich, JETP Lett. **14**, 180 (1971).
[3] C. W. Misner, Phys. Rev. Lett. **28**, 994 (1972).
[4] A. A. Starobinskii, Sov. Phys. JETP **37**, 28 (1973).
[5] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, and J. March-Russell, Phys. Rev. D **81**, 123530 (2010).

[6] A. Arvanitaki and S. Dubovsky, Phys. Rev. D **83**, 044026 (2011).
[7] I. M. Ternov, V. R. Khalilov, G. A. Chizhov, and A. B. Gaina, Izv. Vuz. Fiz. **21N9**, 109 (1978). [Sov. Phys. J. **21**, 1200 (1978)].
[8] T. J. M. Zouros and D. M. Eardley, Ann. Phys. (N.Y.) **118**, 139 (1979).
[9] S. L. Detweiler, Phys. Rev. D **22**, 2323 (1980).
[10] S. R. Dolan, Phys. Rev. D **76**, 084001 (2007).
[11] H. Yoshino and H. Kodama, Prog. Theor. Exp. Phys. **2014**, 043E02 (2014).
[12] A. Arvanitaki, M. Baryakhtar, and X. Huang, Phys. Rev. D **91**, 084011 (2015).
[13] R. Brito, V. Cardoso, and P. Pani, Classical Quantum Gravity **32**, 134001 (2015).
[14] R. Brito, V. Cardoso, and P. Pani, Lect. Notes Phys. **906**, 1 (2015).
[15] A. Arvanitaki, M. Baryakhtar, S. Dimopoulos, S. Dubovsky, and R. Lasenby, Phys. Rev. D **95**, 043001 (2017).
[16] J. G. Rosa and S. R. Dolan, Phys. Rev. D **85**, 044043 (2012).
[17] P. Pani, V. Cardoso, L. Gualtieri, E. Berti, and A. Ishibashi, Phys. Rev. D **86**, 104017 (2012).
[18] P. Pani, V. Cardoso, L. Gualtieri, E. Berti, and A. Ishibashi, Phys. Rev. Lett. **109**, 131102 (2012).
[19] W. E. East, Phys. Rev. D **96**, 024004 (2017).
[20] M. Baryakhtar, R. Lasenby, and M. Teo, Phys. Rev. D **96**, 035019 (2017).
[21] D. Baumann, H. S. Chia, J. Stout, and L. ter Haar, J. Cosmol. Astropart. Phys. 12 (**2019**) 006.
[22] R. Brito, V. Cardoso, and P. Pani, Phys. Rev. D **88**, 023514 (2013).
[23] R. Brito, S. Grillo, and P. Pani, Phys. Rev. Lett. **124**, 211101 (2020).
[24] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).
[25] S. Weinberg, Phys. Rev. Lett. **40**, 223 (1978).
[26] F. Wilczek, Phys. Rev. Lett. **40**, 279 (1978).
[27] H. Yoshino and H. Kodama, Prog. Theor. Phys. **128**, 153 (2012).
[28] H. Yoshino and H. Kodama, Classical Quantum Gravity **32**, 214001 (2015).
[29] H. Fukuda and K. Nakayama, J. High Energy Phys. 01 (2020) 128.
[30] M. J. Stott, arXiv:2009.07206 [Phys. Rev. D (to be published)].
[31] A. Gruzinov, High Energy Astrophys. Phenom. (**2016**), arXiv:1604.06422.
[32] S. Detweiler, Phys. Rev. D **22**, 2323 (1980).
[33] A. B. Gaina and I. M. Ternov, Sov. Phys. J. **31**, 830 (1988).
[34] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, England, 1995).
[35] J. E. McClintock, R. Narayan, and J. F. Steiner, Space Sci. Rev. **183**, 295 (2014).
[36] K. K. Y. Ng, O. A. Hannuksela, S. Vitale, and T. G. F. Li, Phys. Rev. D **103**, 063010 (2021).
[37] K. K. Ng, S. Vitale, O. A. Hannuksela, and T. G. Li, Phys. Rev. Lett. **126**, 151102 (2021).
[38] R. Brito, S. Ghosh, E. Barausse, E. Berti, V. Cardoso, I. Dvorkin, A. Klein, and P. Pani, Phys. Rev. D **96**, 064050 (2017).
[39] R. Brito, S. Ghosh, E. Barausse, E. Berti, V. Cardoso, I. Dvorkin, A. Klein, and P. Pani, Phys. Rev. Lett. **119**, 131101 (2017).
[40] S. J. Zhu, M. Baryakhtar, M. A. Papa, D. Tsuna, N. Kawanaka, and H.-B. Eggenstein, Phys. Rev. D **102**, 063020 (2020).
[41] P. W. Graham, J. M. Hogan, M. A. Kasevich, S. Rajendran, and R. W. Romani, Mid-band gravitational wave detection with precision atomic sensors, arXiv:1711.02225.
[42] D. Baumann, H. S. Chia, R. A. Porto, and J. Stout, Phys. Rev. D **101**, 083019 (2020).
[43] J. Zhang and H. Yang, Phys. Rev. D **101**, 043020 (2020).
[44] J. Zhang and H. Yang, Phys. Rev. D **99**, 064018 (2019).
[45] D. Baumann, H. S. Chia, and R. A. Porto, Phys. Rev. D **99**, 044001 (2019).
[46] G. G. di Cortona, E. Hardy, J. P. Vega, and G. Villadoro, J. High Energy Phys. 01 (**2016**) 034.
[47] H. Yoshino and H. Kodama, Prog. Theor. Exp. Phys. **2014**, ptu029 (2014).
[48] V. Cardoso, O. J. C. Dias, G. S. Hartnett, M. Middleton, P. Pani, and J. E. Santos, J. Cosmol. Astropart. Phys. (2018), 03 043.
[49] M. C. Miller and J. M. Miller, Phys. Rep. **548**, 1 (2015).
[50] R. Brito, V. Cardoso, and P. Pani, Phys. Rev. D **88**, 023514 (2013).
[51] R. Brito, S. Grillo, and P. Pani, Phys. Rev. Lett. **124**, 211101 (2020).
[52] C. S. Reynolds, Space Sci. Rev. **183**, 277 (2014).
[53] C. S. Reynolds, Classical Quantum Gravity **30**, 244004 (2013).
[54] M. Middleton, in *Astrophysics of Black Holes* (Springer Berlin Heidelberg, 2016), pp. 99–151.
[55] R. Brito, S. Ghosh, E. Barausse, E. Berti, V. Cardoso, I. Dvorkin, A. Klein, and P. Pani, Phys. Rev. D **96**, 064050 (2017).
[56] M. J. Stott and D. J. Marsh, Phys. Rev. D **98**, 083006 (2018).
[57] B. P. Abbott *et al.*, Phys. Rev. X **9**, 031040 (2019).
[58] B. P. Abbott *et al.*, Astrophys. J. **882**, L24 (2019).
[59] B. Zackay, T. Venumadhav, L. Dai, J. Roulet, and M. Zaldarriaga, Phys. Rev. D **100**, 023007 (2019).
[60] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, Phys. Rev. D **101**, 083030 (2020).
[61] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), arXiv:2010.14533.
[62] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), arXiv:2010.14527 [Phys. Rev. X (to be published)].
[63] B. Abbott, R. Abbott, T. Abbott, M. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari *et al.*, Phys. Rev. X **6**, 041015 (2016).
[64] R. Brito, S. Ghosh, E. Barausse, E. Berti, V. Cardoso, I. Dvorkin, A. Klein, and P. Pani, Phys. Rev. D **96**, 064050 (2017).
[65] A. Arvanitaki, S. Dimopoulos, M. Galanis, L. Lehner, J. O. Thompson, and K. Van Tilburg, Phys. Rev. D **101**, 083014 (2020).

[66] D. V. Semikoz and I. I. Tkachev, Phys. Rev. Lett. **74**, 3093 (1995).
[67] P. Sikivie and Q. Yang, Phys. Rev. Lett. **103**, 111301 (2009).
[68] L. Hui, J. P. Ostriker, S. Tremaine, and E. Witten, Phys. Rev. D **95**, 043541 (2017).
[69] D. Levkov, A. Panin, and I. Tkachev, Phys. Rev. Lett. **121**, 151301 (2018).
[70] B. Bar-Or, J.-B. Fouvry, and S. Tremaine, Astrophys. J. **871**, 28 (2019).
[71] L. Tsukada, T. Callister, A. Matas, and P. Meyers, Phys. Rev. D **99**, 103015 (2019).
[72] M. Isi, L. Sun, R. Brito, and A. Melatos, Phys. Rev. D **99**, 084042 (2019).
[73] H. Yoshino and H. Kodama, Prog. Theor. Exp. Phys. **2015**, 61E01 (2015).
[74] L. Sun, R. Brito, and M. Isi, Phys. Rev. D **101**, 063020 (2020); **102**, 089902(E) (2020).
[75] V. Dergachev and M. A. Papa, Phys. Rev. D **101**, 022001 (2020).
[76] C. Palomba *et al.*, Phys. Rev. Lett. **123**, 171101 (2019).
[77] A. Arvanitaki and A. A. Geraci, Phys. Rev. Lett. **110**, 071105 (2013).
[78] N. Aggarwal, G. P. Winstone, M. Teo, M. Baryakhtar, S. L. Larson, V. Kalogera, and A. A. Geraci, arXiv:2010.13157 [Phys. Rev. Lett. (to be published)].
[79] S. Dolan, Phys. Rev. D **76**, 084001 (2007).
[80] J. Aasi *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **32**, 115012 (2015).
[81] B. Behnke, M. A. Papa, and R. Prix, Phys. Rev. D **91**, 064007 (2015).
[82] B. Steltner, M. Papa, H.-B. Eggenstein, B. Allen, V. Dergachev, R. Prix, B. Machenschalk, S. Walsh, S. Zhu, and S. Kwang, Astrophys. J. **909**, 79 (2021).
[83] B. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. D **100**, 024004 (2019).
[84] B. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Astrophys. J. Lett. **882**, L24 (2019).
[85] C. S. Reynolds, Space Sci. Rev. **183**, 277 (2014).
[86] M. C. Miller and J. M. Miller, Phys. Rep. **548**, 1 (2015).
[87] S.-C. Yoon and N. Langer, Astron. Astrophys. **443**, 643 (2005).
[88] S. Woosley and A. Heger, Astrophys. J. **637**, 914 (2006).
[89] B. P. Abbott *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **34**, 044001 (2017).
[90] D. Reitze *et al.*, Bull. Am. Astron. Soc. **51**, 035 (2019).
[91] M. Punturo *et al.*, Classical Quantum Gravity **27**, 194002 (2010).
[92] S. Hild *et al.*, Classical Quantum Gravity **28**, 094013 (2011).
[93] B. Sathyaprakash *et al.*, Classical Quantum Gravity **29**, 124013 (2012); **30**, 079501(E) (2013).
[94] M. Maggiore *et al.*, J. Cosmol. Astropart. Phys. 03 (2020) 050.
[95] S. Hod, Phys. Rev. D **86**, 104026 (2012).
[96] C. A. Herdeiro and E. Radu, Phys. Rev. Lett. **112**, 221101 (2014).
[97] B. Ganchev and J. E. Santos, Phys. Rev. Lett. **120**, 171101 (2018).
[98] J. C. Degollado, C. A. Herdeiro, and E. Radu, Phys. Lett. B **781**, 651 (2018).
[99] J. C. Bustillo, N. Sanchis-Gual, A. Torres-Forn, J. A. Font, A. Vajpeyi, R. Smith, C. Herdeiro, E. Radu, and S. H. W. Leong, Phys. Rev. Lett. **126**, 081101 (2021).
[100] C. D. Bailyn, R. K. Jain, P. Coppi, and J. A. Orosz, Astrophys. J. **499**, 367 (1998).
[101] F. Ozel, D. Psaltis, R. Narayan, and J. E. McClintock, Astrophys. J. **725**, 1918 (2010).
[102] L. Kreidberg, C. D. Bailyn, W. M. Farr, and V. Kalogera, Astrophys. J. **757**, 36 (2012).
[103] K. Belczynski, G. Wiktorowicz, C. L. Fryer, D. E. Holz, and V. Kalogera, Astrophys. J. **757**, 91 (2012).
[104] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Astrophys. J. Lett. **896**, L44 (2020).
[105] T. A. Thompson, C. S. Kochanek, K. Z. Stanek, C. Badenes, R. S. Post, T. Jayasinghe, D. W. Latham, A. Bieryla, G. A. Esquerdo, P. Berlind, M. L. Calkins, J. Tayar, L. Lindegren, J. A. Johnson, T. W.-S. Holoien, K. Auchettl, and K. Covey, Science **366**, 637 (2019).
[106] B. Margalit and B. D. Metzger, Astrophys. J. Lett. **850**, L19 (2017).
[107] P. Amaro-Seoane *et al.* (LISA Collaboration), arXiv:1702.00786.
[108] J. Baker *et al.*, arXiv:1907.06482.
[109] S. Kawamura *et al.*, arXiv:2006.13545.
[110] K. Wette, Phys. Rev. D **85**, 042003 (2012).
[111] D. F. J. Kimball, S. Afach, D. Aybas, J. W. Blanchard, D. Budker, G. Centers, M. Engler, N. L. Figueroa, A. Garcon, P. W. Graham, H. Luo, S. Rajendran, M. G. Sendra, A. O. Sushkov, T. Wang, A. Wickenbrock, A. Wilzewski, and T. Wu, Springer Proc. Phys. **245**, 105 (2020).
[112] K. Blum, R. T. D'Agnolo, M. Lisanti, and B. R. Safdi, Phys. Lett. B **737**, 30 (2014).
[113] A. Hook and J. Huang, J. High Energy Phys. 06 (2018) 036.
[114] T. Sleator, E. L. Hahn, C. Hilbert, and J. Clarke, Phys. Rev. Lett. **55**, 1742 (1985).
[115] J. H. Chang, R. Essig, and S. D. McDermott, J. High Energy Phys. 09 (2018) 051.
[116] P. Carenza, T. Fischer, M. Giannotti, G. Guo, G. Martínez-Pinedo, and A. Mirizzi, J. Cosmol. Astropart. Phys. 10 (2019) 016; 05 (2020) E01.
[117] K. Hamaguchi, N. Nagata, K. Yanagi, and J. Zheng, Phys. Rev. D **98**, 103015 (2018).
[118] A. Sedrakian, Phys. Rev. D **93**, 065044 (2016).
[119] L. D. Luzio, F. Mescia, and E. Nardi, Phys. Rev. Lett. **118**, 031801 (2017).
[120] Y. Kahn, B. R. Safdi, and J. Thaler, Phys. Rev. Lett. **117**, 141801 (2016).
[121] S. Chaudhuri, K. D. Irwin, P. W. Graham, and J. Mardon, arXiv:1904.05806.
[122] R. Lasenby, Phys. Rev. D **102**, 015008 (2020).
[123] A. Berlin, R. T. D'Agnolo, S. A. R. Ellis, C. Nantista, J. Neilson, P. Schuster, S. Tantawi, N. Toro, and K. Zhou, J. High Energy Phys. 07 (**2020**) 088.
[124] C. A. Thomson, B. T. McAllister, M. Goryachev, E. N. Ivanov, and M. E. Tobar, Phys. Rev. Lett. **126**, 081803 (2021).

[125] A. Berlin, R. T. D'Agnolo, S. A. R. Ellis, and K. Zhou, arXiv:2007.15656.
[126] W. DeRocco and A. Hook, Phys. Rev. D **98**, 035021 (2018).
[127] I. Obata, T. Fujita, and Y. Michimura, Phys. Rev. Lett. **121**, 161301 (2018).
[128] H. Liu, B. D. Elwood, M. Evans, and J. Thaler, Phys. Rev. D **100**, 023548 (2019).
[129] A. Payez, C. Evoli, T. Fischer, M. Giannotti, A. Mirizzi, and A. Ringwald, J. Cosmol. Astropart. Phys. 02 (2015) 006.
[130] M. P. Hertzberg and E. D. Schiappacasse, J. Cosmol. Astropart. Phys. 11 (2018) 004.
[131] A. Dima and E. Barausse, Classical Quantum Gravity **37**, 175006 (2020).
[132] S. Sen, Phys. Rev. D **98**, 103012 (2018).
[133] H. Witek, V. Cardoso, A. Ishibashi, and U. Sperhake, Phys. Rev. D **87**, 043513 (2013).
[134] S. R. Dolan, Phys. Rev. D **87**, 124026 (2013).
[135] W. E. East and F. Pretorius, Phys. Rev. Lett. **119**, 041101 (2017).
[136] W. E. East, Phys. Rev. D **96**, 024004 (2017).
[137] W. E. East, Phys. Rev. Lett. **121**, 131104 (2018).
[138] A. Mathur, S. Rajendran, and E. H. Tanin, Phys. Rev. D **102**, 055015 (2020).
[139] J. Bonifacio, K. Hinterbichler, A. Joyce, and R. A. Rosen, J. High Energy Phys. 06 (2018) 075.
[140] J. Bonifacio and K. Hinterbichler, Phys. Rev. D **98**, 085006 (2018).
[141] M. M. Sternheim and J. F. Walker, Phys. Rev. C **6**, 114 (1972).
[142] L. Landau and E. Lifshitz, in *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed. (Pergamon Press, New York, 1977), pp. 121–124.
[143] N. Kampel, R. Peterson, R. Fischer, P. Yu, K. Cicak, R. Simmonds, K. Lehnert, and C. Regal, Phys. Rev. X **7**, 021008 (2017).
[144] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, Rev. Mod. Phys. **82**, 1155 (2010).